

# Regulation and Perception Concerning the Use of Health Data for Research in Australia

# Christine M O'Keefe<sup>1</sup>, Chris Connolly<sup>2</sup>

<sup>1</sup>CSIRO Mathematics, Informatics and Statistics, Canberra ACT, Australia <sup>2</sup>Galexia, Pyrmont NSW, Australia

#### Abstract

The primary objective of this review is to provide an overview of the issues involved in balancing privacy and access in the context of health research. Appropriate collection, management, linkage and interrogation of health data can play a vital role in improving individuals' health and wellbeing. However, the assembly and use of linked population, clinical and genetic health databases in the research and policy analysis environments raises privacy, confidentiality and ethical concerns.

The topic of our review is of current interest in the context of the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS) investment in the Population Health Research Network (PHRN), which aims to provide improved accessibility to health-related data for the research sector. This initiative is likely to attract new researchers to the field of population health, and the current review may assist them in taking account of privacy regulation and perceptions when designing study and consent processes.

Although there is little evidence of privacy complaints or breaches in health research, it seems clear that 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. In particular, significant concerns regarding consent and de-identification remain in the community.

Recent Australian Law Reform Commission recommendations 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 information provide some hope that de-identification can occur without a negative impact on data quality.

#### Keywords: Confidentiality; Privacy; Privacy Act

### **1** Introduction

As the health care industry moves from paper-based to electronic records, electronic data archives are accumulating in health care facilities and administrative agencies. Analysis of these health-systemusage and clinical data can yield information vital to effective health policy development and evaluation, as well as to enhanced clinical care through evidence-based practice and safety and quality monitoring. The result would be improved health and wellbeing. At the same time, the analysis of these health data archives must be conducted in such a way as not to compromise standards of privacy and confidentiality for individual health care consumers and providers, health care facilities and health data custodians. In this context, (information) privacy can be defined as the interest individuals have in controlling who can access and use their personal information, while confidentiality can be defined as an ethical principle (normally embodied in regulation) ensuring that information is accessible only to authorised users.

The electronic Journal of Health Informatics (ISSN: 1446-4381) is dedicated to the advancement of Health Informatics and information technology in healthcare. eJHI is an international Open Access journal committed to scholarly excellence and has a global readership in all health professions and at all levels.

© Copyright of articles originally published in www.ejhi.net under the Creative Commons Attribution 3.0 License is retained by the authors.

Clearly, the use of health information in research has both privacy and confidentiality implications.

Health information is amongst an individual's most sensitive and private information. However, because many diseases and conditions have a hereditary component, an individual's health information is also part of the health information of family members and so is sensitive and private to them as well. As discussed in [8], "The damaging effects brought on by a breach in the security of this information are endless. Third parties - employers, bankers, neighbours - could use this information to discriminate against and potentially ostracize an individual diagnosed with an "unpopular" disease or condition." Indeed discrimination by health insurance companies is an example, which could result in considerable harm to an individual. Privacy legislation and codes of practice must be adhered to as a minimum requirement and health data custodians' responsibilities to protect confidential data must be supported.

Data analysis and data mining tools are constantly being developed to be more powerful and to extract more information from data. Even if an analyst does not have direct access to the data, just the results of the analyses or mining can be enough to reveal private information [1-7].

In this paper we review privacy regulation and privacy perception in Australia in the context of the use of health data in research and policy analysis. We observe that privacy concerns regarding the use of health data extend beyond strict privacy law compliance. Note that we do not consider privacy and confidentiality issues regarding health databases except in the context of the use of health data in research and policy analysis, normally involving statistical and other analysis of single or linked data sets.

In the rest of this section, we review three examples of current and potential future initiatives, which depend on achieving an appropriate balance between access and use of data and privacy and confidentiality protection. Section 2 provides an overview of the privacy regulatory environment in Australia and Section 3 reviews available evidence of privacy perception in Australia. Section 4 provides a more detailed discussion of de-identification, consent and bias, as well as a review of arguments that (excessive) privacy regulation has a negative effect on public health research. Section 5 gives a conclusion.

### 1.1 Example - Population Health Research Network

Through its National Collaborative Research Infrastructure Strategy (NCRIS), the Australian Government is making significant investments over 2005-2011 to provide researchers with major research facilities, supporting infrastructures and networks necessary for world-class research, see [9]. Investment in the Population Health Research Network (PHRN) recognises that Australia is an international leader in the scope and extent of health-related data collected at the population level. New and emerging technologies underpin the potential to integrate and link these data sets to provide a valuable new resource for monitoring the health of the population and the effectiveness of health services, and for research. The PHRN has therefore been established to provide Australian researchers with access to linkable, de-identified data from a diverse and rich range of health datasets, across jurisdictions and sectors. This will support nationally and internationally significant population-based research that will improve health and enhance the delivery of health care services in Australia, see [10]. The research outcomes will have both clinical and administrative impact.

# **1.2 Example - Pharmacovigilance and Post-Market Drug Surveillance**

Pharmacovigilance is defined as the detection, assessment, understanding and prevention of adverse effects, particularly long-term and short-term side effects, of medicines. The case for a routine pharmacovigilance system is becoming stronger as the number of examples of unexpected drug effects resulting in recalls or litigation increases [11]. This could be part of a wider post-market drug surveillance system, which would routinely evaluate the safety, efficacy and cost-effectiveness of drugs once marketed and sold [12].

A pharmacovigilance or post-market drug surveillance system would best operate by analysing population-wide, linked data on medications, hospitalisations, cancers, birth defects and other selected disabilities, cardiovascular diseases, emergency department visits and deaths. Such data exist in very few countries, notably including Australia [13]. Proofof-concept studies to demonstrate the feasibility of automated signalling of potential adverse events associated with medicine use from Australian administrative health data have already been conducted, see for example [14].

#### 1.3 Example – National Bowel Cancer Screening Program

Population screening programs have been established in Australia for breast, cervical and bowel cancer [15]. These programs aim to reduce morbidity and mortality through early detection.

The collection of complete, high quality data is essential for follow-up of positive tests within each program and for evaluating overall program efficacy, performance and quality. Analysis of the collected program data can contribute vital knowledge to the underlying evidence base. However, information relating to any program participant is collected in several different locations and so is spread across a number of different databases. The assembly of the relevant information for an individual depends on the completion and transfer of a paperbased or web-based form, with loss of information if the form is not completed or is lost. Both types of forms impose a compliance burden on health professionals.

Data linkage has the potential to provide an effective way to achieve routine collection of high quality data for population screening program follow-up and evaluation. At the same time, it would reduce the compliance burden on health professionals by removing the need for four of the current five paperbased forms. Several implementation options and the impact of each on confidentiality and privacy are discussed in [16].

# **2 Privacy Regulation in Australia**

In this section, we give an overview of the privacy regulatory environment in Australia, including the additional layer of health-specific privacy regulation.

The regulation of privacy in Australia is complex. Relevant privacy laws include: Commonwealth, state and territory privacy legislation, health specific privacy legislation, privacy and confidentiality provisions within other laws, codes of conduct; research guidelines and the common law.

In 2006, the Australian Law Reform Commission began an inquiry into the extent to which the Privacy Act 1988 (Commonwealth) and related laws continue to provide an effective framework for the protection of privacy in Australia. The final report, For Your Information: Australian Privacy Law and Practice (ALRC 108) was delivered to the Australian Attorney General on 30 May 2008 [17] and the government released the first stage of its response on 14 October 2009 [18]. Exposure draft legislation including an important element of the first stage response, the Australian Privacy Principles, was released on 24 June 2010 [19].

In this section, we provide a discussion of common themes and issues that emerge from this patchwork of privacy regulation in Australia. We also provide a summary of the main recommendations related to health information contained in the ALRC's final report and the Government's responses to these.

### 2.1 General Privacy Laws

The general privacy legislation currently in place in Australia is shown in Figure 1. The State and Territory legislation in this list generally applies to the activities of State and Territory public sector agencies, while the Commonwealth legislation applies to both the Commonwealth public sector, and significant parts of the private sector. However two different standards of privacy protection exist in the Commonwealth legislation, namely the Information

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

On 1 November 2010 the Office of the Privacy Commissioner was integrated into the Office of the Australian Information Commissioner (OAIC). See http://:www.oaic.gov.au.

*Figure 1: General privacy legislation currently in place in Australia.* 

Privacy Principles (IPPs) which apply to Commonwealth and ACT government agencies and the National Privacy Principles (NPPs) which apply to parts of the private sector (those that earn more than \$3 million annually) and all health service providers. There is some inconsistency between the IPPs and the NPPs, which was addressed in the Australian Law Reform Commission recommendation that the IPPs and NPPs be replaced by a new set of Uniform Privacy Principles (UPPs). The issue is addressed by the proposed new Australian Privacy Principles [19].

The general privacy laws regulate the collection and handling of personal information such as credit information and medical and government records. They do not address the protection of people's physical selves against invasive procedures, the security and privacy of mail, telephones, e-mail and other forms of communication, and the setting of limits on intrusion into the domestic and other environments such as the workplace or public space.

### 2.2 Common Themes and Issues

Despite the patchwork of laws, some common themes and issues emerge. Key themes relevant to this paper are discussed in the next paragraphs.

• **Definition of personal information**. There are some minor inconsistencies between the definitions of personal information in the different Acts, but generally it is defined to be any form of information about an individual whose identity is apparent or can reasonably be ascertained from the information. This definition is regularly tested in privacy complaints, and different interpretations of the word "reasonably" appear to have emerged in different jurisdictions. For example, in a recent complaint regarding the use of individuals' deidentified health data in which there was a small possibility of re-identification using crossmatching, the Privacy Commissioner concluded that because the data was de-identified it did not fall within the definition of personal information outlined in the Privacy Act [20].

- **Definitions of use and disclosure**. Within the privacy regulation framework, there is a difference in the provisions concerning use and disclosure. For example, personal information may be used without explicit consent under certain notice, data quality and security requirements. Although it can be difficult to determine whether a given scenario involves use, disclosure or even both, the judicial interpretation of the term use to date has been very strict, and may in some jurisdictions include data encryption.
- Approaches to de-identification. Some privacy laws include specific provisions for deidentification and de-identified data, limited to certain types of research. De-identification is discussed below.
- Consent for disclosure. There are considerable inconsistencies in Australian privacy legislation surrounding the concept of consent for disclosure of personal information to a third party. Although most laws contemplate both explicit and implied consent, this has been the subject of varied interpretation by Privacy Commissioners and the NSW courts [21-23]. Strict application of the consent provisions in some cases has forced researchers to seek alternative methods of gaining access to data without triggering consent provisions. Some implications of consent provisions are discussed below.

### 2.3 Health-Specific Privacy Laws

In addition to specific provisions for health information in the general privacy laws, there is a further layer of health-specific privacy legislation which

Jurisdiction	Legislation	Regulator	
Commonwealth	Privacy Act 1988	Australian Privacy Commissioner	
Australian Capital Territory	Health Records (Privacy and Access) Act 1997	Community and Health Services Complaints Commissioner	
New South Wales	Health Records and Information Privacy Act 2002	Public Sector – internal review Private Sector – Privacy NSW	
Northern Territory	None currently in place	Not applicable	
Queensland	Information Privacy Act 2009	Health Quality and Complaints Commission	
South Australia	Code of Fair Information Practice	Not applicable	
Tasmania	None currently in place	Not applicable	
Victoria	Health Records Act 2001	Health Services Commissioner	
Western Australia	None currently in place	Not applicable	

Figure 2: Health-specific privacy legislation currently in place in Australia.

adds complexity to the consideration and analysis of the general privacy laws. The main such laws are shown in Figure 2. These health-specific laws seek to regulate health-related personal information.

We now discuss some of the implications and issues related to these health specific privacy laws.

- Sensitive information. In some laws health information is included in a special class of sensitive information and is thus subject to stricter provisions, including requirements for explicit (not implied) consent and additional restrictions on disclosure.
- **Health information**. Health information is represented differently in different laws, either included in personal information or defined separately. The definitions generally include genetic information and family hereditary information.
- **Health research**. Detailed provisions (including references to additional guidelines) appear in some laws for health research and for health research on de-identified data.
- **Health identifiers**. Additional requirements for the use and disclosure of health identifiers may be included. These generally prohibit private sector providers from using public sector health identifiers.
- **Health data linkage**. Additional requirements for health data linkage may be included. For example, NSW laws require consumers to opt-in to a data linkage program.

It is important to note that the provisions on deidentification in health privacy legislation are closely linked to the debate concerning the definition of personal information.

### 2.4 Other Legislative Privacy Requirements

Additional privacy requirements are contained in specific health laws. These may place additional burdens on health data custodians and in some circumstances may place specific restrictions on linking data sets. Data custodians will usually be very aware of the specific restrictions that apply to their data.

An additional layer of privacy regulation is in place for health research without individual consent of participants, in the form of enforceable guidelines. These guidelines are applied in the first instance by ethics committees and data custodians, but are also the subject of regulatory oversight by Privacy Commissioners. The relevant key guidelines are:

- Privacy Act Section 95 Guidelines (Commonwealth agencies). Under Section 95 of the Privacy Act, the National Health and Medical Research Council (NHMRC) has issued guidelines for the protection of privacy in the conduct of medical research. These Guidelines apply to medical research using information held by Commonwealth agencies where identified information needs to be used without consent.
- Privacy Act Section 95A Guidelines (private sector). These Guidelines apply to private sector organisations in circumstances where for the purposes of research, compilation or analysis of statistics, relevant to public health or public safety, an organisation must collect, use or disclose health information.
- Medicare and Pharmaceutical Benefits Program Privacy Guidelines. These Guidelines are made under Section135AA of the National Health Act 1953. Guideline 4 provides that disclosure for research purposes must conform to secrecy provisions in the Health Insurance Act 1973 and National Health Act 1953. In addition, identified claims information may only be disclosed to researchers where either individuals have given their free and informed consent to the use of the information in the research; or disclosure is made for medical research conducted in accordance with Section 95 Medical Research Guidelines issued by NHMRC.

An NHMRC analysis of the use of these Guidelines in practice [24] found that:

- Among consumers, there is a low awareness of privacy legislation and difficulty in distinguishing between confidentiality and privacy.
- Consumers were uncertain about providing consent for the use of their data.
- Health professionals tend to equate confidentiality with privacy and always maintain patient confidentiality. They were concerned that privacy legislation could delay correspondence between practitioners.
- Researchers reported difficulties in getting access to registries and inconsistent decision making by human research ethics committees regarding access and release of information.
- Data custodians believe that there is no need for researchers to have access to identified data and feel they get the same benefit from de-identified information.
- Ethics committees believe that interpreting privacy legislation is complex and have the strongest

opposition to researchers having access to health information without consent.

The broad theme in these Guidelines is that they enable health research to be carried out without consent, in circumstances where obtaining consent would be impracticable, and with a preference for strong de-identification. However, 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.

### 2.5 Australian Law Reform Commission Privacy Inquiry

The main recommendation related to health research contained in the ALRC's final report [17] is Recommendation 65-1 that the various existing Guidelines on privacy and research should be replaced by a formal set of Research Rules issued by the Privacy Commissioner.

In their "first stage" response to the ALRC's proposals [18], the Government accepted this central reform with an amendment, recommending that the National Health and Medical Research Council (NHMRC) in conjunction with other appropriate bodies (such as the Australian Research Council and Universities Australia) should write the Research Rules, rather than the Privacy Commissioner. The Government agreed that these Research Rules should conform to the National Statement on Ethical Conduct in Human Research and noted that the Privacy Commissioner should be responsible for approving the Research Rules. A second central proposal supported by the Government is to expand the research provisions to allow such handling for any research in the public interest, not just health and medical research. One important parameter of the current regime will be maintained, namely: that the public interest in research must 'substantially outweigh' the protection of privacy.

The key accepted recommendations in the context of health research can be summarised as follows:

- The definition of 'personal information' in the Privacy Act should be revised to '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 Privacy Commissioner should develop and publish guidance on the meaning of 'identified or reasonably identifiable'.
- The National Health and Medical Research Council should lead the issuing of one set of rules under

the research exceptions to the 'Collection' and 'Use and Disclosure' principles.

- The arrangements relating to the collection, use or disclosure of personal information without consent in the area of health and medical research should be extended to cover the collection, use or disclosure of personal information without consent in human research more generally.
- 'Research' should include the compilation or analysis of statistics
- For exceptions to the 'Collection' and 'use and Disclosure' principles, it should be sufficient that the public interest in the research outweighs the public interest in maintaining the level of protection provided in the Privacy Act.
- Where 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.
- The Research Rules issued by the Privacy Commissioner 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.
- That infrastructure to allow the linkage of personal information for research purposes should be subject to a Privacy Impact Assessment.
- The Privacy Commissioner should develop Research Rules that address the question of 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.

Thus, the Australian Government accepted most of the ALRC's proposals in principle regarding Health Research. The main qualification to the recommendations simply noted that the National Health and Medical Research Council with other appropriate bodies would be responsible for addressing most concerns in the creation of Research Rules, which the Privacy Commissioner would have the role of approving. The main impacts of the reforms are to simplify and standardise the regulations, broaden their applicability and streamline the associated procedures. Several of the reforms will result in reduced uncertainty about the scope of applicability of the Privacy Act in health-related research.

An exposure draft of the first, and most fundamental, part of the rewritten Privacy Act, the Australian Privacy Principles, was released on 24 June 2010 [19].

# 3 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. Three main issues emerged from the review, namely de-identification, consent and participation (related to moral rights) and we focus on them.

### **3.1 De-Identification**

In reporting on Australian Government Department of Health and Ageing qualitative research, Taylor [25] noted: "It's really important I think here to stress 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 or are that familiar with. But interestingly here, they were expecting that consent for this kind of data would be expected at ethics committee approval level and that they would not need to be approached personally if their data which are truly de-identified is to be shared. Basically, they see no threat to their privacy if data are de-identified and see it as a sensible and an efficient use of that data." There was an indication that a smaller, but still significant proportion of the population, believed that their permission should be sought, even where the data would be de-identified.

In a poll conducted by the Australian Medical Association in 2005 [26], 60% of respondents reported that they were slightly or very concerned about the de-identification process.

### 3.2 Consent

The Office of the Privacy Commissioner has conducted surveys into privacy attitudes in Australia, in 2001, 2004 and 2007 [27-29]. In the first two surveys, about 64% of respondents said that consent should be sought for the use of de-identified data for research, while 33% said that use without consent was fine. In the 2007 survey, 51% said that consent should be sought, while 46% said that consent should not be sought.

In the AMA poll [26], 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 only concerned de-identified data. For results of the US National Consumer Health Privacy Survey 2005, see [30].

In contrast, the Australian Government Department of Health and Ageing research [25] found that consumers supported the use of data in research and registers provided the data was de-identified and the purpose was legitimate and worthwhile. If identified data was to be used, consumers expected to be informed and their consent sought.

The Australian Consumer's Association (ACA) call for notification rather than consent: "*People also* have the right to know if their records are being used for any other purpose even in a de-identified form" [31].

It is interesting to compare individuals' clear preference for consent to the use of de-identified data for research with the legal requirements, where health research guidelines permit the use of de-identified and even sometimes identifiable health data for research without consent, subject to certain conditions. This difference may influence ethics committee decisions, data custodian attitudes and privacy law reform.

High quality health and related data are extremely expensive to collect, and therefore it is important that the maximum benefit is gained from all collected data. This aim brings with it real challenges regarding compliance with consent provisions in privacy regulations, including that at the time of collection of the data it is sometimes not possible to know the full range of uses for which it will be desirable to use the data. Because of the nature of research, new technologies emerge and new research questions follow from the answers to initial questions. A too-restrictive requirement for informed consent can seriously limit the utility of a valuable resource and increase the total burden of survey and study participation.

### **3.3 Participation**

In this context, the use of an individual's health data for research is viewed as participation by that individual in the research. An individual may have an objection to participation in research on moral grounds even when there is no risk of identification or personal consequences. It is the purpose of the research that is important.

This concern has been well-expressed in the Recommendations from PRIVIREAL to the European

Commission [32], "... First, there is a "narrow" conception of privacy, according to which privacy is concerned merely to protect the identity of the data subject. But there is a contrary "broad" concept of privacy according to which privacy seeks to give data subjects control over personal information on them that can negatively affect their physical, psychological and moral integrity. Under the narrow conception, to render personal data anonymous is to remove any interest that the data subject has in the use of it. Under the broad conception, rendering it anonymous merely protects against certain abuses of that data. (It does not, e.g., prevent personal information obtained from devout Catholics being used to develop chemical contraceptive methods, which is arguably contrary to their moral integrity)... (It needs to be emphasised that under the broad conception, anonymisation [de-identification] of which the data subject is not informed, in principle, threatens privacy rather than protecting it, because it results in the data subject losing all possi*bility of control of processing*) ..."

The AMA poll [26] 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.

As well as raising concerns about consent and deidentification, Christopher Newell [33] argued that the use of data for research may raise ethical and spiritual issues. He gave examples in which the consequences to a community of participation in research were extremely serious but had nothing to do with personal privacy.

# 4 Discussion - 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, and thus the rights of researchers with respect to access to health data for research to benefit the wider community. However there is a range of views on where the appropriate balance lies (see, for example, [12, 34-35].

The debate largely focuses on different ways to rate respective interests, to strike a fair balance see, for example [36]. These approaches could be characterised as policy-centric. The complementary technology-centric approaches are just beginning to attract attention. For example, a review of current Australian technological approaches to the problem of enabling the use of health data for research and policy analysis while protecting privacy and confidentiality is provided in [37]. In fact, a mutually satisfactory balance is likely to be achieved by a combination of policy-centric and technology-centric measures.

It is interesting to note that the ALRC recommended a shift of the balance towards the public interest in the research and consequently away from individual privacy protection. However, it should be noted that the Government's response to this recommendation preferred a test, which required the public interest in the research activity to substantially outweigh the public interest in maintaining the level of privacy stipulated in the Privacy Act. This test provides a balance in favour of research proceeding, while taking into account the circumstances where personal data might be handled without an individual's consent.

In this section, we provide a more detailed discussion of de-identification, consent and bias, as well as a review of arguments that (excessive) privacy regulation has a negative effect on public health research. These are all factors, which must be considered in any debate about the balance between privacy and research.

# 4.1 De-Identification

De-identification is a very complex issue surrounded by lack of clarity and standard terminology. This is important because it underpins many health information privacy guidelines and legislation.

First, it is often not at all 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 data set cannot be identified from the data – though in turn it can be unclear what this means. Of course, simply removing nominated identifiers is often insufficient to ensure that individuals represented in a data set cannot be identified - it can be a straightforward matter to match some of the available data fields with the corresponding fields from external data sets, and thereby obtain enough information to determine individuals' names either uniquely or with a low uncertainty. In addition, sufficiently unusual records in a database without nominated identifiers can sometimes be recognised. 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 [38] avoids the term 'de-identified data' as its meaning is unclear. Instead, it proposed that data may be collected, stored or disclosed in three mutually exclusive forms, as follow:

1. individually identifiable data, where the identity of a specific individual can reasonably be ascertained. Examples include the individual's name, image, date of birth or address;

2. re-identifiable data, from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets;

3. non-identifiable data, which have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individuals can be identified. A subset of non-identifiable data is those that can be linked with other data so it can be known that they are about the same data subject, although the person's identity remains unknown.

One problem here is that is not difficult to imagine datasets, which do not fit into any of these categories. For example, a dataset of detailed health information from which all identifiers have been permanently removed may still allow the identification of an individual by matching to an external database, so these data could not fit into any of these categories.

On the other hand, the US Health Insurance Portability and Accountability Act 1996 (US) (HIPAA) [39] provides a useful legislative test for deidentification that provides certainty for the research community and for ethics committees. Under the test, protected health information can be determined to be non-identifiable if either a suitably qualifited person certifies that the risk is very small that the information could be used to identify an individual, or if specified identifiers are removed and the custodian does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual.

The de-identification test contained in HIPAA is a useful example of a legislative test that provides certainty for the research community. It allows for a small risk of re-identification through reverse engineering or multiple, complex queries. The lack of such a test in Australia places the onus on individual judges and Privacy Commissioners to interpret the "reasonable" test in the definition of personal information in privacy legislation. In practice, the HIPAA approach would be simpler to implement, and provide greater certainty for researchers.

On the other hand there may be a significant burden of compliance – if an organisation has many data sets then it would take a great deal of time for a person to perform the tasks outlined.

### 4.2 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.

Informed consent and opt-in is a common model for clinical trials, for example, where the risk is normally predominantly to the participating individual. However, in the case of population health research, the findings will often be implemented for the whole population. In these cases informed consent and opt-in may not be appropriate models because nonparticipation can introduce bias and therefore affect the applicability of the 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 et al. [40] 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 [41-42] 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."

It is important that population health researchers seek to minimise selection bias arising from consent processes in their analysis. It is also important that population health data access managers and users of population health research results seek to minimise the occurrence and impact of such bias.

# 4.3 4.3 Arguments that (excessive) Privacy Regulation has a Negative Effect on Public Health Research

- Efficiency of health research: The perception is that overheads resulting from privacy regulation hamper research efficiency in Australia [43-44]. Implementation of ALRC Recommendations 65-3, 65-6 and 66-2 may potentially exacerbate this situation. In an increasingly risk-averse environment and consequent conservative interpretation of regulations, it is likely that privacy considerations will continue to impact research efficiency as resources are diverted from research to administration and governance.
- Quality of health research: The fear is that selection effects from privacy-related processes including consent will lead to results bias. Implementation of ALRC Recommendations 66-1 and 66-3 may potentially exacerbate this situation. The issue of bias was discussed in Section 4.2.
- 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. The argument was expressed in [45] as follows: ...``imagine the case of a person who had been treated with a certain drug in childhood and a hypothesis arose that the treatment might have adverse effects down the track that could perhaps be related to cancer, one would be asking people, with very little evidence yet of a real effect, to have their data on their earlier treatment linked with the cancer registry. This can put people in a position of fairly helpless anxiety in the intermediate period."
- Interests of the community versus the individual: The perception is that excessive privacy regulation results in the interests of the individual being placed above those of the community, by denying the community the full potential benefits of health research based on more complete data. The ALRC Recommendation 65-4 potentially ameliorates this situation. 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 ..." [41].
  - "This relates to human rights because generally the people who are excluded from studies are the most marginalized" [42].

- "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" [43].

# **5** Conclusion

Although there is little evidence of privacy complaints or breaches in health research, it seems clear that 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. In particular, 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 a quarter and a third of the population) for some time to come.

Will these community concerns impact upon health research? Ultimately, decisions on research are made by ethics committees applying guidelines that allow some measure of "balance" between privacy and research. The decision is therefore taken out of the hands of individual consumers. Nevertheless, 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 Australian Law Reform Commission, 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 Australian Law Reform Commission 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 information [37] provide some hope that de-identification can occur without a negative impact on data quality.

These recommendations should be viewed in light of the Australian Government's first stage response, which suggest that the Research Rules would be created by the National Health and Medical Research Council and adhere to the National Statement on Ethical Conduct in Human Research.

This article has provided an overview of the issues involved in balancing privacy and access in the context of health research. It is hoped that this will be useful to researchers in population health in particular, in that it may assist them in taking account of privacy regulation and perceptions when designing study and consent processes.

# Acknowledgements

The main part of this work was conducted as a project commissioned by the CSIRO Preventative Health National Research Flagship. The authors thank the project stakeholder representatives who gave valuable inputs during the consultations. Additional research and guidance was provided by Galexia consultants Peter van Dijk and Francis Vierboom. The authors also thank the anonymous reviewers whose thoughtful comments and questions have led to an improved article.

An abridged version of this article has appeared in the Medical Journal of Australia [46]. The current paper provides details and information, which are not included in the abridged version.

# References

- 1 Gomatam S, Karr AF, Reiter JP, Sanil A. Data dissemination and disclosure limitation in a world without microdata: a risk-utility frame-work for remote access servers. Statistical Science 2005; 20: 163-177.
- 2 O'Keefe CM, Good NM. Regression output from a remote analysis server. Data & Knowledge Engineering 2009; 68: 1175-1186.
- 3 Reiter JP. Model diagnostics for remote-access regression servers. Statistics and Computing 2003; 13: 371-380.
- 4 Reiter JP, Kohnen CN. Categorical data regression diagnostics for remote servers. Journal of Statistical Computation and Simulation 2005; 75: 889-903.
- 5 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.pd
  - f (accessed Sep 2010).

- 6 Sparks R, Carter C, Donnelly J, Duncan J, O'Keefe CM, Ryan L. 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).
- 7 Sparks R, Carter C, Donnelly J, O'Keefe CM, Duncan J, Keighley T, McAullay D. Remote access methods for exploratory data analysis and statistical modelling: privacy-preserving analytics. Computer Methods and Programs in Biomedicine 2008; 91: 208-222.
- 8 Carl, K., 2010. It's Personal: Privacy Concerns Associated With Personal Health Records. IS-JLP, 5, p.533–603.
- 9 Australian Government Department of Innovation, Industry, Science and Research. Strategic Roadmap for Australian Research Infrastructure, 2008. http://ncris.innovation.gov.au/Documents/2008\_

Roadmap.pdf (accessed Sep 2010). 10 Population Health Research Network: building

- 10 Population Health Research Network: building data linkage infrastructure for Australia. http://www.phrn.org.au (accessed Sep 2010).
- 11 World Health Organization. The importance of pharmacovigilance: Safety monitoring of medicinal products, Geneva 2002. http://whqlibdoc.who.int/hq/2002/a75646.pdf (accessed Dec 2010).
- 12 Kelman CW, Pearson S-A, O'Day R, Holman CD'A, Kliewer EV, Henry DA, Evaluating medicines: let's use all the evidence, Medical Journal of Australia 2007; 186: 249-252.
- 13 Prime Minister's Science, Engineering and Innovation Council (PMSEIC) working group on data for science. From data to wisdom: pathways to successful data management for Australian science, Canberra 2006 http://www.dest.gov.au/NR/rdonlyres/D15793B2
  -FEB9-41EE-B7E8-C6DB2E84E8C9/15103/From\_Data\_to\_Wisdom \_Pathways\_data\_man\_forAust\_scie.pdf
- 14 Jin H-D, Chen J, He H, Williams G, Kelman C, O'Keefe CM Mining Unexpected Temporal Associations. IEEE Transactions on Information Technology in Biomedicine. 2008; 12(4): 488-500.
- 15 Department of Health and Ageing. http://cancerscreening.gov.au (accessed Dec 2010).

- 16 O'Keefe CM, O'Dwyer M, Hansen D, Young GP and Macrae F. Using data linkage to put the patient at the centre of the Australian National Bowel Screening Program. In H. Grain (ed.), Proceedings of the Health Informatics Conference HIC2008, Health Informatics Society of Australia, (ePoster).
- 17 Australian Law Reform Commission. For Your Information: Australian Privacy Law and Practice (ALRC Report 108). Sydney: ALRC, 2008. http://www.alrc.gov.au/publications/report-108 (accessed Sep 2010).
- 18 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).
- 19 Parliament of Australia, Senate. Exposure Drafts of Australian Privacy Amendment Legislation. Canberra: SFPAC, 2010. http://www.aph.gov.au/senate/committee/fapa\_ct te/priv\_exp\_drafts/ (accessed Sep 2010).
- 20 Office of the Privacy Commissioner. Privacy Commissioner concludes investigation into CAMM Pacific and Health Communications Network Limited [media release 11 May 2005]. http://www.privacy.gov.au/news/media/05\_02.ht ml (accessed Sep 2010).
- 21 New South Wales Administrative Decisions Tribunal. Vice Chancellor, Macquarie University v FM (2003) NSWADTAP 43. Sydney: Office of the NSW Privacy Commissioner, 2003. http://www.lawlink.nsw.gov.au/lawlink/privacyn sw/ll\_pnsw.nsf/pages/PNSW\_07\_cnadtap43 (accessed Sep 2010).
- 22 New South Wales Administrative Decisions Tribunal. KJ v Wentworth Area Health Service (2004) NSWADT 84. Sydney: Office of the NSW Privacy Commissioner, 2004. http://www.lawlink.nsw.gov.au/lawlink/privacyn sw/ll\_pnsw.nsf/pages/PNSW\_07\_cnadt84 (accessed Sep 2010).
- 23 New South Wales Administrative Decisions Tribunal. MT v Director General, NSW Department of Education and Training (2004) NSWADT 194. Sydney: Australasian Legal Information Institute, 2004. http://www.austlii.edu.au/au/cases/nsw/NSWAD T/2004/194.html (accessed Sep 2010).

- 24 Hill D. Consumer attitudes to privacy in an ehealth environment [speech]. Sydney: National Health Information Summit, 2004.
- 25 Taylor J. Consumer attitudes to privacy in an ehealth environment [speech]. Sydney: National Health Information Summit, 2004.
- 26 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).
- 27 Office of the Privacy Commissioner, Australia. Privacy and the Community, 2001. Sydney: Office of the Privacy Commissioner, 2001. http://www.privacy.gov.au/materials/types/resear ch/view/6614 (accessed Sep 2010).
- 28 Office of the Privacy Commissioner, Australia. Community Attitudes Towards Privacy 2004. Sydney: Office of the Privacy Commissioner, 2004. http://www.privacy.gov.au/materials/types/resear ch/view/6615 (accessed Sep 2010).
- 29 Office of the Privacy Commissioner, Australia. Community Attitudes to Privacy 2007, Sydney: Office of the Privacy Commissioner, 2007. http://www.privacy.gov.au/materials/types/down load/8820/6616 (accessed Sep 2010).
- 30 Forrester Research. National Consumer Health Privacy Survey 2005. Oakland, CA: California HealthCare Foundation, 2005. http://www.chcf.org/publications/2005/11/nation al-consumer-health-privacy-survey-2005 (accessed Sep 2010).
- 31 Ballenden N. Transforming Australia's Health System [speech]. Sydney: National Health Information Summit, 2004.
- 32 Privacy in Research Ethics and Law. Recommendations from PRIVIREAL to the European Commission. Sheffield, UK: PRIVIREAL, 2003.
  http://www.privireal.org/content/recommendatio ns/#Recc (accessed Sep 2010).
- 33 Newell C. Health IM&ICT for public health and population research: a consumer perspective [speech]. Sydney: National Health Information Summit, 2004.
- 34 Peto J, Fletcher O, Gilham C. Data protection, informed consent, and research. British Medical Journal 2004; 328: 1029-1030.

- 35 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\_fin al.cfm (accessed Sep 2010).
- 36 Room S. Data protection, informed consent, and research: Data Protection Act does not bar medical research [letter]. British Medical Journal 2004; 328: 1437.
- 37 O'Keefe CM. Privacy and the use of health data— reducing disclosure risk. Electronic Journal of Health Informatics 2008; 3: e5.
- 38 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).
- 39 Health Insurance Portability and Accountability Act 1996 (US) http://www.legalarchiver.org/ hipaa.htm (accessed 2 Aug 2011)
- 40 Woolf SH, Rothemich SF, Johnson RE, Marsland DW. Selection bias from requiring patients to give consent to examine data for health services research. Archives of Family Medicine 2000; 9: 1111-1118.
- 41 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).
- 42 Stanley F. Australian Health Information Coun-

cil (AHIC) conference 2004 [speech]. Sydney: National Health Information Summit, 2004.

- 43 Gun R. Privacy law is kneecapping epidemiological research. Australasian Epidemiologist 2005; 12.1: 2-4.
- 44 Skene L. Proliferating ethics committees and privacy legislation: new fetters on scientific research. Australasian Epidemiologist 2005; 12.1: 16-18.
- 45 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/commttee /S8383.pdf (accessed Sep 2010).
- 46 O'Keefe CM, Connolly, CJ. Privacy and the use of health data for research. Medical Journal of Australia. 2010; 193(9): 537-541.

# Correspondence

Dr Christine M O'Keefe BSc MBA PhD

CSIRO Mathematics, Informatics and Statistics GPO Box 664 Canberra, ACT 2601, Australia

Phone: +61 2 6216 7021 Fax: +61 2 6216 7111

www.csiro.au/people/Christine.OKeefe.html

Christine.OKeefe@csiro.au