

Public Support for Data-based Research

To Improve Health

**A discussion paper based on the proceedings of a
Menzies Foundation Workshop
16th August, 2013**

Table of Contents

PREFACE	3
EXECUTIVE SUMMARY	5
Why is this discussion paper needed?	5
What are the issues?	5
Recommendations	6
WHAT IS DATA-BASED RESEARCH?	8
Historical foundations	8
Data-based research today	8
Benefits from data-based research in Australia	9
FUTURE CHALLENGES FOR AUSTRALIA	10
Understanding opportunities and risks	10
Recognising and responding to opportunities	13
Jurisdiction issues	14
PRIVACY MATTERS	16
Safeguards in the current system	16
Personal identifiers	17
Privacy protection strategies	17
Re-identification	18
The scope of data release	19
Data release in practice	19
ADDRESSING PROBLEMS WITH DATA RELEASE	20
Factors which may affect decisions by data custodians	20
Procedural complexities	22
THE WAY AHEAD	23
Table 1. Public benefits from data linkage and data-based research in Australia	25
REFERENCES	28
GLOSSARY AND NOTES	33
ANNEXES	36
Workshop program	36
List of attendees	38
Acknowledgements and thanks	40

PREFACE

Provenance of this discussion paper

The Menzies Foundation was established in 1979 to perpetuate the ideals of Sir Robert Menzies, Australia's longest-serving prime minister. The Foundation has subsequently made its mark on Australian society by providing non-partisan leadership through prestigious Menzies Memorial scholarships, through the outstanding health research carried out by its daughter institutions, and through a series of workshops on matters of national importance.

This paper is based on the proceedings of a Menzies Foundation workshop held on 16th August 2013, attended by more than 50 people, including experts in data-based health research, in ethics, in legal and privacy issues, as well as community representatives, journalists and observers from a number of government agencies.

Objectives of the workshop

1. To inform the community and the media about the public benefits of research based on the sharing and linkage of publicly-funded data, currently held in trust by various government agencies
2. To reassure the community and the media that such research can proceed without risk to privacy
3. To encourage a frank discussion of the reasons for the long delays in approval and release of data for such research projects
4. To seek community support for the idea that publicly-funded data should be more readily available for public interest research projects that are ethically approved and scientifically sound.

Why was this workshop needed?

Australia spends some \$140 billion each year on health, including significant sums on the electronic recording of births and deaths, health services, medications, health outcomes, and other relevant data.

The Australian community can reasonably expect that information collected on its behalf is being well used for public-benefit purposes, including health research. Wisely used, the sharing of public sector information will help to identify previously unsuspected risks to health, improve health and social outcomes for many people, and improve the efficiency and effectiveness of the entire health system.

The value of health research based on public sector information is widely recognised, and is justified by the many important findings from earlier research projects. Unfortunately, it is still

the case that health researchers external to government can have great difficulty in obtaining approvals for access to public sector data for projects that have been ethically approved, and are judged to be of high scientific quality.

This discussion paper explains the rationale for making public sector data more readily available for research purposes, and the measures needed to bring this about: adequate resources, protection of privacy, and streamlining of approvals processes. In the short-term it may also be necessary to reassure government agencies and jurisdictions that they will not be disadvantaged by making their data holdings more available to external researchers.

The Foundation is pleased to have been able to host the workshop, and to support the release of this discussion paper on such an important topic. We would certainly welcome any comments on the matters raised.

John D Mathews
Executive Director (to 23 August, 2013)
Menzies Foundation
mathewsj@unimelb.edu.au

Brian Doyle
Chairman
Menzies Foundation
menzies@vicnet.net.au

Important Note and Disclaimer

The workshop discussions were held under the Chatham House rule, with a guarantee that comments would not be attributed to individuals. The presentations at the Workshop can be viewed via the Menzies Foundation website: www.menziesfoundation.org.au. This discussion paper, based on the workshop proceedings, was prepared by the Foundation and finalised after receiving comments on the draft circulated to all attendees. The content of this document is the responsibility of the Foundation, and is not necessarily endorsed by all or any of the other attendees.

EXECUTIVE SUMMARY

Why is this discussion paper needed?

The public benefits of data-based health research¹ are not being fully realised in Australia.

Many ethically approved research projects of high quality have been unable to proceed because one or more of the relevant government agencies has not provided essential data to researchers.

This impasse is at great cost in terms of knowledge forgone, and in lost productivity of researchers, and of the health sector as a whole.

What are the issues?

Public sector data, collected by government agencies and held in trust on our behalf, is a valuable community resource for research to improve the efficiency and effectiveness of the entire health system and to improve health and social outcomes for all Australians.

Benefits have already flowed from data-based research carried out in accordance with ethical guidelines of the National Health and Medical Research Council (NHMRC) and relevant privacy protection legislation.

Although the value of data-based health research is widely understood by experts, the public benefits will not be fully realised unless urgent action is taken to improve research access to government sector data.

This discussion paper explores the problems with current arrangements and makes recommendations that would maintain privacy protection while making public sector data more readily available for public benefit research.

¹ Although this report emphasises health-related research, similar issues arise in relation to data-based research in other areas of public policy and practice.

Recommendations

Community involvement

1. Community representatives, consumers, professional bodies, researchers and other interested parties should re-affirm their support for data-based research for public benefit purposes, and engage with the media and with politicians to discuss the issues raised in this paper.
2. Mechanisms should be strengthened, in partnership with agencies such as NHMRC, to help community representatives to become more involved in the setting of relevant priorities for data-based research, in the monitoring of processes for data release and privacy protection, and in the assessment of research outcomes and consequential public benefit.

Government responsibilities

3. Government agencies should acknowledge that the research use of their data is central to their core business, and implement a whole-of government approach, with data release policies and practices that are standardised across agencies and jurisdictions.
4. Governments should allocate resources to support core infrastructure for data-based health research, including infrastructure for data linkage and for the secure storage and analysis of de-identified data; such resources would allow agencies to conduct intramural research, to cooperate with external researchers, and to provide data promptly to external researchers for approved projects.

Legislative protection for data linkage

5. Legislation should be amended as necessary to establish clear and transparent procedures to support the release of personal information to recognised data linkage organisations for linkage using privacy-preserving protocols.
6. Legislation should provide legal penalties for breaches of data-privacy and also mandate the reporting of any such breaches.

Release of data

7. Government agencies should be obliged to release data for ethically approved and scientifically valid research unless there are strong countervailing reasons, such as those based on national security.
8. Decisions about release of data should be made in the public interest and in accordance with published criteria. Criteria should not be based on the special interests of particular agencies or jurisdictions.
9. Decisions on the release of data should be made and communicated within a specified time-frame and agencies should be required to provide reasons for any refusal and a mechanism for reviewing decisions.

Approval processes

10. The processes for ethical approval and for approvals of data release should be simplified to minimise duplication of effort by researchers, ethics committees, government agencies and data linkage organisations, and to maximise the sharing and re-use of existing linked data resources. For example:
 - a. Ethical approval from a lead Human Research Ethics Committee specialising in data release and linkage applications should normally be accepted by all parties without the need for additional ethical review.
 - b. A favourable review by NMHRC or another relevant research funding agency should normally suffice as evidence of the scientific quality and relevance of the proposed research.
 - c. A single organisation should coordinate data linkage projects that require cooperation between several government agencies or jurisdictions.
 - d. Linked data-sets, once created and de-identified, should be made available for re-use for new research purposes whenever it is in the public interest to do so, subject to approval from the relevant ethics committees and data-custodians.

The way ahead

11. The issues raised in this paper should be reviewed in further discussions involving consumers and community representatives, privacy authorities, data-holding agencies, researchers and research agencies such as NHMRC.
12. Leadership is required to refine these recommendations and to establish an enhanced Australian commitment to data sharing and data linkage, to consider the possibility of a national privacy plan to facilitate such data-based research, and to seek bipartisan and jurisdictional support for any necessary changes in legislation.
13. Such a national commitment should also ensure:
 - a. More explicit recognition of the public interest
 - b. Greater involvement of consumers and community representatives
 - c. More adequate resourcing
 - d. Greater cooperation between government agencies and external researchers
 - e. Greater transparency of streamlined approvals processes
 - f. Monitoring of data requests, approvals, and releases
 - g. Monitoring for any privacy breaches
 - h. Monitoring of research outcomes.

WHAT IS DATA-BASED RESEARCH?

Historical foundations

14. For many centuries, doctors have learnt about disease by carefully describing individual patients and by looking for things in common between patients with similar symptoms. For example, Percival Pott used this epidemiological² method to show, in 1775, that occupational exposure to soot seemed to explain the high incidence of cancer of the scrotum in London chimney-sweeps.
15. Since the 19th century, in registering deaths, statisticians in the tradition of William Farr have linked cause of death to age and sex, to place of residence and to occupation, thus providing important clues about how diseases such as cholera, tuberculosis, liver cirrhosis, hypertension, asbestosis and mesothelioma are spread or caused, and how they might be prevented.
16. Although the identity of each person has to be known when such data are collected, the statistical tables published by Registrars have always been in aggregated and de-identified form to protect privacy.

Data-based research today

17. The research value of electronic data has been widely recognised in many authoritative reports from government and academia [1-18; W1-W4]³. Many databases of relevance to health are available in electronic form in the government domain, and there is great potential value in using such data for public-benefit research, often after bringing together (linking) information for the same set of individuals from different data sources.
18. Electronic data linkage, with safeguards to protect privacy [12, 14], now makes it possible to answer important health questions which would be impossible if very large numbers of records had to be scanned manually.
19. Data linkage supports privacy because researchers are able to use merged data from different sources from which personal identifiers have been removed [19]. Access to linked data is only allowed after stringent ethical review [12, 14] to ensure that privacy is protected.

² Some of the technical terms used in this report are explained in the Glossary and Notes appended.

³ Numbers in square brackets refer to publications and sources listed in the Reference List; Workshop presentations have a W prefix.

Benefits from data-based research in Australia

20. Many valuable projects (see Table 1 for a selection) have already used data linkage to identify and quantify health outcomes or health risks that were otherwise poorly understood eg:
- a. Birth records for natural and assisted conceptions were linked to the register of major birth defects in Western Australia. Infants conceived as a result of in-vitro or intra-cytoplasmic technologies were twice as likely to suffer major birth defects [20] (Table 1A).
 - b. The increased risk for deep venous thrombosis (blood clots in the leg) following long-haul air travel was assessed by linking international passenger arrivals into Perth airport to subsequent hospital admissions in Western Australia; admission risk for DVT had returned to normal by 3 weeks after arrival [21] (Table 1B).
 - c. National data linkage was a vital strategy in a ground-breaking study (Table 1), linking de-identified Medicare records of CT scan (medical X-ray) exposures to de-identified cancer notifications for the same individuals in the ensuing years. This study showed that with an average follow-up of only 10 years, the average risk of cancer increased by 16% for each additional CT scan in a person under the age of 20 years [22] (Table 1C). These findings have already led to changes in CT policy and practice to reduce future exposures of young people to low-dose ionising radiation from CT scans.
 - d. The risk of birth defects caused by maternal exposure to medications prescribed in early pregnancy was assessed by linking PBS records of medications to WA records of maternity and birth defects [23] (Table 1D). Linkage at the national level, with larger numbers, would have had the potential to identify any other medications, not yet known to cause birth defects if taken during early pregnancy.
 - e. Caesarean section rates in NSW varied between 11.8 and 47.4% of births depending on the hospital [24] (Table 1E); the differences were unexplained, but may reflect social factors as well as differences in obstetric practice between hospitals.
 - f. In the Longitudinal Study of Australian Children, the costs of community-based healthcare during childhood were greatly increased for those who had been low birth-weight and/or pre-term infants [25] (Table 1F), emphasising the need for prevention through maternal education and ante-natal care.
 - g. Linkage of results from the Health in Men study to WA mortality data showed that low vitamin D status was an independent predictor of increased frailty and all-cause mortality in older men [26] (Table 1G); randomised

studies would be needed to finally prove whether supplementary vitamin D will reduce frailty and mortality.

21. Although data linkage is a powerful research tool (see also [27-33]), it is not always necessary to link information from separate sources; other research projects have yielded new knowledge by simply relying on detailed analyses of de-identified data from single data-bases, eg see [34, 35] (Table 1P and 1Q) and [36].
22. In future years, data-based research will deliver much new knowledge to support evidence-based policy and practice, and also deliver additional community benefits through improvements in productivity, particularly in the health sector [W1-W4].

FUTURE CHALLENGES FOR AUSTRALIA

Understanding opportunities and risks

23. Governments in UK, Europe and North America have long recognised the importance of public sector data and linkage [16, 17, 18], and have recently invested in relevant infrastructure. For example, UK has invested over 100 million pounds to support research using electronic health data, including linked data
24. For more than a century, information collected by Australian governments has been used to create de-identified data-sets and tabulations (eg infectious disease notifications, cause of death statistics) which have been accessible for epidemiological research.
25. In recent decades, new public sector data-sets have been created by state and territory governments, often with support from the Australian Government, to support health activities such as cancer registration, screening for breast cancer and cervical cancer, and to monitor hospital activity or maternal mortality or birth defects.
26. With this growth of government business, and with public sector information now routinely stored electronically, there are thus many new opportunities for public benefit research by using single data-sets for new research purposes, and by linking information from separate data-sets for the same individuals [1-16, 37-39].
27. However, such research opportunities can also pose risks: for example, privacy may be threatened if research data are not adequately de-identified or stored securely [12-16, 19, 37-39, W5-W6], or the interests of a government agency or jurisdiction may be affected if administrative data are exposed to external research scrutiny or evaluation.
28. Western Australia has been a national leader in applying data-linkage to advance health knowledge, policy and practice, while involving health consumers actively in

the process and applying stringent measures for privacy protection [3, 4, 5, 8, 15, 19, 21, 27, 28, 32, 40, W5].

29. The Australian Institute of Health and Welfare (AIHW) was established in 1987 [41] to provide an institutional framework for the national compilation of health information and statistics (eg cancer and death registrations, hospital morbidity statistics), mostly collected by the states and territories. AIHW has been involved in data-linkage projects since the mid-1990s, and has been recently accredited as an Integration (data-linkage) Authority for Commonwealth data.
30. Australian Government databases recording fee-for-service transactions under the Medicare Benefits Scheme (MBS), and the Pharmaceutical Benefits Scheme (PBS), were established primarily for administrative purposes. However, these databases also have great research value, as demonstrated in the many linkage studies carried out in WA [eg 4, 8, 15, W2, W5], in some national studies [W10], including one that obtained subject-consent eg [25]⁴, and in at least one national linkage study that was approved without subject-consent [22, W1].
31. The Australian Government established the Australian Childhood Immunisation Register (ACIR) to support the timely vaccination of all Australian children [42], and the National Notifiable Diseases Surveillance System (NNDSS) to enable national coordination of notifications of communicable diseases from states and territories [43].
32. Health agencies in the states and territories have been able to make increasing use of their own data for research purposes. The value of linking different public sector data-sets to address particular research questions has been proven at state level by ground-breaking projects in Western Australia [3, 4, 5, 8, 15, 19, 21, 27, 28, 32, 40, W5], and by more recent projects involving other jurisdictions [eg 29-31, 33, W3, W4, W7].
33. The Population Health Research Network was established in 2009, with funding from Commonwealth, state and territory governments and academic partners, to help coordinate national infrastructure for data-based research and to support state-based data linkage activities. The Commonwealth and all Australian states now have active data linkage units and there is a national facility in Sydney which allows researchers to securely access designated Australian data sets [38, W2]. (<http://www.phrn.org.au/about-us>)
34. The Personally Controlled Electronic Health Record (PCEHR), recently introduced by the Australian Government with patient consent, will eventually allow *identified*

⁴ For example, in the Longitudinal Study of Australian Children [25], involving some thousands of families, it was feasible to seek consent, at the time of recruitment, to link Medicare records to other data. In contrast, in the CT scan study [22], it was not feasible or desirable to seek consent from all 11 millions subjects, so that Medicare records of CT exposures were linked to cancer notifications without obtaining subject consent, but only after approval had been obtained from ethics committees working to NHMRC guidelines.

health records to follow each patient through the health system, with stringent safeguards to protect privacy. In the longer-term, as more people sign-up to the PCEHR, as the quantity and quality of captured data improve, and as the public is reassured about privacy protection, *de-identified* records from the PCEHR could be aggregated to provide a much richer vein of clinical information to identify adverse drug events and other risks, to evaluate diagnostic criteria and treatments, and to improve the efficiency and effectiveness of the health system as a whole. (See <http://www.ehealth.gov.au/internet/ehealth/publishing.nsf/content/home>)

35. Discoveries in molecular science and genetics are posing a new generation of challenges for medical care and health information [44].
 - a. Genetic testing is becoming cheaper and cheaper, although as testing becomes more useful and accepted, it will inevitably add to the cost of health care;
 - b. The increasing availability of genetic information is adding complexity in the areas of ethics and privacy. For example, genetic information for one person also has relevance for close relatives, which raises complex ethical questions about privacy and responsibilities for disclosure to relatives.
 - c. At a more subtle level, genetic testing threatens the very notion of de-identifiability: a single sample of blood, saliva or tissue, tested appropriately, provides so much genetic information that the profile would be unique, unless there was an identical twin.
36. In Australia the private sector makes very effective use of personal information to target and promote products and services to the Australian community. Personal and financial information collected in the course of business by stores, financial and credit institutions, and by on-line services such as Google, is sold, shared and linked to create personal profiles that can be targeted by individual businesses. This state of affairs is accepted by the Australian community, even though it effectively creates a double-standard between the freer access to personal information in the private sector, and the more restricted access to public sector personal information for public benefit research purposes.
37. Access to public records for health research purposes should also be considered in the context of the access provided for other public benefit purposes. For example, political parties are given access to the personal details in electoral records for campaign purposes, whereas in recent years, the Electoral Commission in several jurisdictions has refused access to electoral records for epidemiological research purposes.

Recognising and responding to opportunities

38. Consumer organisations and representatives [6, 33, 39, 46, W9] and members of the public [8, 9, 33, 45, 46, W8] have expressed support for data-based health research, with few reservations.
39. The Prime Minister's Science, Engineering and Innovation Council [9], the Academy of the Social Sciences in Australia [10], The Consumers Health Forum [39], The Population Health Research Network [38], NHMRC [12], the McKeon Strategic Review of Health and Medical Research [11] and others have emphasised the value of data integration, data-linkage and data-based research.
40. The Australian government has recently issued an *Australian Public Service Big Data Strategy* [13] to emphasise the value of research based on the integration and linkage of data held by government, and to shape an enabling framework for future research and evaluation.
41. Government agencies have long had the opportunity to use their publicly-funded administrative data for evaluation purposes, particularly in areas such as health. Unfortunately, the culture of evaluation has not yet been firmly established within all of the bureaucracy, perhaps in part because of the lack of resources and expertise, and partly because of a focus on short-term outcomes, with implicit concerns that rigorous evaluation might reflect upon agency performance.
42. Health consumers in Western Australia have become active participants in the Developmental Pathways project; 160 community representatives are sitting on 25 committees for data linkage projects, and contributing to decisions about research priorities, policies and practices [8, 46].
43. Health consumers have long been concerned about long queues and waiting times and other adverse experiences of patients, and there is a growing need to measure health outcomes and agency performance from the consumer perspective, to measure what the public is paying for, to identify strategies that can be shown to improve performance, and to publicise the success stories as they emerge.
44. As few members of parliament have backgrounds in health or research, there is a continuing need for national leadership to showcase the benefits of data-based health research for the public, for all agencies and for all jurisdictions, and to put the risks in perspective.
 - a. The media can play an important role, through the emerging cohort of data journalists, as well as through web-based and modern media modalities.
 - b. The Parliamentary Library could provide a useful venue for communication with backbenchers, and for discussions to consolidate bipartisan support for data-based health research.

- c. If controversy about data-based research should emerge, then the issues could be usefully referred to a Senate Inquiry.

Jurisdiction issues

- 45. Under Australia's federal system, delivery of publicly-funded hospital services and most public health services is a responsibility of the states and territories. The Australian government funds private medical services via Medicare and pharmaceutical services via the PBS; it also contributes to hospital funding through cost-sharing agreements with states and territories, and provides additional funding for health research and other public health services.
- 46. The disconnects between financial responsibility and service responsibility have led to cost-shifting and "blame-games" between the states and Canberra, with a number of other consequences that make it difficult to establish the research and evaluation culture that could improve the effectiveness and efficiency of the health system:
 - a. Because the role of the Australian government is essentially financial, most of its data collections (eg MBS and PBS) have been designed to ensure financial accountability, rather than to support outcome-based research and evaluation [W1, W10]. For example, MBS provides no diagnostic classifications for primary care consultations. Nevertheless, de-identified MBS records have been used to great effect in a national linkage study that showed an increased risk of cancer following exposure to low-dose radiation from Medicare-funded CT scans [22].
 - b. The Australian government has used funding incentives to obtain diagnostic information on hospital activity and morbidity from the states and territories. As states and territories are not always adequately involved in the analyses of such data at AIHW, their sense is often that data go down a one-way street, to be warehoused in Canberra, rather than being used optimally in the public interest.
- 47. Although all jurisdictions have legislation that enables the sharing of publicly-funded data for research and evaluation purposes, with appropriate measures for privacy protection [12, 14, W5, W6], the release of data for research and evaluation purposes often seems to be resisted, rather than facilitated, by data custodians [W1]. For example, the Department of Health and Ageing had a long-standing policy of not releasing Medicare data for national linkage purposes. More recently, Medicare data has been made available for selected projects (see paragraph 30).
- 48. Access to data is most problematic for national studies requiring data from seven or eight different jurisdictions.
 - a. For example, because of complex arrangements between jurisdictions, national death records may not be available for up to five years after the death has occurred. This is because death certificate information provided

by state and territory registrars goes to the Australian Bureau of Statistics (ABS) to ensure uniform national coding, and the ICD codes then go back to the states to be linked to person identifiers, before being submitted to the national death index at AIHW. State registrars, led by the coordinating registrar in Queensland, are trying to simplify and expedite these processes, to provide more timely access to national death records.

- b. Despite the creation of AIHW to produce national data-bases from data provided by states and territories, access to *unit data* at the national level (eg for cancer incidence) can still require separate approvals from each state or territory. While such approvals are rarely refused when data have already been provided to AIHW, the need to go back to ethics committees and data custodians in every state and territory introduces procedural complexity and costly delays.
 - c. There have also been difficulties in identifying Indigenous Australians in official records, reflecting both past differences in recording practices between jurisdictions, and changes over time in the proportion of people of Indigenous descent who self-identify as such. Difficulties persist within jurisdictions [27, 28, 47, 48, W7] and at the national level [30, 37]. Data linkage projects can help to reduce the ambiguities in Indigenous identification by exploring the consistency between different data-sets, and by testing the assumption that if a person is identified as Indigenous in one or more data sets, this is likely to be correct (48).
49. Some agencies are required to recoup the costs of providing data from their files. For example, for many years, state registrars of births and deaths have charged fixed fees for copies of individual certificates requested by persons who could show a legitimate interest. In the era of manual records, most registrars were also prepared to facilitate epidemiological research through arrangements that charged no more than the marginal cost of searching for larger numbers of records. As most records are now electronic, the marginal costs are small, and the public interest would not be served by asking researchers to pay the full costs of databases.
50. Health agency cultures and values also vary by jurisdiction. State and territory agencies, with service accountability, tend to employ more health professionals, and to be more directly responsive to patient and consumer needs. Commonwealth health agencies see the world through the prisms of information and money, and professional or research expertise can be sidelined, particularly if it is seen as representing special interests. In some jurisdictions or agencies, there is a natural tendency for the prevailing culture to be self-protective and self-perpetuating, to see publicly-funded data as “belonging” to the agency, and to resist transparency and comparisons with other agencies or jurisdictions; this may help to explain the reluctance to release data for research and evaluation purposes.

PRIVACY MATTERS

51. Communities and individuals value privacy and strike a balance between personal needs and goals, and what others need or want to know about them. Privacy underpins human dignity and gives people a measure of control in everyday interactions, and over the use of their personal information in the wider world [14].
52. The 1980s saw increasing concerns about data protection and privacy, perhaps partly because of the sensitivity of personal information about HIV/AIDS, and partly because of concerns that the digitisation of personal information could increase the risks of privacy abuse [14].
53. New legislation and regulatory mechanisms for privacy protection were introduced in the public sector, giving data custodians a discretion to release information for research or other public benefit purposes, provided that privacy was adequately protected [12, 13, 14, 37, W5, W6].
54. The concept of community privacy has also been recognised and protected, most particularly for Indigenous Australians, where identifiable communities may be at risk of being stigmatised [12-14, 37, 47].

Safeguards in the current system

55. In Australia, health research projects require prior ethical approval by a Human Research Ethics Committee (HREC) working to NHMRC guidelines⁵ [12]. NHMRC guidelines are also consistent with the Declaration of Helsinki and the guidelines of the World Medical Association, currently under revision [50].
56. Privacy protection legislation in a number of jurisdictions makes HREC approval a statutory requirement for projects seeking to make use of public sector or personal data.
57. For such data-based projects, the ethical review by the HREC will focus on the adequacy of the proposed measures for protection of personal and community privacy.
58. Furthermore, in reviewing the ethics, the HREC will also make a judgment about the scientific quality of each proposal submitted, and the competence of the investigators to successfully complete the work. Ethical approval will be refused or deferred for weak proposals.
59. Ethical approval by an HREC thus provides an endorsement of both the privacy protection measures and the scientific quality of the proposal and a judgement that the project is ethically justified.

⁵ It is possible that in some jurisdictions, projects internal to government can still proceed without such a formal ethical review.

60. The peer review process for projects that seek competitive funding provides evidence of the scientific quality of projects. Only a small proportion of applicants for competitive funding from NHMRC will actually be successful; nevertheless even for unfunded grants the ranking process provides evidence of the scientific quality and relevance of the research proposed, and some of these projects are able to proceed with alternative sources of funding.

Personal identifiers

61. In 1985, the Hawke government attempted to introduce the Australia Card, a personal numbering system for all Australian residents, with the principal purpose of minimising taxation fraud. The measure was defeated twice in the Senate and never introduced, but it left enduring concerns about privacy and big government.
62. Special legislation was introduced to protect against unauthorised use of the universal Medicare number, and the tax-file number system was introduced to minimise taxation fraud.
63. Electronic data aggregation, using probabilistic linkage techniques, requires access to personal details (usually sex, full names, date of birth, and address). Linkage will often fail if data items are missing or if names or addresses have changed [5,7,37,38,W2].
- a. Government agencies at state level are more easily able to monitor changes in name or address or vital status by accessing marriage or death registrations and other state data-bases [W2];
 - b. The Electoral Register is maintained by the Australian Electoral Commission on behalf of all jurisdictions, but as many younger adults are unregistered, the register would be of limited value for linkage. However, agencies with the necessary approvals can access the Register to update addresses for persons who have moved.
 - c. Australian Government data-bases (eg the Medicare register) are also updated for changes of name and address. The Medicare register is more complete than the electoral register. Linkage for research purposes would be more efficient if it could make use of the Medicare number⁶ or another universal personal identification number, especially for cross-jurisdiction linkage.

Privacy protection strategies

64. Government legislation provides legal penalties for public servants who disclose personal information for unauthorised purposes, or who are otherwise in breach of privacy legislation [W2, W5, W6].
65. Researchers who receive government data for research purposes are required to sign confidentiality agreements which bind them and their organisation to protect

⁶ The Medicare number is not always unique to each individual.

the information provided; legal penalties can apply to any breach. New legislation would be required to mandate uniform national penalties for breaches of data-privacy [W2, W5, W6].

66. De-identification, the most effective strategy for privacy protection, requires not only that names, and addresses and dates of birth are removed or changed in files provided to researchers, but that other data items that might allow re-identification should also be removed [14, 37,38, W2, W5, W6].
67. Because of the need to access personal identifiers for linkage purposes, such access must be restricted to a “trusted” linkage organisation, usually within government. Typically, such linkage can be done without anyone in that organisation having access to any potentially sensitive information about individuals. Once the linkage is complete, personal identifiers are removed and the de-identified file containing the linked information is provided to researchers eg [22].
68. In clinical studies, with sensitive personal information involved, the abstraction or interpretation of medical records by a “trusted” clinician may precede or even be part of the linkage process itself [W5]. In such cases, the file containing the final linked records would be de-identified before being passed to the final research team.

Re-identification

69. In exceptional circumstances it may be necessary to re-identify one or more research subjects for clinical/ethical reasons: for example, if the research findings show that persons with certain characteristics would be likely to benefit from a particular test or treatment, there would be an ethical obligation to alert the responsible clinician and/or the patient directly eg [12, 14, W5]. Accordingly, in clinically oriented projects, the possibility of re-identifying individuals would often be built into the privacy protection plan (eg by retaining a file which could be linked back to the personal information by a trusted agent if there were an ethical requirement for re-identification.) The rationale for such deliberative re-identification of research subjects is less in large-scale data-linkage projects carried out without subject consent, as in [22].
70. A potential privacy concern relates to the re-identification of records for reasons that have no ethical justification. For example, it is theoretically possible to use multiple data items relating to a person to create a unique set of items (eg age over 95, male, living in a small (named) country town) that would serve to identify the person. However, for such a possibility to lead to an actual breach of privacy, it would be necessary for someone in the research team to take deliberate steps to determine the actual identity of the person with those characteristics. Such steps would constitute a breach of trust by the researchers, a breach of the contractual confidentiality undertakings and in some jurisdictions would be in breach of privacy law. There is no instance known of such a breach in data-based research projects in Australia.

71. There is also a risk of re-identification of some deceased individuals through media reports of the circumstances of the death.

The scope of data release

72. For most data-based research projects, the knowledge gain will increase in proportion to the amount of data that is made available:
 - a. In linkage studies, the knowledge gain will tend to increase with the number of data items released for each (de-identified) person. This often means that researchers will ask for as many data items as possible, and that this will be resisted by data custodians on the grounds that if multiple items are released, there is an increased risk of “re-identification” of individuals. Negotiations over this point can lead to long delays in data release.
 - b. It will often be preferable to use data from the whole of Australia, rather than from a single state or territory; in addition, with whole of country data, additional insights may emerge from comparisons between states or territories. However, the benefits of using national data, arising from the larger numbers (extra statistical power), and from contrasts between the states, currently need to be traded off against the administrative burden of obtaining approvals and data from multiple jurisdictions.

Data release in practice

73. Procedures for the protection of privacy have proven to be very effective, and there has been no evidence of any privacy breaches from data-based health research in Australia.
74. The only known breaches of privacy legislation have resulted from data-base access by public servants for an unauthorised purpose – eg to target the personal details of a public figure. Sophisticated procedures are now in place to track each access or transaction, so that such privacy breaches can be more easily detected.
75. The discretion to release government data for research by external researchers is vested in a data custodian within each of the relevant agencies or jurisdictions [14 W2, W5, W6]. Ideally, a data custodian would approve the release of data for any project that was recognised as scientifically valid and relevant through an appropriate advisory committee or funding agency such as NHMRC, and approved by a Human Research Ethics Committee, with appropriate measures to protect privacy [12, 14, W5, W6].
76. Data custodians have the discretion to refuse or defer a decision to release data for external research, usually without an obligation to give reasons. Possible reasons are considered from paragraph 81.

ADDRESSING PROBLEMS WITH DATA RELEASE

77. Unfortunately, despite widespread support for data-based research, many Australian researchers still report lengthy delays in obtaining approvals for access to government data-sets, and many potentially important projects have not been able to progress at all because data has not been provided. This impasse is at great cost to the productivity of the research community and of the health system as a whole.
78. Research projects requiring access to national data (eg Medicare or PBS) or data from multiple agencies or jurisdiction are more likely to be delayed or refused, in part because of the complexity of current approvals processes, and also because of the overly cautious approach of data-custodians within individual agencies and jurisdictions [W1, W10]. For example, with a project requiring access to de-identified Medicare data and national cancer and death registrations, it was still necessary for the researcher to seek and obtain as many as 20 separate approvals from data custodians and ethics committees acting for DoHA and Human Services, for AIHW, for the eight state and territory jurisdictions, and for the HREC of the home institution [22]. The approval from DoHA alone required several years of negotiation – most researchers can't afford to give up that much time to a single project!
79. The public interest position should be for all parties to welcome and facilitate data-based research that might inform policy and practice eg Table 1, [25, 35]. This implies that the default position should be for release of data for research unless there are strong countervailing reasons such as national security, a failure to meet ethical standards or that the research question is judged to be of insufficient quality or relevance by research-funding agencies such as NHMRC.
80. Decisions on data release by data custodians are not transparent, and reasons for refusing or deferring data release are not routinely given.

Factors which may affect decisions by data custodians

81. *Resource limitations and perceived priority:* Policies or practices that restrict or delay data release or linkage may reflect limited government resources or a view that specific research projects have limited relevance or priority. If government agencies were provided with more resources to support data-based research, this would be more than justified by the productivity gains in the longer term [49]. Questions of relevance and priority would be best addressed by requiring that individual projects not only have ethical approval, but that they have been judged to be of high scientific quality by NHMRC, ARC, or a relevant institutional body.
82. *Concerns about privacy* may be offered to justify the non-approval of projects requiring data release. However, privacy and related legislation in all jurisdictions supports research and recognises the need to consider privacy interests together

with the public interest in research. There is no instance of any privacy breach resulting from data-based research in Australia that would justify delays by data custodians in approving the release of data to researchers. Data custodians and ethics committees also protect privacy by minimising the amount of information provided and by imposing strict conditions on researchers [14, W1, W2, W5, W6].

83. *National projects in a federal system:* Projects that aspire to national linkage usually need data from multiple jurisdictions, thereby running the risk that one or other jurisdiction will not agree, or will delay approvals unnecessarily. Fortunately, PHRN with the support of AIHW and Australian jurisdictions, is developing and operating a cooperative infrastructure to facilitate data sharing and linkage within and between jurisdictions [38, W2].
84. *Potential implications of research:* Some in government may be cautious about providing data for external research because they believe that the results may reflect adversely on policy or practice in their jurisdiction or agency or otherwise threaten the policy agenda [W1]. Governments, hospitals and service providers may be reluctant to provide data for research purposes if there is a risk of litigation about health risks identified by the research.
85. *Concerns about the scope of research projects:* Researchers may seek to address 'open-ended' questions using datasets that include numerous data-items. Governments, through their data custodians, give preference to research that is focussed on a very specific question that requires only a limited number of data-items. Data custodians give force to this by resisting the release of numerous data-items. Lengthy negotiations between researchers and data custodians may ensue.
 - a. Over the longer-term, it can be argued that there could be greater gains in new knowledge and productivity by not restricting data release to the minimum number of items, simply because if a new or implicit question arises in the course of an initial research project (as frequently occurs), it is very wasteful to have to repeat the whole process of data release and extraction to get extra variables.
 - b. On the other hand, the public interest in the release of data cannot be adequately justified if the use of the data variables cannot be explained or assessed in advance.
86. *Concerns about enduring datasets:* If there is a public interest in establishing an enduring dataset of variables brought together from a number of different datasets then this should be established explicitly, with appropriate ethical approvals and managed in the public interest. Careful governance and management arrangements are required to ensure proper protection, fair access and the ethical use of enduring data sets of this kind [W2, W5]. In situations where researchers wish to use project data to examine a new question which was outside the scope of the original approvals is sufficient to obtain agreement for the new research question in a revised proposal.

87. *Concerns about the use of data for unauthorised purposes:* Released data must be used only for the purposes approved in the project proposal, and a new approval must be obtained before any data are used for a new purpose. However, in some circumstances there may be ambiguity about research questions stated in the proposal. The scope for what might be seen as “unauthorised use” is obviously greater when numerous data items are released, which helps to explain the practice of releasing the minimum number of items (see also paragraph 85). Unauthorised use will attract criminal, civil and professional penalties and these issues can also be addressed through contractual arrangements between data custodians and researchers.
88. *Legal complexity:* Legislative arrangements are complex because each data collection can have a unique act which covers the collection, use and disclosure of information as well as the processes for release of data and the penalties for misuse. There are significant differences between the acts for different agencies and jurisdictions, and data custodians need to have a good knowledge of the legislation underpinning their collections [W5, W6]. Legislation could be amended to clarify the powers and processes for release of data, most particularly for the release of data identifiers to recognised data-linkage units.
89. *Concerns about public reactions to data linkage:* Governments may also worry that members of the public may object to the use of data-linkage, regardless of privacy considerations. On the contrary, evidence from the Consumers Health Forum and others suggests that most members of the public can see the value of data-linkage when it is explained to them, and support its use for public-benefit purposes [4-6, 8, 9, 33, 45, 46, W8, W9].
90. *Government commitment:* Government agencies are becoming more used to the idea that public-benefit data linkage projects should be welcomed, rather than feared, and that any perceived risks to privacy or to agencies can be managed by building more cooperative arrangements with the research sector [W2, W3, W4]. There is also a growing recognition of the benefits to governments and agencies from the research that is undertaken.

Procedural complexities

91. The efficiency of data-based research is greatly reduced by factors such as:
- a. *Complexity of approval processes:* For national linkage projects, it is necessary to obtain approvals from multiple data custodians and ethics committees acting for (several) Commonwealth agencies, all states and territories, and the institutional ethics committees for each of the researchers. For the CT scan study, some 20 separate approvals were necessary, and it took almost five years for the Commonwealth to agree to the release of national Medicare data, which then enabled the flow-on of

most other approvals [22, W1, W2].
(<http://www.bmj.com/cgi/doi/10.1136/bmj.f2360>)

- b. *Data custodian control*: Decision-making about access is vested in a data-custodian, who may defer a favourable decision indefinitely if there are unresolved concerns about the policy position at more senior levels, about resource implications, about privacy or the priority to be attached to the project, or about the number of data items to be released.
- c. *Lengthy negotiations between data custodians and researchers*: Because of the tensions between data custodians, who are concerned to protect privacy, and researchers committed to advance knowledge, there may be time-consuming negotiations that add to the delay, and sometimes lead to projects being abandoned.

THE WAY AHEAD

92. An enhanced Australian commitment to data sharing and data linkage would ideally be embedded in more progressive national policies, with bipartisan and jurisdictional endorsement. Such a commitment would involve:
- a. *Public consultation and involvement*: Health consumers, privacy advocates and other interested persons should become more involved in discussion of policy frameworks, in the setting of relevant research priorities, and in the monitoring of data release practices and privacy protection [4-9, 46, W8, W9].
 - b. *More explicit recognition of the public interest in research using available data*: Members of the public would expect that, subject to stringent ethical and privacy safeguards, publicly-funded information should always be available for public benefit research purposes unless there are powerful countervailing arguments such as those based on national security. This may require legislative changes, perhaps with an appeals process if data access is unreasonably denied by a government agency.
 - c. *More adequate resourcing*, so that government agencies can improve data systems, develop the necessary procedures for privacy protection, carry out their own research and also be able to release government data promptly for approved projects of external researchers.
 - d. *Increased cooperation between government agencies and the research community* on data-linkage projects so that government agencies become more involved with external partners in projects likely to inform policy and practice. The WA Development Pathways Project provides an excellent example of how Treasury and human services agencies including health,

education and police have developed cooperative projects with researchers and health consumers [8, 46, W2].

- e. *Streamlining of project approvals processes* would bring great productivity benefits both for researchers and data custodians [49]. For example, with projects involving unit data from multiple jurisdictions or health services, it is sometimes necessary to obtain more than 20 separate approvals. If the ethical and privacy decisions were devolved to a single decision-making authority, and if the decision were favourable, the expectation would be that data would be released unless one or more of the relevant data-custodians could show cause why the release should not be made. In late 2013, NHMRC will be meeting with government agencies, researchers, AIHW, PHRN, other research funding agencies, and representatives of the public to discuss, amongst other things, how approvals processes could be facilitated.
- f. *Greater transparency*, so that the public, the media and the research community are well informed about projects pending approval, projects in progress, and about the results of completed projects [8, 39, 46, W1, W2, W8, W9].
- g. *Accountability in data access decisions* to build trust between data custodians, the research community and the public.

Privacy by design

- 93. Australian measures for privacy protection are working well, albeit with minor inconsistencies in legislation between jurisdictions [W5, W6]. There could be value in developing a national privacy plan (*Privacy by Design*) to facilitate data-based research for the future benefit of Australians. Such a plan would identify the potential risks to privacy from data-based research, outline the range of accepted measures for privacy protection and for monitoring of compliance, and seek legislative consistency between jurisdictions.
- 94. Such a privacy plan would provide additional guidance for researchers in understanding the flow of personal information in their projects and the legal and ethical ramifications for the privacy of individuals. Researchers would thus be better able to develop and describe their plans to address and manage privacy issues. A national privacy plan would also help to support national consistency in the processes for ethical review and for approvals of data release.

Table 1. Public benefits from data linkage and data-based research in Australia

Study	Data linked	Results	Comments
A. Hansen et al. (2002) <i>The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization</i> [20]	Birth records (natural or assisted conception) linked to major birth defects	Infants conceived as a result of in-vitro or intracytoplasmic technologies were twice as likely to suffer major birth defects	Methodological problems in previous studies meant that few others had identified this increased risk
B. Kelman et al. (2003) <i>Deep vein thrombosis and air travel: record linkage study</i> [21]	Long-haul airline arrivals into Perth, linked to later hospital admissions in WA for deep venous thrombosis	Increased risk of deep venous thrombosis had returned to normal by 3 weeks after arrival	Provided definitive data about the magnitude and duration of DVT risk following long-haul air travel
C. Mathews et al. (2013) <i>Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians</i> [22]	Medicare records of CT scan exposures were linked to the Australian Cancer database	Cancer risk was increased by 16% per CT following exposures at younger ages	Confirmed previous theoretical predictions, and led to a review of CT policies and practices to reduce future radiation exposures.
D. Colvin et al. (2010) <i>Linking a pharmaceutical claims database with a birth defects registry to investigate birth defect rates of suspected teratogens</i> [23]	PBS drug prescriptions for WA women in early pregnancy linked to WA birth defects register	Identified drugs which caused birth defects when taken by pregnant women	Additional drug risks would likely be identified if national PBS data could be linked to a national registry of birth defects.
E. Lee et al. (2013) <i>Unexplained variation in hospital caesarean section rates</i> [24]	Hospital birth records linked to caesarean section(CS) records in NSW	CS rate among hospitals ranged from 11.8% to 47.4%.	A better understanding of reasons behind heterogeneity in CS rates may lead to improved maternity care.
F. Westrupp et al. (2013) <i>Community-based health-care costs for children born low birthweight, preterm and/or small for gestational age: data from the Longitudinal Study of Australian Children</i> [25]	Birthweight records (Longitudinal Study of Australian Children) linked to Medicare	An additional 32m was spent per year for children in the higher risk (lower birthweight) categories (aged 0-9 years)	Perinatal risk is a major public health issue; implementation of practices to reduce risk should be a priority
G. Wong et al. (2013) <i>Low vitamin D status is an independent predictor of increased frailty and all-cause mortality in older men: the Health In Men Study</i> [26]	Detailed health data (Health in Men Study - HIMS) linked to mortality data (Western Australian Data Linkage System - WADLS)	Low vitamin D level is associated with and predictive of frailty, and predictive of death independent of frailty	Randomised clinical trials could assess the potential of vitamin D supplements in older men
H. Moore et al. (2013) <i>Diverging Trends in Gastroenteritis Hospitalizations during Two Decades in Western Australian Aboriginal and Non-Aboriginal Children</i> [27]	Aboriginal and non-Aboriginal birth records in WA linked to gastroenteritis-coded hospitalisations	Rates were higher in Aboriginal children compared to non-Aboriginal children, but the disparity declined over time	Ethnicity, age, climate and seasonality are important considerations in evaluating programs of rotavirus vaccination

I.	Watkins et al. (2013) <i>Aboriginal and non-Aboriginal sexually transmitted infections and blood borne virus notification rates in Western Australia: using linked data to improve estimates</i> [28]	STI and blood borne virus (BBV) notifications linked to administrative health data collections (for Indigenous status)	Proportion of notifications with unknown Indigenous status decreased after linkage of data (10.2 to 2.7%)	Accuracy in estimates of STIs and BBVs by Aboriginality may be improved by routine usage of existing linkage resources
J.	Jorm LR et al. (2012) <i>Smoking and use of primary care services: findings from a population-based cohort study linked with administrative claims data</i> [29]	Questionnaire data (45 and up study) linked to Medicare data for 12mths following study entry	Identified differential patterns in usage of health services between current smokers, recent quitters and never-smokers	Results suggest current smokers are less likely to seek health care, and may miss out on beneficial preventative health services
K.	Randall et al. (2013) <i>Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia</i> [30]	Hospital morbidity data linked to mortality data	Aboriginal Australians are 37% less likely than other Australians to have revascularization procedures after heart attacks	This difference is explained by greater likelihood of admission to smaller rural hospitals, higher rates of comorbidity, and lower rates of private health insurance
L.	Hobday et al. (2013) <i>Potential for the Australian and New Zealand paediatric intensive care registry to enhance acute flaccid paralysis surveillance in Australia: a data-linkage study</i> [31]	Non-Polio acute flaccid paralysis (AFP) cases in national AFP dataset linked to acute presentations in ANZPIC registry	Identified 10 ANZPIC cases that were not notified to AFP surveillance system	Need to increase awareness of the AFP surveillance program in ICUs; ANZPIC could identify additional AFP cases
M.	Colvin et al. (2013) <i>Are women with major depression in pregnancy identifiable in population health data?</i> [32]	Delivery records linked to hospital morbidity & PBS records during pregnancy	7495 pregnant women were identified using either data set. 899 were found in both groups	80% of women dispensed an antidepressant were not identified via admission data
N.	Berry et al. (2013) <i>Parent perspectives on consent for the linkage of data to evaluate vaccine safety: a randomised trial of opt-in and opt-out consent</i> [33]	Immunisations to hospital records (analysis of consent preferences in relation to linkage)	Majority of parents supported data linkage for evaluating vaccine safety, and trusted in measures of privacy protection	Most subjects indicated a preference for opt-out/no consent, there was a limited understanding of data linkage
O.	Condon et al. (from 2013) <i>Cervical screening participation and outcomes for Indigenous Australian women. Work in progress - NHMRC-funded project commencing in 2013</i>	Screening records linked to Australian Cancer Database/Hospital admission data	N/A – work in progress	Will help to clarify indigenous identification and why mortality from cervical cancer is high in Indigenous women
P.	Muscatello et al. (2005) <i>An automated, broad-based, near real-time public health surveillance system using presentations to hospital Emergency Departments in New South Wales, Australia</i> [34]	Hospital admission data collected from Sydney Emergency Departments during 2003 Rugby World Cup	The surveillance system had the capacity to identify admission trends and public health threats, although no major threats were identified during this period	Demonstrated the potential of systems which use routinely collected data, and are highly automated, timely, and broad in scope
Q.	Howell et al. (2009) <i>Using routine inpatient data to identify patients at risk of hospital readmission</i> [35]	Hospital admissions and re-admissions	Algorithm used to predict readmission performed moderately, with a high percentage of false negatives	Accurate identification of patients at risk of re-admission could lead to better case-management and lower costs of hospital services

REFERENCES

Workshop Papers

- W1. Mathews, J. Public support of data linkage for better health. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W2. Smith, M. Data linkage: project approval and data access. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W3. Jorm, L. Expanding opportunities for data linkage. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W4. Roder, D. Data Linkage Opportunities and Requirements in Australia – NHMRC and Personal Perspectives. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W5. Allen, J. Public support of data linkage for better health: Legal and ethical issues. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W6. Thomson, C. Ethical review & approval of data linkage research projects. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W7. Whop, L. National Indigenous Cervical Screening Project: a data linkage study. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W8. Xafis, V. Community attitudes to data linkage research. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W9. Bennett, C. Data Linkage for Better Health: A Consumer Perspective. Presented at the Menzies Data Linkage Workshop, 16th August 2013.
- W10. Pearson, S. Pharmaceutical Benefits Scheme. Presented at the Menzies Data Linkage Workshop, 16th August 2013.

Link to workshop papers

The papers presented at the workshop can be found on the Menzies Foundation web-site at <http://www.menziesfoundation.org.au/health/health.html>

References

1. Dunn HL. Record Linkage*. Am J Public Health. 1946; **36**(12): 1412-6.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1624512/>
2. Acheson ED. Oxford Record Linkage Study: A Central File of Morbidity and Mortality Records for a Pilot Population. Br J Prev Soc Med. 1964; **18**: 8-13.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1058929/>
3. Hobbs MST, McCall MG. Health Statistics and Record Linkage in Australia. J Chronic Dis. 1970; **23**(5-6): 375-381.
<http://www.sciencedirect.com/science/article/pii/0021968170900202>
4. Stanley FJ, Croft ML, Gibbins J, Read AW. A population database for maternal and child health research in Western Australia using record linkage. Paediatr Perinat Epidemiol. 1994; **8**(4): 433-47.
<http://www.ncbi.nlm.nih.gov/pubmed/7870627>
5. Kelman CW, Bass AJ, Holman CD. Research use of linked health data-a best practice protocol. Aust N Z J Public Health. 2002; **26**(3): 251-5.

<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-842X.2002.tb00682.x/abstract>

6. Hetzel D. Data linkage research - can we reap benefits for society without compromising public confidence? *The Australian Health Consumer*. 2005; (2):27-28.
<https://www.chf.org.au/australian-health-consumer.php>
7. Holman CD, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, et al. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Australian health review : a publication of the Australian Hospital Association*. 2008; **32**(4): 766-77.
<http://www.publish.csiro.au/paper/AH080766.htm>
8. Stanley FJ, Glauert R, McKenzie A, O'Donnell M. Can joined-up data lead to joined-up thinking? The Western Australian Developmental Pathways Project. *Healthc Policy*. 2011; 6(Special Issue): 63-73.
<http://www.childhealthresearch.org.au/our-research/published-research/2011/impacts-on-Annals-of-epidemiologypolicy-and-practice/can-joined-up-data-lead-to-joined-up-thinking-the-western-australian-developmental-pathways-project.aspx>
9. Working Group on Data for Science. From data to wisdom: pathways to successful data management for Australian science; Report to PMSEIC December 2006.
<http://www.innovation.gov.au/science/PMSEIC/Documents/DataForScience.pdf>
10. Academy of the Social Science in Australia. Facilitating access to routine data for research benefiting the Australian people; Submission of recommendations by the Academy of the Social Sciences in Australia to the Minister for Health and Ageing; 2013.
<http://www.assa.edu.au/programs/policy/submissions/2013-06-20>
11. Department of Health and Ageing. Strategic review of health and medical research; Final report Feb 2013. Canberra, ACT; 2013.
<http://www.mckeonreview.org.au/>
12. National Health and Medical Research Council. National Statement on Ethical Conduct in Human Research 2007 (Updated May 2013): Chapter 3.2: Databanks. 2013 [cited Sep 2, 2013]; Available from: <http://www.nhmrc.gov.au/book/chapter-3-2-databanks>
13. Department of Finance and Deregulation, AGIMO. The Australian Public Service Big Data Strategy: Improved Understanding through enhanced data-analytics capability. Commonwealth of Australia, 2013.
<http://agict.gov.au/big-data>
14. Australian Law Reform Commission. For Your Information: Australian Privacy Law and Practice (ALRC Report 108); Chapter 66 - Research: Databases and Data Linkage. 2008 [cited Sep 2, 2013]; Available from: <http://www.alrc.gov.au/publications/66.%20Research%3A%20Databases%20and%20Data%20Linkage/establishing-databases>
15. Brook EL, Rosman DL, Holman CD. Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System. *Aust N Z J Public Health*. 2008; **32**(1): 19-23.
<http://onlinelibrary.wiley.com/doi/10.1111/azph.2008.32.issue-1/issuetoc>
16. Redman S, Wells RW. Better information to drive health investment decisions. *Med J Aust*. 2013; **199**(4): 235.
<https://www.mja.com.au/journal/2013/199/4/better-information-drive-health-investment-decisions>
17. International Health Data Linkage Network. 2013 [cited Oct 9, 2013]; Available from: <http://www.ihdln.org/>
18. OECD. Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges: OECD Health Policy Studies, OECD Publishing; 2013.
http://www.oecd-ilibrary.org/social-issues-migration-health/strengthening-health-information-infrastructure-for-health-care-quality-governance_9789264193505-en

19. Trutwein B, Holman CD, Rosman DL. Health data linkage conserves privacy in a research-rich environment. *Ann Epidemiol*. 2006; **16**(4): 279-80.
<http://www.ncbi.nlm.nih.gov/pubmed/16257227>
20. Hansen M, Kurinczuk JJ, Bower C, Webb S. The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *N Engl J Med*. 2002; **346**(10): 725-30.
<http://www.nejm.org/doi/full/10.1056/NEJMoa010035>
21. Kelman CW, Kortt MA, Becker NG, Li Z, Mathews JD, Guest CS, et al. Deep vein thrombosis and air travel: record linkage study. *British Medical Journal*. 2003; **327**(7423): 1072-5.
<http://www.bmj.com/content/327/7423/1072>
22. Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, et al. Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *British Medical Journal*. 2013; **346**: f2360.
<http://www.bmj.com/content/346/bmj.f2360>
23. Colvin L, Slack-Smith L, Stanley FJ, Bower C. Linking a pharmaceutical claims database with a birth defects registry to investigate birth defect rates of suspected teratogens. *Pharmacoepidemiol Drug Saf*. 2010; **19**(11): 1137-50.
<http://www.ncbi.nlm.nih.gov/pubmed/20602344>
24. Lee YY, Roberts CL, Patterson JA, Simpson JM, Nicholl MC, Morris JM, et al. Unexplained variation in hospital caesarean section rates. *Med J Aust*. 2013; **199**(5): 348-53.
<https://www.mja.com.au/journal/2013/199/5/unexplained-variation-hospital-caesarean-section-rates>
25. Westrupp EM, Lucas N, Mensah FK, Gold L, Wake M, Nicholson JM. Community-based healthcare costs for children born low birthweight, preterm and/or small for gestational age: data from the Longitudinal Study of Australian Children. *Child Care Health Dev*. 2013; Mar 5 [Epub ahead of print]
<http://www.ncbi.nlm.nih.gov/pubmed/23461342>
26. Wong YY, McCaul KA, Yeap BB, Hankey GJ, Flicker L. Low vitamin D status is an independent predictor of increased frailty and all-cause mortality in older men: the Health In Men Study. *J Clin Endocrinol Metab*. 2013; **98**(9): 3821-8.
<http://jcem.endojournals.org/content/early/2013/06/19/jc.2013-1702.abstract>
27. Moore HC, Manoharan KR, Lim FJ, Shellam G, Lehmann D. Diverging Trends in Gastroenteritis Hospitalizations during Two Decades in Western Australian Aboriginal and Non-Aboriginal Children. *Pediatr Infect Dis J*. 2013; Aug 16. [Epub ahead of print]
<http://www.ncbi.nlm.nih.gov/pubmed/23958809>
28. Watkins RE, Mak DB, Giele CM, Clews S. Aboriginal and non-Aboriginal sexually transmitted infections and blood borne virus notification rates in Western Australia: using linked data to improve estimates. *BMC Public Health*. 2013; **13**: 404.
<http://www.biomedcentral.com/1471-2458/13/404>
29. Jorm LR, Shepherd LC, Rogers KD, Blyth FM. Smoking and use of primary care services: findings from a population-based cohort study linked with administrative claims data. *BMC Health Serv Res*. 2012; **12**: 263.
<http://www.biomedcentral.com/1472-6963/12/263>
30. Randall DA, Jorm LR, Lujic S, O'Loughlin AJ, Eades SJ, Leyland AH. Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia. *Circulation*. 2013; **127**(7): 811-9.
<http://www.ncbi.nlm.nih.gov/pubmed/23319820>
31. Hobday LK, Thorley BR, Alexander J, Aitken T, Massey PD, Cretikos M, et al. Potential for the Australian and New Zealand paediatric intensive care registry to enhance acute flaccid paralysis surveillance in Australia: a data-linkage study. *BMC Infect Dis*. 2013; **13**(1): 384.
<http://www.biomedcentral.com/1471-2334/13/384>

32. Colvin L, Slack-Smith L, Stanley FJ, Bower C. Are women with major depression in pregnancy identifiable in population health data? *BMC Pregnancy Childbirth*. 2013; **13**: 63.
<http://www.biomedcentral.com/1471-2393/13/63>
33. Berry JG, Ryan P, Duszynski KM, Braunack-Mayer AJ, Carlson J, Xafis V, et al. Parent perspectives on consent for the linkage of data to evaluate vaccine safety: a randomised trial of opt-in and opt-out consent. *Clin Trials*. 2013; **10**(3): 483-94.
<http://www.ncbi.nlm.nih.gov/pubmed/23568940>
34. Muscatello DJ, Churches T, Kaldor J, Zheng W, Chiu C, Correll P, et al. An automated, broad-based, near real-time public health surveillance system using presentations to hospital Emergency Departments in New South Wales, Australia. *BMC Public Health*. 2005; **5**: 141.
<http://www.biomedcentral.com/1471-2458/5/141>
35. Howell S, Coory M, Martin J, Duckett S. Using routine inpatient data to identify patients at risk of hospital readmission. *BMC Health Serv Res*. 2009; **9**(96).
<http://www.biomedcentral.com/1472-6963/9/96>
36. Podger A. Putting Citizens First: A Priority That Needs to be Addressed with Care. *Aust J Publ Admin*. 2012; **71**(1): 85-90.
<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8500.2012.00756.x/abstract>
37. Australian Institute of Health and Welfare and Australian Bureau of Statistics. National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People; AIHW Cat. no. IHW 74. Canberra, ACT; 2012.
<http://www.aihw.gov.au/publication-detail/?id=10737422216>
38. Population Health Research Network. Response to the NHMRC Public Consultation on Proposed revisions to the National Statement on Ethical Conduct in Human Research 2007, Chapters 3.4, 3.6, and 3.2. 2012 [cited Sep 2, 2013]; Available from:
http://consultations.nhmrc.gov.au/public_consultations/submissions/NSECHR/1744
39. The Consumers' Health Forum of Australia. Toward the best use of personal health information - a consumer perspective; 1998.
<https://www.chf.org.au/reports.php>
40. Jablensky AV, Morgan V, Zubrick SR, Bower C, Yellachich LA. Pregnancy, delivery, and neonatal complications in a population cohort of women with schizophrenia and major affective disorders. *Am J Psychiatry*. 2005; **162**(1): 79-91.
<http://ajp.psychiatryonline.org/article.aspx?articleID=177275>
41. Australian Institute of Health and Welfare Act 1987.
<http://www.comlaw.gov.au/Series/C2004A03450>
42. Australian Childhood Immunisation Register
<http://www.humanservices.gov.au/customer/services/medicare/australian-childhood-immunisation-register>
43. National Notifiable Diseases Surveillance System
<http://www.health.gov.au/internet/main/publishing.nsf/Content/cda-surveil-nndss-nndssintro.htm>
44. Australian Law Reform Commission. Essentially Yours: The Protection of Human Genetic Information in Australia (ALRC Report 96); 2003.
<http://www.alrc.gov.au/publications/report-96>
45. Wellcome Trust. Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data. London; 2013.
<http://www.wellcome.ac.uk/About-us/Publications/Reports/Public-engagement/WTP053206.htm>
46. McKenzie, A. Personal Communication. Menzies Data Linkage Workshop, 16th August 2013.
47. Onemda VicHealth Koori Health Unit (University of Melbourne), Victorian Aboriginal Community Controlled Health Organisation (VACCHO). Victorian Aboriginal Child Mortality Study. 2013

[cited Oct 9, 2013]; Available from: <http://www.healthinfonet.ecu.edu.au/key-resources/programs-projects?pid=1494>

48. Randall DA, Lujic S, Leyland AH, Jorm LR. Statistical methods to enhance reporting of Aboriginal Australians in routine hospital records using data linkage affect estimates of health disparities. *Aust N Z J Public Health*. 2013; **37**(5): 442-9.
<http://www.ncbi.nlm.nih.gov/pubmed/24090327>
49. McLachlan R, Gilfillan G, Gordon J. Deep and Persistent Disadvantage in Australia; Productivity Commission Staff Working Paper. Canberra, ACT; 2013.
<http://www.pc.gov.au/research/staff-working/deep-persistent-disadvantage>
50. World Medical Association, WMA Declaration of Helsinki Working Group. World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects; Draft revised text for public consultation, 15 April-15 June 2013; 2013.
<http://www.wma.net/en/20activities/10ethics/10helsinki/15publicconsult/>

GLOSSARY AND NOTES

Aggregation, integration or linkage of data records - the terms are virtually synonymous.

Today, many data bases of relevance to health, most often in the government domain, are available in electronic form, and there is great potential value in using such data for public-benefit research, often after bringing together (linking) information for the same set of individuals from different data sources. Electronic data linkage now makes it possible to answer health questions which would be impossible if very large numbers of records had to be scanned manually. Data linkage supports privacy because researchers are able to use merged data from different sources from which personal identifiers have been removed. Such data linkage is only allowed after stringent ethical review to ensure that privacy is protected.

Data custodian

Under most acts governing data release for research purposes, an official of the relevant agency has a discretionary power to release official data for approved research purposes. Although data custodians act in good faith, and in accordance with their legislation, it sometimes appears that the discretion is exercised to restrict or delay the release of data, or to unnecessarily restrict the number of data items to be released for research projects that are scientifically relevant and ethically approved. Delays are even more problematic for data linkage projects involving data held by several different agencies or jurisdictions as this requires the favourable exercise of their discretion by a number of different data custodians.

The public interest would be better served if the default position was an obligation on data custodians to release data unless there were public interest reasons for not doing so.

Data linkage refers to the task of finding records or data items that refer to the same person or entity in different data sources, data files or databases and bringing those records together in a linked, matched or aggregated file. Data linkage requires personal identifiers, or a unique personal number that is common to the data-bases to be linked. Personal identifiers are removed to protect privacy after the linkage has been completed. Data Linkage is called Record Linkage in many jurisdictions, but is the same process.

Data linkage for Australia - WA beginnings: The potential value of data linkage (also known as record linkage) was recognised by many pioneers, and there were developments in Canada even before the advent of electronic computing. The Oxford Record Linkage Study, established by Donald Acheson in 1962, was positioned as a pilot study of what would be possible with the availability of digitised records and increasingly powerful computing.

Michael Hobbs, one of Acheson's first post-graduate students in Oxford, had the imagination to transplant the ideas and capacity for record linkage back to Perth in WA in the late 1960s. In subsequent decades, through the leadership of Hobbs, Michael McCall, Bruce Armstrong, Fiona Stanley, D'Arcy Holman and many others, Western Australia became the epicentre for record linkage in Australia; it was the first jurisdiction to establish a birth defects registry and it broke new ground in linking data from different government and hospital collections, and in linking WA with Commonwealth data.

Data linkage – complexities for Australia: The success of earlier Australian linkage projects in WA and elsewhere, which have been well accepted by the community, has provided an important precedent for all jurisdictions to make greater use of data linkage, particularly in researching health problems. The case for sharing and linkage of data for research in related areas (such as health economics, evaluation, social or family research) has also been well argued in a separate document [10].

National data linkage has presented particular challenges, as many data sets (eg birth, death and cancer registrations, hospital statistics) are the responsibility of the states and territories, while others (eg Medicare, PBS, census, cause of death coding) belong to the Commonwealth. A major advance was the establishment of the Australian Institute of Health and Welfare in 1987 to collate data from all jurisdictions, and to produce national health statistics and reports. For example, AIHW oversees and reports on the National Death Index and the Australian Cancer Database. However, if researchers need linked records based on unit data from either AIHW collection, they must still obtain separate approval from data custodians in each of the states and territories.

De-identification and linkage: The complexities can be illustrated by the steps taken in linking Medicare records of CT scan exposures to the national cancer notifications held by AIHW [22]. In the first instance, Medicare records were tagged with a unique PIN number (not the Medicare number) for each person, and two daughter files were created inside government by authorised personnel.

One daughter file was stripped of the sensitive data items relating to medical consultations and then passed to the “trusted” linkage unit at AIHW, which used personal identifiers (names, DOB, sex, address) to look for matches in the national cancer data base. When a match was found, the cancer details were tagged with the PIN, and the identification details were removed.

The second daughter file was stripped of personal identifiers and passed to the research team. The research team also received the de-identified file of cancer matches, and then linked them to de-identified records of Medicare services (including CT services) using the PIN.

Epidemiology - understanding health and disease

The discipline of epidemiology (*epi* = upon, *demos* = the people) studies diseases at the population level, particularly outbreaks of infectious or communicable disease (epidemics). The epidemiological method has also been used to discover what environmental causes, behaviours or measured risk factors (eg cholesterol level) can predict the incidence of non-infectious diseases. To finally prove causality, it may be necessary to show that an intervention to reduce or remove a putative cause will prevent or delay the onset of disease, ideally in a clinical trial where the intervention is allocated at random to half the persons at risk, if it is ethical to do so. For many situations (eg smoking and lung cancer, or asbestos and mesothelioma) such an intervention would have been unethical and also impractical, because of the long delay between exposure and cancer outcome. Those causal connections were quickly accepted because of the strong associations between exposure and outcome, and biological plausibility. In other circumstances, it can be misleading to make inferences from associations alone.

Population Health Research Network (PHRN): PHRN was established in 2009, with funding from Commonwealth, state and territory governments and academic partners, to help coordinate national data linkage infrastructure and to support state-based data linkage activities. (<http://www.phrn.org.au/about-us>)

Privacy protection

Data linkage improves privacy because it reduces the need for researchers to use identified data. Although personal names and other identification details are required to establish linkage in the first place, they are scientifically irrelevant, and identifying details are routinely removed from the final linked data sets used by researchers.

All research projects depending on data access or data linkage need prior ethical approval, with each project subject to a contract specifying the security arrangements and the methods that must be used to protect individual privacy.

Record linkage (see data linkage)

Unit data refers to data organised on an individual by individual basis, so that for each individual, data from one source can be linked to data from another source. Personal identifiers are removed before the final linked file is used by researchers.

ANNEXES

Workshop program

Public Support of Data Linkage for Better Health

Workshop: Friday 16th August 2013

10.00 am	Welcome, background and linkage examples	John Mathews <i>Menzies Foundation</i>
	<i>Public benefits and future potential</i>	
10.20	Pharmaceutical Benefits Scheme- Case-studies	Sallie Pearson <i>The University of Sydney</i>
10.35	Aboriginal cervical cancer – work in progress	Lisa Whop <i>Menzies School of Health Research</i>
10.45	Expanding opportunities for data linkage	Louisa Jorm <i>University of Western Sydney</i>
11.00	Morning tea/ coffee & informal discussion	
	<i>Current arrangements and related issues</i>	
11.30	Project approvals & data access	Merran Smith <i>Population Health Research Network</i>
11.45	Legal and ethical issues	Judith Allen <i>University of Western Australia</i> Colin Thomson <i>University of Wollongong</i>
	<i>Community and researcher perspectives</i>	
12.15	Community perspectives	Carol Bennett <i>Consumers Health Forum of Australia</i>
12.30	Community attitudes to data linkage research	Vicki Xafis <i>University of Adelaide</i>
12.45	NHMRC and personal perspective	David Roder <i>NHMRC - Prevention and Community Health Committee</i>
1.00	Sandwich lunch – informal discussion	

SEE OVER FOR AFTERNOON PROGRAM

2.00 Panel discussion⁷ (Chaired by Terry Nolan, University of Melbourne)

Primary Discussants

Kay Patterson	<i>Australian Health Minister (2001-2003)</i>
Carol Bennett	<i>Consumers Health Forum</i>
Anne McKenzie	<i>Consumer advocate, WA</i>
Andrew Podger	<i>Australian National University</i>
Stephen Duckett	<i>Grattan Institute</i>
Melissa Sweet	<i>Freelance journalist & health writer</i>
Mark Metherell	<i>Consumers Health Forum journalist</i>
Mukesh Haikerwal	<i>World Medical Association</i>

The panel discussion will start with the workshop objectives (see below), identify the issues and the questions that arise, and suggest what might be done.

3.00	Comments on draft recommendations – led by	Carol Bennett David Roder
3.30	Tea and coffee break- informal discussion	
4.00 pm	Thanks and workshop close	

Workshop Objectives

1. To inform the community and the media about the public benefits of research based on the sharing and linkage of publicly-funded data, currently held in trust by various government agencies
2. To reassure the community and the media that such research can proceed with no risk to privacy
3. To encourage a frank discussion of the reasons for the long delays in approval and release of data for such research projects
4. To seek community support for the idea that publicly-funded data should be more readily available for public interest research projects that are ethically approved and scientifically sound.

⁷ Fiona Stanley, University of Western Australia, had agreed to chair the panel discussion, but had to cancel unexpectedly.

List of attendees

Presenters

Judith Allen	The University of Western Australia
Carol Bennett	Consumers Health Forum of Australia
Louisa Jorm	University of Western Sydney
John Mathews	Menzies Foundation
Sallie Pearson	The University of Sydney
David Roder	NHMRC / The University of South Australia
Merran Smith	Population Health Research Network
Colin Thomson	University of Wollongong
Lisa Whop	The Menzies School of Health Research
Vicki Xafis	The University of Adelaide

Participants

Carolyn Adams	Macquarie University
Michael Barton	CCORE /The University of New South Wales
James Best	The University of Melbourne
Annette Braunack-Mayer	The University of Adelaide
Nicholas Callinan	Menzies Foundation
Ying Chen	Victorian Data Linkages, Department of Health
Steven Darlington	Australian National University
Stephen Duckett	Grattan Institute
Sophie Dyson	Capital Markets CRC
Anna Forsythe	The University of Melbourne
Mukesh Haikerwal	World Medical Association
Anne McKenzie	Telethon Institute for Child Health Research
Mark Metherell	Consumers Health Forum of Australia
Terry Nolan	The University of Melbourne
Joan Ozanne-Smith	Monash University
Kay Patterson	Monash University
Dora Pearce	The University of Melbourne
Andrew Podger	Australian National University
James Popple	Office of the Australian Information Commissioner
Christopher Reid	Monash University
Sheena Reilly	Murdoch Childrens Research Institute
David Studdert	The University of Melbourne
Melissa Sweet	Sweet Communication / The University of Sydney
Margaret Whitstock	Independent Researcher

Observers

Phil Anderson	Australian Institute of Health and Welfare
Katy Balmaks	Department of Human Services
Anthony Bendall	Privacy Victoria
Chris Brook	Department of Health, Victoria
Kim Browne	Bureau of Health Information
Teresa Dickinson	Australian Institute of Health and Welfare
Kylie Jonasson	Department of Health and Ageing
Jane London	NPS MedicineWise
Alison McClelland	Productivity Commission
Gary Moorhead	Department of Human Services
Marcus Nicol	NHMRC
Melanie Taylor	National Health Performance Authority
Gemma Van Halderen	Australian Bureau of Statistics
Pia Waugh	Department of Finance and Deregulation

Panel Discussants

Terry Nolan (Chair)
Carol Bennett
Stephen Duckett
Mukesh Haikerwal
Anne McKenzie
Mark Metherell
Kay Patterson
Andrew Podger
Melissa Sweet

Apologies

Peter Carver	Victorian Data Linkages, Department of Health
Jane Freemantle	The University of Melbourne
Christine Giles	Cancer Australia
Sarah Hardy	Menzies Foundation
Fiona Stanley	The University of Western Australia

Acknowledgements and thanks

The Menzies Foundation is very grateful to all those who attended the Workshop, and most particularly to those involved in planning, those who presented papers, those who contributed to discussions on the day, and those who provided helpful comments on the draft report.

Anna Forsythe and Dora Pearce (University of Melbourne), Pam Shearman (Menzies Foundation) and Anne-Maree Arthur (Menzies Foundation) provided excellent support with Workshop planning, arrangements on the day, and in the preparation of the workshop report.