QIMR Berghofer
Medical Research Institute

SUBMISSION ON THE PRODUCTIVITY COMMISSION'S ISSUES PAPER ON

DATA AVAILABILITY AND USE

QIMR Berghofer Medical Research Institute

July 2016

Introduction

QIMR Berghofer is a world-leading translational medical research institute focussed on cancer, infectious diseases, mental health and chronic disorders. On most measures, it is in the top two medical research institutes in Australia. The Institute is a statutory body established under the *Queensland Institute of Medical Research Act 1945* and is also a registered charity.

QIMR Berghofer is pleased to make a submission on the Productivity Commission's issues paper on data availability and use.

Access to data is vital to the Institute's research into the prevention, diagnosis and treatment of diseases and disorders across its four research programs. QIMR Berghofer recognises that there are existing and potential sources of data that could be invaluable for research directed at improving understanding of diseases and disorders that cause incalculable human suffering and economic cost both in Australia and overseas.

In preparing this submission, we have addressed the questions raised in the issues paper relevant to medical research.

QUESTIONS ON HIGH VALUE PUBLIC SECTOR DATA

What public sector datasets should be considered high value data to the medical research sector?

The following datasets have been identified as containing data of high value to medical research:

- Ehealth records (including longitudinal data from primary health care – as available for the UK and other western countries)
- Medicare records (MBS, PBS data)
- Hospital inpatient data
- Notifiable diseases registries
- Cancer registration data
- National Death Index
- Registry of Births, Deaths and Marriages.

What characteristics define high value datasets?

In the health sector, high-value datasets are those that capture sensitive, accurate health information that can be linked to individuals. These characteristics permit health data to be linked across multiple datasets,

allowing researchers to build profiles of risk or survival based on health characteristics. Such data mining permits exquisitely sensitive models to be built that track health status over time, and model the effects of changes in exposure or treatment on patients.

At a higher level, such data also permit analysis of use of health services, thereby identifying patterns of care, use of resources, waiting times and outcomes.

What benefits would the community derive from increasing the availability and use of public sector data?

Assuming research is conducted properly, the potential benefits of such output are considerable, and would include applications as varied as:

- population-wide surveillance of rare but serious adverse drug reactions
- systemic inefficiencies in the delivery of services
- very early notification of epidemics (and their possible source)
- definition of genetic risk factors
- variations in treatment outcomes
- predictors of survival.

The list of opportunities is constrained only by imagination.

One practical example is QIMR Berghofer's involvement in the International Pseudomonas Consortium Database, which links metadata, genomics and human health for the development of treatments for cystic fibrosis infections (see https://ipcd.ibis.ulaval.ca/index). QIMR Berghofer has contributed about 25 per cent of the isolates in this BioBank. The benefits already gained from this study include rapid access to genome data on the largest global collection of Pseudomonas strains, rapid sequencing of strains for our future analysis linking with existing clinical and microbiological metadata and the development of a global network of researcher collaborations to allow knowledge transfer and opportunities for our research team to further their work in internationally respected laboratories. However, uncertainty over the BioBank ownership of samples and constraints in transferring strains and clinical metadata, and other data access issues, have led to a more limited range of isolates being able to be transferred to this international sequencing project, resulting in slower and more limited potential benefits of our contribution to this project to be realised. This is just one of many similar examples.

Which rules, regulations or policies create unnecessary or excessive barriers to linking datasets?

Many datasets are poorly annotated or not well described, greatly limiting awareness of their utility for researchers. For example, metadata for the full PBS dataset are not publicly available so it is difficult for researchers to know what they can ask for.

On a related issue, the requirements for individual consent for linkage to MBS/PBS data change over time such that approvals granted now are often not valid for subsequent linkages. This means that participants need to be re-consented for use of their data. This is a grossly inefficient practice and imposes enormous burdens on researchers seeking to capture complete, high-quality data for their studies.

Similarly, the processes to obtain linked **de-identified** data are extremely complex, especially if it is cross-jurisdictional (involving Commonwealth and state-based data sets). For example, accessing information from

the Australian Cancer database (which is held by the AIHW) requires multiple approvals, including from each state-based cancer registry. To then link this data set with others of interest (deaths or MBS/PBS) requires many more approvals such that the process can take a year or more.

Even when MBS/PBS data access is granted, under current policies, researchers can only access a maximum of five years of data, which makes longitudinal and long-term epidemiological and genetic analyses impossible.

Sharing of genetic data has become a condition on many US grants. The process to submit data to these repositories typically takes more than 12 months and takes up a huge amount of researcher time. It would be a major hurdle to our research if a similar process was required in Australia.

What private sector datasets should be considered high value data to the medical research sector?

All biomedical data in the private sector, including imaging and pathology data, GP records, private hospital records and private specialist records, and the contents of biobanks, should be considered high-value data.

Regarding principles, protocols or legislative requirements to manage concerns of private sector data owners, in both the US and in Europe (and probably in Australia, too, where in most states the hospital morbidity data sets contain information from both private and public hospitals) there are initiatives underway to establish ICT platforms for linking data in which ethical, legal and governance issues that relate to the data sources can be mapped such that the custodians retain control over availability of their data (as mandated by their respective legal and ethical constraints) in a way that is transparent to the end user. Advances in bioinformatics have led to the development of these initiatives. An example is the Observational Health Data Sciences and Informatics (OHDSI) Network – an international collaboration creating open-source solutions for performing large-scale analysis using observational health data (http://www.ohdsi.org/).

What would be the public policy rationale for any associated government intervention?

Government intervention is necessary to free up access to high-value data. Access to such data would lead to greater knowledge across many areas, such as:

- disease prevalence and incidence
- treatment patterns for diseases at the population level
- use and efficacy of pharmaceuticals
- adverse events.

This knowledge would have wide application for public benefit. For example, it would inform decisions about which pharmaceuticals should be listed and remain on the PBS.

What benefits would the community derive from increasing the availability and use of private sector data?

As many health services are provided in the private sector, information on patterns of treatment derived only from the public sector data would lead to a very skewed picture of health service use.

The benefits the community would derive from increasing the availability and use of private sector data in combination with public sector data include:

- economic benefits, such as the more efficient use of services and medicines
- health benefits, such as safer prescribing and dispensing of medicines.

Many of the health conditions that consume large proportions of the health budget are managed outside of the hospital system. Access to eHealth records and other data held in the private sector (e.g. radiology and pathology results) would provide much more complete information on prevalence, geographic variation, and treatment patterns for common, non-acute conditions.

What impediments currently restrict consumers' access to and use of public and private sector data about themselves? Is there scope to streamline individuals' access to such data and, if there is, how should this be achieved?

Consumers' access is most likely restricted due to lack of knowledge or awareness about existence of stored data – as non-specialists, most members of the public naturally have very little knowledge of what data are available. Apart from having an accurate record of their health, the direct benefits to the individual might be relatively small.

At a societal level, providing researchers and clinicians with greater access to data would likely provide indirect benefits to health consumers as the result of, for example, fewer inefficiencies in the health system, a better understanding of the long-term consequences of particular treatments, and a better understanding of the causes and patterns of disease in our community.

How should the costs associated with making more public sector data widely available be funded?

A user pay levy seems most appropriate. As long as the prices were reasonable (e.g. cost recovery), then these charges this could be built into research budgets as legitimate expenses involved in conducting the research.

Is availability of skilled labour an issue in areas such as data science or other data specific occupations? Is there a role for government in improving the skills base in this area?

There is definitely a shortage of skilled analysts who have the requisite training and understanding of the issues. Specialists in data linkage, data analysis and biostatistics are all in short supply in Australia and need to be trained. There is also a shortage of people trained in the ethics of data linkage and data sharing. The absence of ethicists educated in this area contributes to some of the perceived barriers. It is easier for committees and custodians to deny access to data than to think through solutions to safe release of data.

What types of data and data applications (public sector and private sector) pose the greatest concerns for privacy protection?

The following types of data pose the greatest privacy concerns:

- financial data
- health data (granular consultation information)
- genetics data

(For an extensive review on data privacy concerns in Australia, including commissioned research of the Australian population involving quantitative and qualitative surveys, see the report entitled *"The Impact of Privacy Legislation on NHMRC Stakeholders"* prepared for the National Health and Medical Research Council of Australia by Campbell Research and Consulting (July 2004).)

How can individuals' and businesses' confidence and trust in the way data is used be maintained and enhanced?

Confidence and trust could be maintained and enhanced by:

- ensuring that the Australian privacy principles are upheld
- transparent open ledger of data use agreements
- strong penalties for breaches
- increasing the use of secure computing environments such as SURE (Secure Unified Research Environment, Sax Institute) for analysis of linked, sensitive data sets.
- ensuring that the benefits of data usage are clearly demonstrated and can occur in a timely fashion.

How effective are existing approaches to confidentialisation and data security in facilitating data sharing while protecting privacy?

Existing approaches to confidentialisation and data security are very effective in protecting privacy. Data linkage units exist in many parts of the world and their operating principles are publicly available (e.g. WARLS, UK ONS, NSW, and QLD health linkage units). Basically, a third-party linkage group performs a link between two independent datasets using newly created linkage keys. Once the linkage has been performed and checked, the other identifiers are stripped and the new linked dataset cannot be re-identified. Further data security can be ensured with the use of secure computing environments for data analysis that mean that data sets cannot be copied and removed from the environment, meaning that only aggregate data can be held by individual researchers.

Data access issues

Health economics research requires complex data needs which would be significantly simplified and vastly improved if access to personal data and linkage across organisations was streamlined. Currently, modelling studies for examining cost-effectiveness or cost savings across interventions require a 'jigsaw' approach (1, 2). Multiple sources of data are pieced together requiring assumptions about 'averages' for the population. For example, to gain the full cost of a particular disease (3, 4), to assess whether a new intervention can improve efficiency, we typically need resource use and cost information to the health system (e.g., hospitalisation records, GP records, Medicare records or other health organisations), patient or family costs (e.g. surveys, interviews, Medicare out-of-pocket records) and third-payer costs (e.g. pharmacies, health insurance reimbursements, non-government organisation expenses). Using surveys is currently unavoidable but problematic for strong reliable evidence as they are subject to recall, selection and other biases (5, 6). None of these various sources are linked and enormous effort and expense is required within research projects to gain consent, access and manually link these sources to produce the information needed. Because of this very high research burden, in most cases only partial resources are collated, which can be insufficient. Critically, if we were to have greater linkage across health resources/cost sources, we could make huge strides to inform authorities and decision-makers of wastage in the health economy to potentially reverse these inefficiencies

(7, 8), translating to real gains for Australians.

Participants are willing to release personal data on financial burden of health (6, 9) (no privacy issues or breaches) under routine ethical approval and consent processes upheld at all research institutions. This has been via online, telephone and paper methods (6, 9). If ownership rights lie with patients to release data held at various health institutions, it would be a massive improvement for health economics research if patients had one electronic card and could log all their visits and consumption of healthcare services which could be released to researchers (all linked). Passive consent to access MBS and PBS data is used in our studies, however, this is partial data and does not include community, hospitalisation visits etc.

Similar arguments to the above could be made about epidemiological research. For example, identifying how particular exposures affect the risk of premature mortality might require information from five or more different data sets to ensure that all relevant information for analysis is captured.

In conclusion

QIMR Berghofer supports the Productivity Commission's inquiry into data availability and use. We trust the information above is of assistance and would be happy to provide more detail to the Commission if this would be of further assistance.

1. Tuffaha HW, Rickard CM, Webster J, Marsh N, Gordon L, Wallis M, et al. Cost-effectiveness analysis of clinically indicated versus routine replacement of peripheral intravenous catheters. Applied health economics and health policy. 2014;12(1):51-8.

2. Koh R, Pukallus M, Kularatna S, Gordon LG, Barnett AG, Walsh LJ, et al. Relative cost-effectiveness of home visits and telephone contacts in preventing early childhood caries. Community Dentistry and Oral Epidemiology. 2015;43(6):560-8.

3. Gordon LG, Eckermann S, Hirst NG, Watson DI, Mayne GC, Fahey P, et al. Healthcare resource use and medical costs for the management of oesophageal cancer. The British journal of surgery. 2011;98(11):1589-98.

4. Gordon L, Tuffaha HW, James R, Scuffham P. Economic Modelling of Healthcare Services for Prostate Cancer. Prostate Cancer Foundation of Australia, 2015.

5. Gordon LG, Patrao T, Hawkes AL. Can colorectal cancer survivors recall their medications and doctor visits reliably? BMC health services research. 2012;12:440.

6. Gordon LG, Walker SM, Mervin MC, Lowe A, Smith DP, Gardiner RA, et al. Financial toxicity: a potential side effect of prostate cancer treatment among Australian men. European journal of cancer care. 2015.

7. Thompson BS, Coory MD, Gordon LG, Lumley JW. Cost savings for elective laparoscopic resection compared with open resection for colorectal cancer in a region of high uptake. Surgical endoscopy. 2014;28(5):1515-21.

8. Kondalsamy-Chennakesavan S, Gordon LG, Sanday K, Bouman C, De Jong S, Nicklin J, et al. Hospital costs associated with adverse events in gynecological oncology. Gynecologic oncology. 2011;121(1):70-5.

9. Gordon LG, Beesley VL, Lynch BM, Mihala G, McGrath C, Graves N, et al. The return to work experiences of middle-aged Australian workers diagnosed with colorectal cancer: a matched cohort study. BMC public health. 2014;14:963.