



ACADEMY OF
THE SOCIAL SCIENCES
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 THE MINISTER FOR AGEING

- The Department of Health and Ageing is the custodian of significant data holdings
- Data access regimes involve long delays and costs to both data custodians and data users. Current efforts to remedy this should be expedited
- Improving data access procedures will enhance health services research and lead to increased understanding of how the health system works, potentially improving public policy
- The Department of Health and Ageing should establish a timetable to facilitate researcher access to routine data

With support from the Public Health Association of Australia and
the Health Services Research Association of Australia and New Zealand

Public health data and research

1. Payment systems for government benefits and services lead to the accumulation of large data sets, containing millions of data items. The primary purpose of the collection of the data is for payment and accountability. But an important secondary purpose is to use the data to understand patterns and trends. Secondary analysis of data sets collected for 'routine' or 'administrative' purposes is now a well accepted type of research.
2. Secondary analysis of government datasets has a number of benefits:
 - a. Because the analysis is conducted on government data, the research is almost always policy-relevant. It can shed insight into patterns of spending or service use that would otherwise not come to the attention of policy makers; and
 - b. Secondary analysis leverages investments that have already been made in data collection and so is generally a less expensive form of research. Secondary analysis of Australian data sets has addressed a range of issues including quality and efficiency of, and access to, health care.
3. Secondary analysis is increasingly common and procedures have been established by many data agencies to ensure confidentiality of data to ensure that individual privacy is not infringed by allowing further (secondary) use of the routinely collected data. Data released is de-identified (to protect privacy) and small cells are also suppressed to prevent any possible identification. Data release agreements generally require researchers to certify that they will not attempt to identify individuals and/or specific services.
4. Government data holdings, derived from claims processing, should be seen as an important public resource to assist in policy-relevant research which will benefit the Australian community. Failure to harness fully the potential of these data sets represents a significant lost opportunity both for policy development and research.
5. In this submission the Academy of the Social Sciences in Australia proposes a four point plan to facilitate access to routinely collected data to enhance policy-relevant research:
 - a. Consolidate information about the availability of data;
 - b. Streamline special access requests to focus on speed and ease of access;
 - c. Resource departmental data linkage efforts properly; and
 - d. Common use data sets
6. Advising researchers of the existence of data sets, and streamlining access procedures, would be an important way of facilitating health services research in Australia. Expansion of health services research has been an object of policy for the last few decades, most recently as outlined in the McKeon Review of health and medical research (<http://www.mckeonreview.org.au/>).

Consolidating information

1. Routine or 'administrative' data holdings are complex, with vast numbers of data elements collected. Some of these will be obvious and known to researchers e.g. Medicare Benefits Schedule (MBS) item numbers, but others may be system created e.g. information could be extracted about medical services based on year the service was delivered or year a claim is processed.
2. Information about the variables held and their definitions ('metadata') could be standardised and made publicly available, provided that necessary resources are supplied to maintaining this metadata.
3. Publication of information on data elements should be accompanied by standardisation of de-identification methods and publication of information about the rules used to protect privacy (e.g. suppression or randomisation of low frequency events). This could include allowing unit record data to be accessed in a secure manner and then applying confidentiality practices to the tables to be released.
4. Both publication of metadata and information about de-identification methods will help to make researchers aware of the data available and will also improve the quality of requests for data extraction as researchers would be better placed to specify requests precisely in terms of the data elements held.

5. *It is recommended that detailed information on data holdings, including data element definitions, be consolidated and made available publicly.*
6. Current administrative arrangements provide that claims processing for major government functions (e.g. claims against the Medicare Benefits Schedule) be undertaken by the Department of Human Services whilst policy responsibility for these functions is vested in the relevant policy line agency, in the MBS case, the Department of Health and Ageing.
7. The current arrangements for separation of payment and policy means that although claims data is held by the Department of Human Services, decisions about data release are made by the Department of Health and Ageing.
8. Data release decisions require co-ordination between DoHA, FaHCSIA and DEEWR and at present this process tends to be slow and cumbersome. A more streamlined management process should be implemented for cross-agency data release decisions.

Developing common use datasets

1. Common or public use data sets are standardised extracts of data or reports from surveys which are made available for external use. They can exist at various levels: detailed published tables, data cubes (interactive tables which allow for personalised extraction of aggregate data), confidentialised unit record files based on a sample of records, and researcher access to unit record data under secure conditions.
2. Many overseas jurisdictions release common use data sets. The United States Department of Health and Human Services has a very open policy and research using important data sets (such as the Medicare claims data) have been used in thousands of research papers.
3. Although special data extraction requests will continue to be required, secondary analysis would be facilitated by developing 'common use' data sets. Such data sets may meet up to two thirds of common data extraction requests.
4. Development and release of common use data sets will help to streamline access and provide an important resource for researchers and policy analysis. In addition to facilitating research which would answer defined questions, a common use data set would also facilitate exploratory research, allowing researchers to undertake preliminary analyses to test whether specific research questions may be worth pursuing and whether special data extraction requests are warranted.
5. Although many common use data sets could be considered for early release under a streamlined access proposal, *it is recommended that a data set consisting of confidentialised information of MBS use of a random sample of 10% of the population for a two year period be the first released. A new version of the data set should be issued annually.*
6. This should be seen as a first stage with similar common use data sets created for PBS utilisation and, over time, data sets linking MBS, PBS and hospital use. Additional common data sets (e.g. provider based) should also be developed over time.
7. The Australia Institute for Health and Welfare now has extensive experience with releasing data sets and has routinised its processes, making a customer commitment to handle data request speedily, generally within a fortnight. It is able to meet standards such as this because it has developed common use data sets for release from some of its key collections/surveys. These data sets are already 'confidentialised' (the data sets are called Confidentialised Unit Record Files).
8. Conditions for release of common data sets should be modelled on existing ABS procedures including:
 - a. Approval of the body to which data is released;
 - b. Conditions on secure storage;
 - c. Conditions on not identifying individuals in any publications;
 - d. Prohibitions on on-passing data to third parties; and
 - e. The payment of a nominal charge.
9. Each application for access to the common use data set should be required to provide a short outline of proposed use of the data to allow a determination that the proposed use meets a 'public interest' test as required by existing legislation.

10. *It is recommended that a timetable be published for development and release of a number of common use data sets, commencing with a timetable for release of a person-based sample of the Medicare Benefits Schedule.*

Streamlining access requests

1. Not all research questions will be able to be answered by the common use data sets so arrangements for data release in response to special requests will need to continue.
2. As awareness of the availability of secondary data has increased, so too have data requests. This has created bottlenecks and delays in data release.
3. Although creation of common use data sets should reduce special access workloads, existing data release consideration processes need to be streamlined. Streamlining would be facilitated by publishing metadata as proposed above, but other system standardisation approaches such as those adopted by the Australian Institute of Health and Welfare should be considered.
4. *It is recommended that the Department of Health and Ageing publish customer response standards, similar to those adopted by the Australian Institute of Health and Welfare, covering data access requests.*

Data linkage

1. Linking two or more data sets can be used to investigate more sophisticated policy and service interactions and more complex policy questions than can be answered by use of a single data set.
2. Data linkage needs to occur in a secure environment as it generally requires personally identifiable information to perform the linkage. Once linked, the personally identifiable information can be removed.
3. Protection of privacy concerns are even more acute with linked data sets and these have to be treated very seriously. However, there may be superfluous legislative barriers to linkage which should be addressed.
4. Over time, linked common use data sets should be developed and released (e.g. linking MBS, PBS and hospitals data).
5. Special data linkage requests will continue, especially for linkage to researcher-initiated surveys.
6. Policy and processes about data linkage are currently being redeveloped in Australia and so many requests for data linkage are on hold and some previous data linkage arrangements have been suspended.
7. Data linkage research is extremely powerful and it is recommended that data linkage arrangements (such as at the Australian Institute of Health and Welfare) be appropriately resourced to ensure timely response to linkage requests. Currently, AIHW has sought and received limited funding through the Population Health Research Network for data linkage. However this is on a limited fee-for-service basis. Appropriation funding could make data access for researchers more cost effective.

Summary of recommendations

- A. Detailed information on data holdings, including data element definitions, be consolidated and made available publicly
- B. A data set consisting of confidentialised information of MBS use of a random sample of 10% of the population for a two year period be the first released. A new version of the data set should be issued annually.
- C. A timetable be published for development and release of a number of common use data sets, commencing with a timetable for release of a person-based sample of the Medicare Benefits Schedule
- D. The Department of Health and Ageing publish customer response standards, similar to those adopted by the Australian Institute of Health and Welfare, covering data access requests.
- E. Data linkage arrangements (such as at the Australian Institute of Health and Welfare) be appropriately resourced to ensure timely response to linkage requests.

For further information

Please contact:

Andrew Swanson

Program Manager, Policy and Advocacy, International

Academy of the Social Science in Australia

Email: assa.admin@anu.edu.au or Phone: 02 6249 1788