Electronic Health Data for Primary Care Evaluation

Appendix A

Users Guide to health data access

This brief guide provides some simple approaches to planning a research project and locating appropriate data resources, the attached combined spreadsheets provide a 'browseable' version of the individual worksheets attached to the full report.

Basic approach for a simple project based on administrative data

1. Develop a clear research question
2. List information required to answer the question
3. Assess availability of these data use links provided in this report
4. Refine question according to data availability
5. Develop detailed description of data required
6. Discuss requirements with data custodian if possible
7. Download data from website or prepare a request for data custodian

NB HealthWIZ may be useful in both developing and refining your question and in performing an initial analysis, though more detailed information may be required for the final analysis

1. During the initial stages of research planning, you should develop a clear picture of what level of data you require. Keep in mind that the more detailed or identifiable the information required the more conditional access will be. These issues will have consequences as regards permissions, time resources and funding required for the project.

Levels of data available and their access requirements

- *Freely accessible electronic report/ interactive data cube.* You must seek permission from the custodian if you plan to reproduce data in a publication.
- *Aggregate de-identified data not available on web-site or in a report;* application to the custodian is usually required and a charge may be levied (ethics approval may also be required).
- *Identifiable data* will require ethics approval and if granted may need consent of individuals, depending on the nature and purpose of the research.

Issues for researchers using electronic data collections

- **Data security** - Any receipt of data requires researchers to provide a secure data environment. Provision for secure storage for up to 7 years after completion of research may be necessary. In some cases, de-coding and /or destruction of data may be required.
- **Know your data.** Expert advice on data definitions, peculiarities and limitations should always be sought from the custodian during the planning stage of a project and reviewed prior to commencing analysis. Administrative data are quirky and erroneous results are guaranteed if current and accurate knowledge of the data is not acquired in advance.

2. If funding and time are limited it is probably advisable to request as little personal information as is consistent with the research requirements. Usually it is wise to also limit the number of fields requested; each field should be justified on the basis of the requirements of analysis.

Electronic Health Data for Primary Care Evaluation

Once you have decided what data you require, the following guidelines may assist in locating and accessing relevant datasets.

Federal

Federal data are accessible through three main web-sites:

1 The AIHW – the first site to search as it contains a great deal of accessible Health Statistics. In addition, most of the introductory pages will also describe the data sources so provides a good starting point as to the scope of information that is collected and is available (albeit maybe not in that particular publication). The title pages of these publications usually also provide a good source of contact information.

http://www.aihw.gov.au/

2 The Health Insurance Commission (if you require data that is collected by the government for medical claims (Medicare, PBS and DVA).

http://www.medicareaustralia.gov.au/about/our_organisation/annual_report/03_04/statistics.htm#1

3 Department of Health and Aging (similar data to HIC website though not as interactive, can also get general aged care data)

http://www.health.gov.au/internet/wcms/publishing.nsf/Content/Research+%26+Statistics-1

4 Australian Bureau of Statistics - another excellent source of data. All ABS publications are now available on-line (since December 2005). However if microdata are requested or specific additional requests are made, a fee will be charged.

Although these data sources are free and easily accessible, permission should be sought before using their data in a publication pursuant to usual copyright laws.

If the required information is not available, you will need to apply to the relevant body and confidentialised data can usually be provided, though usually at a cost. In addition, details regarding the nature of your research may have to be provided, Depending on the nature of the project, Ethics Committee approval may also be necessary.

If you require identifiable information then extra steps are necessary:

Ethics Committee approval is required and data release is at the discretion of the governing body.

All the federal data collection bodies generally either collect identifiable information (HIC) or if collecting from the states it is usually de-identified. This means that government departments can link data themselves (aged care datasets have been linked for example) but as a researcher it may be difficult to conduct a data linkage project. Where the federal dataset is a collation of state based reporting such as

hospital morbidity or national cancer statistics the issue of linkage is even more complicated as they individual states own the data. Refer to the NHMRC's guidelines for more detailed information on using identifiable data:

http://www7.health.gov.au/nhmrc/publications/hrecbook/01_commentary/14.htm

The ABS will not release identifiable data as they are bound by legislation which forbids it.

State holdings

In the short term, rather than approaching commonwealth agencies, it may be more expedient to obtain data from state and territory health departments directly. Data are updated more regularly and data supply may be more timely. For example, in NSW the Inpatient Statistics Collection is updated monthly. While data capture may not be complete for recent months (e.g. interstate hospitalisations won't be included) the data might be satisfactory for evaluation purposes.

Occasionally States will publish on-line reports but they can be a little hidden but are generally under the "publications" section of most web-sites. Other options that can be investigated include contacting the epidemiology/health research section for that jurisdiction (nomenclature may vary between states), or performing a general search of the web such as "Victorian perinatal statistics report".

Annual reports for health departments also sometimes contain useful data, particularly for the ACT. Reports released by the AIHW can also provide state data and/or information about state data holdings.

Researchers seeking state based data are advised to consider the following when attempting to locate a particular dataset:

- The Department of Health in that state is a good starting point, more specifically most state health departments have an epidemiology unit (it is occasionally called a health surveillance unit).
- Examine the relevant state university web-pages, as they quite often house national and state health databases
- For rarer diseases/conditions refer to the relevant disease support/research organisation.

If after exhausting the options above the required dataset cannot be located, it may be worth searching university web-sites of in that state or contact the appropriate organization directly if you require information regarding a particular disease or condition (for instance Diabetes Victoria, Cystic Fibrosis Australia).

Recommended approach for pharmaceutical outcomes evaluation:

- General practice databases should be used to examine which medicines are commonly used to treat specific conditions
- The DUSC database can be used to explore the extent of capture of medicines in the Medicare Australia data sets and also to examine trends in total community utilisation of medicines of interest

- Medicare Australia data should be used to track changes in service provision and use of investigations specific to the conditions of interest
- Linked Medicare Australia data (where available) should be used for detailed analyses of associations between the treatments and outcomes that are captured (coded) in the available data sets.

Recommended approach for access to hospital and mortality data

In general, the following data resources are recommended depending on purpose:

- The online NHMD data cubes should be used to examine overall annual trends, by age and sex, in hospital separations where the diagnosis of interest is adequately coded in ICD-9/10
- State and territory hospital separation data should be sought to examine short-term trends, by month, age, sex and Division of GP
- For optimal analysis of outcomes, linked data should be sought from the WA linkage unit
- AIHW NHMDS data should be used to examine national short-term trends, by month, age, sex and DGP
- An overview of deaths information is available from the ABS, individual information from state registrars or the AIHW's National Death Index.
- It may be preferable to request mortality data from state health departments or registrars as they may be more liberal with respect to releasing unit record data. However, this may vary from state to state and data release is likely to be dependent on the reputation of the organisation requesting data.

Resource Browsing

The following amalgamated spreadsheets are provided to allow researchers to browse for useful data holdings in a manner similar to browsing a library shelf for useful references. (All information is duplicated from the separate spreadsheets linked to the main document).

A list of data fields in each of the major data sets held by Medicare Australia
→ HIC.xls

A compendium of all the State Health Department data sets
→ State 1.xls

A description of all holdings that contain General Practice data resources
→ Primary Care Data Holdings 1.xls