

Appendix A

GUIDELINES FOR THE USE OF UNIQUE PATIENT IDENTIFIERS FOR DATA LINKAGE

The following guidelines are a first step towards a model code of practice for custodians of health data collections that are de-identified (i.e. they do not contain explicitly identifying information such as names and addresses) but which include unique patient identifiers (UPIs). The focus is on managing the capacity to match records in different data collections using the UPI. Mechanisms to control access to the matched information are also proposed.

These guidelines are intended to help data custodians to ensure that the data used in health statistical collections and research projects are de-identified and remain de-identified at all stages of their use, storage and eventual destruction. They illustrate 'best practice' in compliance with and the application of the Federal Information Privacy Principles, the National Privacy Principles and the section 95 and section 95A guidelines approved by the National Health and Medical Research Council under the *Privacy Act 1988*. This legislative framework is technologically neutral and must be complied with in the electronic environment.

These guidelines relate to the handling of de-identified statistical data rather than the collection of such data in clinical and administrative situations. However, privacy breaches can be avoided if organisations which manage data advise individuals about what data they collect and why, and ensure that the organisations and individuals have shared expectations in relation to directly related secondary uses and disclosures of the data including the fact that de-identified data may be used for research or statistical collections.

MINIMISATION OF POTENTIALLY IDENTIFYING INFORMATION

Any use of statistical data resulting from linkage of records from more than one collection must be accompanied by steps to prevent individuals being identified or recognised by users of the data. As a general guide, the following principles should be considered and exceptions documented:

- When a UPI is used to create a data set by linking data from two or more sources, the UPI should be removed from the data set or encrypted before it is made available to the research team.
- Other potentially identifying data items such as the unit record number assigned to the patient by the hospital or other health care provider should be removed or encrypted before the data set is made available to the research team. It may also be necessary to ensure that the hospital cannot be identified, especially for small hospitals or those that serve small communities.
- Detail in data items should be reduced to the level necessary for the research. For example, age would normally be computed from date of birth and length of hospital stay would normally be computed from dates of admission and discharge.
- Where possible, data items should be aggregated to the level that is needed for the research project. For example, Statistical Local Area or postcode of residence should normally be aggregated to larger geographical units such as the Statistical Division or health region unless the focus is on a specific small area. Similarly, country of birth or language should normally be restricted to major groups or specific countries or languages of interest rather than used in a form that identifies every country or language (however uncommon) identified in the collection. In accordance with standard statistical practice, tabulations with less than five individuals in a single cell should be avoided in research work and should never be published.
- Diagnosis and procedure codes should only be released with a three-digit ICD-10-AM level of detail unless there is a specific need for greater detail.
- In addition, cross-tabulations of data items should be limited to those that are strictly necessary for the research. For example, while Indigenous status, place of residence, country of birth and preferred language may all be relevant to a health research project, a four-dimensional cross-tabulation of these variables would usually be unnecessarily cumbersome and would often include an unacceptable number of cells with only one or two individuals.

SUPERVISION OF THE USE OF DATA

The following general principles should be applied to most research projects using data sets that either have been linked or are capable of linkage:

- Projects involving the linkage of client level data should be considered by an institutional or departmental ethics committee established in accordance with the guidelines issued by the National Health and Medical Research Council.
- There should be a clearly documented and agreed method for overseeing the project and monitoring linkage and the use of UPIs. This should include explicit procedures and sanctions designed to ensure confidentiality and adherence to best practice as well as relevant legal obligations.
- Security measures and technical protective measures should be specified. This would include details of precautions taken to ensure the physical security of data and prevent unauthorised access to computer systems. Agreed minimum standards should be specified.
- Regular audit procedures designed to identify unauthorised or inappropriate access to data should be adopted. All access requests and uses of data should be logged to provide audit trail information.

DATA EDITING

Research projects using linked data may need to incorporate consistency checks to detect errors in the original unlinked data sets (e.g. there may be inconsistencies between the dates recorded for hospital episodes or vital events in two data sets which may only become apparent after the data sets have been linked). As far as possible this should be applied before data sets are linked to minimise the backtracking from the linked records to the original data sets.

SUBSEQUENT USE AND DESTRUCTION OF DATA SETS

- Rules governing the retention or destruction of data files or data sets after the analyses have been completed need to be implemented, allowing for time for results to be checked and research reports to be refereed.
- Restrictions need to be placed on linkage to data sets other than those that have been approved.
- A register of data releases, termination and destruction should be maintained and methods for regular reporting on progress of long-running research projects should be incorporated.

Conditions of this type are often imposed by data custodians but may not always be rigorously enforced. For this reason, custodian agencies that handle a large number of data requests may need to adopt proactive procedures to ensure that the use of data sets is terminated on or before an agreed date, including a specified period to destroy or de-identify data and related audit procedures. Typically a data set would be made available for a specific number of months or years after which the custodian agency responsible for custody of the data would contact the recipient if necessary in order to satisfy itself that the research had been completed without any breaches of privacy and that the data had been archived, returned or destroyed in a satisfactory manner. Further research projects or extensions of time could then be considered on their merits rather than taken for granted.

While the guidelines would need to be tailored individually for each project, the following standard conditions of release used by one State health authority (Victoria) provide a useful model:

- The data must not be used, published or disseminated in a way that might enable the identity of individual patients or the service profiles of individual doctors or private hospitals to be ascertained.
- The data file is provided solely to the recipient and must not be communicated to other persons or organisations, or linked with files of personal information of other sources, without the prior agreement of the health authority.
- The data will only be used for the purpose(s) outlined by the recipient in requesting the data or for purposes approved by the health authority's ethics committee.
- Data files are to be maintained and stored in a secure manner in an environment where they cannot be linked (either electronically or by personal inspection) with other patient records or patient—level data or personal information.
- When no longer required, or by an agreed date, the data files are to be destroyed or returned to the health authority and the authority is to be notified of such destruction.

If data files are made available to consultants engaged by the recipient then the consultants must also agree to these conditions and the health authority must be provided with written evidence of such agreement.