Issues in the use of unique patient identifiers to link statistical data

Ric Marshall, Neil Powers and Carmel Heaney
Victorian Department of Human Services

When people go to seek healthcare services there are two concerns that are usually of leading importance to them. Firstly, and usually of most importance, they want healthcare that will work and has been proven to work – that is they expect that the healthcare system will be managed by rigorous best practice and quality research. Secondly, they expect that the information about themselves that they give to the healthcare provider will be kept private.

1. Introduction

This paper describes the current approach adopted by the Victorian Department of Human Services to the challenge of making linked administrative data on health service utilisation available for research while ensuring the optimal protection of individual privacy. It discusses two major projects which are designed to further this objective – the development of a central patient register and the specification of best practice guidelines for the use of unique patient identifiers to link de-identified statistical data from different administrative information systems.

2. The development of a central patient register

Like all other State and Territory health authorities, the Department has made de-identified unit record data from its hospital morbidity collection available for ethically approved research projects for many years. However, the potential for data linkage has been limited by the anonymisation process, which ensures that names and addresses are not provided to the central data collection. Thus linkage is only possible through the probabilistic matching of data items such as the hospital-specific unit record number, the Medicare card number and dates of birth, admission or discharge.

Recent developments in legislative provisions have enabled development and trialling of registers of patients who use services across Victorian public hospitals. While the ultimate objective of these registers is enable patients to build a whole of life record that is accurate and portable, with timely access to this information for clinicians, these registers also have the potential to enable improved linkage of statistical data sets for planning, policy and approved research developments at State and regional levels.

2.1 The use of a central patient register for clinical purposes

It is anticipated that a central patient register will provide for effective care across a continuum of time and service settings, requiring access to a comprehensive and accurate longitudinal clinical record. This is in contrast to the current situation where clinical information is held by various health service providers on a number of different technical and software platforms, making information integration an unnecessarily complex task.

The need to accurately identify patients for the purposes of linking disparately stored components of their clinical record is vital in the clinical setting. This requires the development of a central patient register encompassing healthcare providers across a whole jurisdiction – in this case a master patient index for the Victorian public hospital system.

The patient register would enable patients to register for treatment in a single state level public healthcare system. Patients on the central register would be uniquely identified as patients of the Victorian public hospital system rather than as patients of individual hospitals. These patients would then be able to authorise their health service providers to access relevant clinical information from previous health encounters wherever they have occurred in the public hospital system.

An essential component of a central patient register would be some form of unique patient identifier (UPI) to maintain a link between records held on identified individuals across service settings. This UPI could either made available to patients and service providers on a card or some other paper record so that it could be produced or quoted at each encounter like a hospital unit record number, or it could be allocated and maintained as purely an internal system or encrypted number to ensure patient privacy.

The extent to which privacy precautions such as encryption are applied to the UPI may depend on the structure and content of the register. One option would be for the register to be no more than a simple patient index acting as a pointer or gateway to where information is stored, thus enabling a virtual clinical record system for public hospitals across Victoria. This would allow sensitive information to remain at the source of treatment, allowing individual institutions to maintain control over who has access. A second option would befor the register to extend beyond a patient index to include clinical information shared between service providers and this would probably require a higher degree of data security and privacy protection.

More specifically, a range of issues will need to be addressed before a central patient register including a UPI can be established and used for clinical information exchange. These issues include:

- guidelines to manage the issue of patient consent for participation and access to information held on the register,

• client information on the possible effects on treatment advice and outcomes of not participating in or not including all clinical encounters in the register,

• governance issues around custodianship of the register,

• clearly agreed rules for authorising, recording and auditing access to a fully identified data set for duplicate resolution,

• the methodology to be used by each healthcare site to update its local systems when patient information changes or is corrected,

• establishment of data standards within and across jurisdictions (e.g., HL7, SNOMED, Australian Standard for Client Identification),

• progress towards implementation of comprehensive clinical systems at local level, and

• provision for patients who wish to 'mask' particular events or service facilities and access components of their healthcare anonymously.

2.2 The use of a central patient register for statistical purposes

This will be based on the same fundamental principle as the Department's existing hospital based data collections, which is that information must be de-identified or anonymised before it can be used for purposes other than in the direct treatment and care of an identified individual.

Data linkage between different administrative data sets can be used to enhance epidemiological research (by enabling researchers to count people rather than admissions) and to provide improved information output for resource allocation, purchasing, planning, policy development and reform, as well as the ability to monitor performance and assess quality.

If properly managed, a central patient register with a UPI should make it easier to protect the identity of individual clients when linking service utilisation data from different sources.

Firstly, an electronic register can be designed to control access to the full set of individual identifiers giving the name, address, etc corresponding to each UPI. Each access can be logged and the person accessing the data can be held accountable for the use of the information. It is essential that this information is only accessed in a fully accountable way and only when essential for accurate linkage processes. The risk of patient identification can be further reduced through technical security procedures (such as firewalls, encryption, audit trails) and management procedures (including managing access requests to data for research purposes).

Secondly, the UPI can be used for linkage without the explicit identifiers (name, address, etc) and then either removed or encrypted before the linked data are made available for analysis. This should provide more reliable linked data and thus minimise the need to reverify the data against the original source datasets.

Thus by limiting access to anonymised, linked data a central patient register should lower the level of risk of privacy intrusion and increase community confidence in this type of statistical research.

3. Guidelines for the use of Unique Patient Identifiers

The following guidelines are currently under consideration by the National Health Information Management Group as a first step towards a model code of practice for custodians of health data collections that are de-identified (ie, they do not contain explicitly identifying information such as names and addresses) but which include unique patient identifiers (UPIs). They are based on current practice by the Victorian Department of Human Services in managing de-identified hospital morbidity statistics.

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 (IPPs), the National Privacy Principles (NPPs) and the Section 95 and Section 95A Guidelines approved by the National Health and Medical Research Council (NHMRC) 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.

3.1 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.

3.2 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.

3.3 Data editing

Research projects using linked data may need to incorporate consistency checks to detect errors in the original unlinked data sets (eg. 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 back tracking from the linked records to the original data sets.

3.4 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 datasets 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 pro-active procedures to ensure that the use of data sets is terminated on or before an agreed date, including a specified period to destroy/de-identify data and related audit procedures. Typically a data set would be made available for a specific number of months or years after that 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.

3.5 Standard conditions of release

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.