

## Privacy issues

A primary requirement in any health data collection is to protect the privacy of the individual. In statistical collections this is usually achieved by the use of de-identified records and by adopting a rigorous protocol to minimise the risk of re-identification. The unit records held in de-identified statistical collections are said to be 'anonymous'; in that transparently identifying information (such as names and addresses) is either not collected or is removed before the unit records are made available for statistical or research purposes.

The inclusion of unique patient identifiers in these collections may increase the risk for individual subjects to be identified (or re-identified) in at least two ways.

Firstly, individual UPIs may be matched with transparently identifying information such as names or addresses. This risk can be managed by business rules governing the link between patient's names and their UPIs, supported by technical barriers such as encryption restricting the ability of users to make this linkage.

Secondly, the UPI may be used to link data relating to the same individual in two or more data sets in such a way that the individual, although still 'anonymous', is more easily identifiable through a combination of data items that together uniquely describe the individual. An individual may become recognisable through a combination of data items that may or may not include the UPI. For example, the UPI may be used to link morbidity from different hospitals in different jurisdictions and a file of combined patient level data may be released to a researcher after the UPI has been removed. However, if the file contains dates of admission and discharge for every hospital episode in a person's lifetime there may still be an unacceptable risk that users may be able to identify the individual concerned, especially if some of the hospitals are small or if too much geographical data about the person's place of residence are released; this could have adverse impact not just on the person's privacy but on their standing in the community and willingness to seek further health care. This risk can also be managed by business rules governing the level of aggregation or disaggregation required for data to be released for research and planning purposes.

As the technical scope for linkage increases, these issues will need to be addressed in the context of both existing statistical data sets, such as the hospital morbidity collections and data collected under Medicare and the Pharmaceutical Benefits Scheme, and future data collections that may be established as a result of initiatives such as *HealthConnect*.

In fact, three classes of data collections may be need to be defined:

- the UPI register containing clients' UPIs, names and other demographic information—this would be used for client registration and for resolution of possible duplicates;
- statistical unit record data sets containing individual client records each of which would include the client's UPI or a statistical linkage key, but not the client's name;
- data sets for research (possibly linking data across more than one statistical data set) that include encrypted UPIs as an additional safeguard against identification of individual clients, especially where the user may be able to access the UPI register.

## PRINCIPLES FOR USE OF IDENTIFIED RECORDS

### **Patient master indexes**

Patient master indexes are already maintained by most hospitals and health care providers. Hospital indexes generally include each patient's name, address, date of birth, and so on, as well as a hospital unit record number and some clinical or service-related information such as service dates, diagnoses and medical alerts. Such systems are mainly used to link clinical information within a single hospital; however, there is an increasing tendency to extend linkage across service providers by establishing consolidated patient master indexes at the regional level. This type of index may be maintained by multi-hospital agencies such as an area health service (in New South Wales) or a metropolitan health service (in Victoria). Some State and Territory health authorities are also developing statewide patient indexes with the potential to cover their entire public hospital systems. (This has already been achieved in the Northern Territory.)

### **Population registers**

Population registers are designed to cover an entire population or sub-population without restriction to a particular group of service providers. At the population level the most obvious example is the register of Medicare card numbers and internal personal identification numbers maintained by the Health Insurance Commission. A more restricted example would be the register assigning 'DVA numbers' to persons entitled to benefits from the Department of Veterans' Affairs.

There are also national and State/Territory registers relating to specific health issues, some of which contain names. For example, the Australian Institute of Health and Welfare maintains the National Death Index and the National Cancer Statistics Clearing House, both of which contain explicitly identified information that is protected under the *Australian Institute of Health and Welfare Act 1987*. There are also an increasing number of specific health issues registers that do not contain names but that may contain some form of UPI.

Each of these indexes and registers is an example of a UPI system and in each case access to the names and UPIs contained in the system is governed by business rules or in some cases by legislation (or both). These rules are primarily designed to protect individual privacy while facilitating the clinical and administrative purposes of the information system.

With the growing use of electronic health records and electronic messaging, however, the general trend is:

- for selected agencies to be provided with access to the names and numbers in the index for approved clinical or administrative purposes, but on the other hand
- for this access to be governed by privacy principles or legislation that may include requirements for individual consent (either on an 'opt in' or 'opt out' basis) thus making it difficult to use the UPI for statistical purposes.

For example, the *Medicare and Pharmaceutical Benefits Programs Privacy Guidelines* issued under section 135AA of the *National Health Act 1953* place limits on data linkage between Medicare benefits data and pharmaceutical benefits data. Even with patient consent, the use of the Health Insurance Commission's internal personal identification number to link such data is prohibited except in specific instances such as the Coordinated Care Trials conducted by the Commonwealth Department of Health and Ageing. However, section 2.3 of the Guidelines permits the routine provision of the Medicare card number in an encrypted form and the internal personal identification number to the Department in conjunction with de-identified or anonymised claims data for a range of public policy purposes some of which may involve linking records relating to the same (unidentified) individual. The use of the data by the Department is then governed by Part B of the Guidelines, in particular section 5, which includes safeguards against the re-identification of the claims data.

## PRINCIPLES FOR USE OF RECORDS CONTAINING STATISTICAL LINKAGE KEYS

Statistical linkage keys that consist of date of birth and some of the characters of the client's name were developed to facilitate linkage within and between relatively small or specialised data sets where duplicate keys were unlikely. These keys were intended for linkages for statistical purpose only and were never intended to be used in clinical or client management settings. In addition, if the key is not encrypted the risk of direct identification or re-identification of clients from their SLK is greater than from a numeric UPI. Thus it is recommended that systems containing SLKs adopt the rules described above for name-identified records.

## PRINCIPLES FOR THE USE OF UNIQUE PATIENT IDENTIFIERS

Agencies managing or acting as custodian for statistical collections that include UPIs need to adopt business rules and technical barriers that restrict the capacity of users to match the UPI to the individual's name. The key issue to be addressed by these business rules concerns access to the system or systems containing both UPIs and patient names and addresses (e.g. the patient master index or the population register).

While the precise business rules may differ from collection to collection, the basic principle is that, where the UPI is used for clinical or administrative purposes, as well as to link records for statistical purposes, the personnel who use the UPI for clinical or administrative purposes should not normally be able to access additional information on identified clients who have not consented to this access. This could be achieved by encrypting the UPI before it is used for statistical linkage.

## PRINCIPLES FOR THE USE OF ENCRYPTED UNIQUE PATIENT IDENTIFIERS

The encrypted UPI could be used in the same way as the statistical linkage key (SLK) has been used to match records in statistical data sets. The result using UPIs should, however, be more accurate than linkage using an SLK because of the possibility of fewer missed matches and mismatches due to the increased discriminating power of UPIs.

The basic principle is that de-identified information used for statistical, research or planning purposes is not used or disclosed in such a way that an individual's identity can be ascertained. Provided that it remains de-identified, information used in this way does not fall within the definition of 'personal information' incorporated in all current privacy legislation.

However, it is essential that patient privacy is maintained in any de-identified statistical collection, even if there is a possibility that a small number of records may be identifiable by particular users due to the unusual nature of the records. This places a responsibility on the custodians and users of data sets to rigorously manage data to minimise the risk of identification and ensure ongoing ethical handling and disposal of all unit record data. It also provides them with a clear specification of reasonable steps to manage risk of identification. Ethical data handling practices also need to be specified and assured to guide users of data in situations where potential or actual recognition occurs as a result of unpredictable circumstances or a conscious attempt to breach the spirit of the privacy principles.