

The use of linked data in health surveillance in New South Wales

L. Taylor, K. Lim, T. Churches

Epidemiology and Surveillance Branch, New South Wales Department of Health, Sydney.

Abstract

The Epidemiology and Surveillance Branch of the NSW Health Department began experimenting with probabilistic record linkage in 1993, and has carried out a data linkage program since 1994. The program has two components: annual linkage of selected data sets that are of particular importance for population health surveillance; and ad hoc linkages to meet specific needs.

Within the NSW public health system, data linkage occurs in a complex legal, ethical and policy environment. The collection and use of personal health data are governed by the NSW Public Health Act 1991, the NSW Health Administration Act 1982 and the NSW Privacy and Personal Information Protection Act 1998. Policy for linkage of personal health data is described in the Health Department's Information Privacy Code of Practice 1998. The policy provides for referral of certain proposed linkage projects to the Department of Health Ethics Committee, which operates in accordance with NHMRC guidelines.

The Epidemiology and Surveillance Branch of the NSW Health Department started experimenting with probabilistic record linkage in 1993 and has carried out a data linkage program since 1994.

This paper describes some of the data linkage projects that we have carried out, some of the legal and ethical complexities of carrying out health data linkage in NSW, and will touch on two recent developments.

In 1994, we started off carrying out a series of ad hoc linkages to answer specific questions. As time passed, it became apparent that some issues are of continuing interest and importance. The linkage program now comprises a combined approach of annual linkage of selected health-related data sets that are of particular importance for population health surveillance plus ad-hoc linkages as the need arises.

Linkage of the following datasets occurs annually

Linkage of the NSW Inpatient Statistics Collection (ISC) and the NSW Central Cancer Registry (CCR) data is carried out to provide information on patterns of treatment for cancers. In this linkage the Cancer Registry data contributes detailed information on the types and morphologies of cancers and the ISC contributes information on hospital-based treatments. This particular linked dataset has been used for a variety of projects, examples of which are shown in Table 1.

Secondly internal linkage of the ISC is carried out annually. The ISC is a database of hospitalisations not of persons. Internal linkage of the ISC allows us to estimate the numbers of persons with certain diseases. For example, the person-based rate of haemodialysis (a proxy measure for renal failure) is very high among indigenous men and women living in rural areas (Figure 1).

Information on notifiable infectious diseases is collected under the NSW Public Health Act and compiled into the Notifiable Diseases Database. The NDD is a database of cases, rather than notifications, and information on cases is compiled by Public Health Units throughout the State. However duplicate cases can be created if patients are tested or treated in different parts of NSW. Internal linkage of the NDD data at state level allows these duplicates to be reconciled. For example, when we carried out internal linkage of the NDD in 2000, the estimated number of cases of hepatitis B and hepatitis C were reduced by about 15%. So internal linkage of the NDD helps us to get better estimates of the true number of cases of infectious diseases, particularly chronic infectious diseases.

Linkage of the NSW Midwives Data Collection (MDC) to the Registry of Births, Deaths and Marriages birth registration data. These linked data have been used for two purposes.

First, some births are found to have been reported to the MDC but are not registered with the Registry of Births, Deaths and Marriages. Information on unregistered births is forwarded to the Registry.

Second, we can obtain improved estimates of the number of babies born to Aboriginal and Torres Strait Islander mothers using capture-recapture methods and information on maternal Aboriginality from each data collection. Aboriginality is generally not well reported in routinely collected data in NSW. While it is estimated that reporting of maternal Aboriginality has improved from 75 to 88 per cent in rural areas between 1994/1995 and 1999/2000, in urban areas in 1999/2000 it was estimated that only 45 per cent of Aboriginal mothers were reported as Aboriginal (Figure 2). I'd like to acknowledge the work of the Victorian Perinatal Data Collection Unit, who first came up with the idea of using linked data for this purpose.

Finally, linkage of the MDC to the ISC is carried out routinely. These linked data have been used for two main purposes in surveillance. As the MDC has the baby's date of birth and the ISC has the mother's date of discharge, the linked dataset enables reporting of post-natal length of stay in hospital. The ISC contributes reliable information on health insurance status and the linked data can be used to examine obstetric interventions by public/private health insurance status.

In addition to linkage of these datasets each year, over the past three years a variety of ad-hoc projects have been or are being carried out:

- Linkage of the Royal North Shore Spinal Injury database with the ISC to examine long term morbidity following spinal injury
- Linkage of pregnancy ultrasound data from the Bankstown-Lidcombe Hospital with the MDC and NSW Birth Defects Register to examine false positive and false negative results of prenatal diagnosis with ultrasound for quality assurance purposes.
- Linkage of the St George Hospital Hypertension in Pregnancy Database with the MDC to support a study of the risk of recurrent hypertension in pregnancy.
- Internally linked ISC data are being used to examine pregnancy rates following tubal sterilisation.
- Linkage of the Sydney Obstetric and Gynaecological Ultrasound database with the MDC to examine miscarriage rates following amniocentesis in a specialised obstetric ultrasound practice.

In addition to the technical complexities of having a number of different data linkage projects going on over time; data linkage occurs in a complex legal, ethical and policy environment.

The collection and use of personal health data are governed by the NSW Health Administration Act 1982, the NSW Public Health Act 1991 and the NSW Privacy and Personal Information Protection Act 1998. Data collections from other sources may be covered by the other Acts, such as the Registry of Births, Deaths and Marriages Registration Act 1995.

Policy for linkage of personal health data is described in the Health Department's Information Privacy Code of Practice 1998. The policy provides for referral of certain proposed linkage projects to the Department of Health Ethics Committee, which operates in accordance with NHMRC guidelines.

The decision chart as to whether a project should be referred to an ethics committee is shown at Figure 3. The Code of Practice allows linkage of data collections owned and managed by the Department to be carried out with the approval of the Director-General (shown as CEO) if the purpose of linking the dataset is consistent with the original purpose of collecting the data. However, linkage of data collections owned and managed by the Department for purposes other than that for which they were originally collected must be reviewed by an ethics committee, as does linkage of Health Department Data Collections with data collections owned by other organisations.

This process is best illustrated by an example. Figure 4 shows a schematic representation of a project linking the NSW Midwives Data Collection, the Inpatient Statistics Collection and the Registry of Births, Deaths and Marriages (RBDM) birth and death registration data for children up to one year of age.

The reasons for carrying out the linkage were to create a dataset that allows us to:

- Provide identified information on unregistered deaths to the RBDM.

- Examine the association between infant morbidity and mortality and perinatal factors

- Obtain estimates of maternal postnatal length of stay and obstetric interventions by health insurance status
- Assess the quality of reporting of Aboriginality to the MDC

In the example here the linkage includes external data owned by the Registry of Births, Deaths and Marriages. Hence the linkage needs to be considered by the Department's Ethics Committee.

The legal and ethical issues in this project relate to the use of personally identifying information on individuals.

In regard to the MDC/ISC linkage, the MDC data are required to be collected under the NSW Public Health Act and the ISC is covered by the provisions of the NSW Health Administration Act. The linked data will be used for epidemiologic purposes, that is monitoring patterns of diseases in the population and determinants of health, which are consistent with the purposes for which the data were originally collected. So the proposed linkage is permitted by law.

The linkage of birth and death registration data with health department data collections will assist in monitoring the quality of reporting of Aboriginality on the MDC and will provide information on outcomes of hospital care in terms of mortality. This is also consistent with use of Departmental data for purposes for which they were originally collected and therefore permitted by law.

The release of MDC data on unregistered births will assist the Registrar in fulfilling his obligations under the Registration of Births, Deaths and Marriages Act in which the Registrar is required to cause all births and deaths to be registered, and the data release is permitted under the NSW Public Health Act for this purpose. So that's lawful too.

Confidentiality is ensured by:

- restricting access to fully identified data to a minimum number of people during the linkage process
- ensuring electronic data is securely stored
- access to de-identified unit record data is password protected and subject to signing a confidentiality agreement
- release of identifying data to the RBDM is for a lawful purpose and limited to information necessary for provisional registration to occur (mother's name and address, baby date of birth, sex, plurality, birth order and hospital of birth).
- Information will be published in such a way that individuals cannot be identified.
- Deletion of identifiers after 5 years and deletion of de-identified data after 15 years.

In terms of benefits, most individuals whose information is being used would not benefit directly from this project. Some parents will benefit by being able to obtain birth certificates for children from the Registry of Births, Deaths and Marriages for previously unregistered births in an easier manner than previously. Previously, the Registry would seek confirmation of the

birth in writing from the hospital of birth. The main benefit is to the health system in being able to monitor trends in patterns of illness and health care among mothers and infants and the effectiveness of health care services in reducing preventable morbidity and mortality.

The project was recommended for approval by the Department of Health Ethics Committee.

In summary, carrying out a data linkage program is legally and ethically complex, in addition to the technical complexities of getting the linkage itself right.

In terms of current developments, there are two issues I'd like to touch on.

The first is the access to software. Most health data linkage programs currently use Automatch software. However, Automatch has now been taken over by Vality. This means there are no more updates of Automatch and the software has become very expensive. Other linkage software packages that are able to cope with large datasets also tend to be very expensive or difficult to deploy and use. The Epidemiology and Surveillance Branch of NSW Health Department and the Australian National University Data Mining Group are working on developing improved techniques for probabilistic record linkage that will be published under an Open-source software license. Peter Christen from the ANU Data Mining Group is presenting this project in another session at this symposium.

The second development is the NSW Health Records and Information Privacy Bill 2001, which is currently in exposure draft form.

The proposed legislation will:

- establish health information privacy principles to regulate the collection, retention, use, disclosure transfer and linkage of health information about individuals in the private and public sectors, and;
- provide individuals with a means to gain access to health information about them, whether held in the private or public sector, and to alter such information if it is inaccurate, out of date, incomplete or misleading.

Among other things, the draft Bill provides a privacy framework in which organisations may assign identifiers to individuals, for these identifiers to be used by both public and private sector service providers, and for information collected to be used for secondary purposes such as evaluation of health services. However, an organisation is not permitted to include health information about an individual in a 'health records linkage system' unless the individual has expressly consented to their information being included in the system. Systems that are considered 'health records linkage systems' will be described in the regulations to the Act.

If the NSW public health system implements a unique health identifier then the Act, if it is passed in its current format, has the potential to enhance population health surveillance efforts if the majority of the population consent to participate in the system.

The draft Bill is available on the Health Department's web site at www.health.nsw.gov.au/csd/lisb/HealthRecordsPrivacy/index.html.

In summary, we have found probabilistic record linkage to be ethically and technically challenging. Record linkage in NSW has grown from a bit of an experiment to a core component of our surveillance work – and a component that we wish to continue to promote and expand.

1. Breast cancer:
<p>i) Kricker A, Haskill J, Armstrong BK. Breast conservation, mastectomy and axillary surgery in New South Wales women in 1992 and 1995. <i>British Journal of Cancer</i> 2001, 85:668–73.</p> <p>ii) McGeechan K, Kricker A, Armstrong B, Stubbs J. Evaluation of linked cancer registry and hospital records of breast cancer. <i>Australian and New Zealand Journal of Public Health</i>, 1998, 22: 765–70.</p> <p>iii) Adelson P, Lim K, Churches T, Nguyen R. Surgical treatment for breast cancer in New South Wales 1991,1992. <i>Australian and New Zealand Journal of Surgery</i> 1997, 67:9–14.</p> <p>iv) Study of patterns of surgical care of carcinoma in situ of breast (in progress, National Breast Cancer Centre, NSW Cancer Council)</p>
2. Ovarian cancer:
<p>i) Study of survival for ovarian cancer by whether treatment provided at a specialist gynaecological oncology centre or elsewhere (in progress, NSW Cancer Council).</p>

Table 1 Projects using linked data of the NSW Inpatient Statistics Collection and NSW Central Cancer Registry

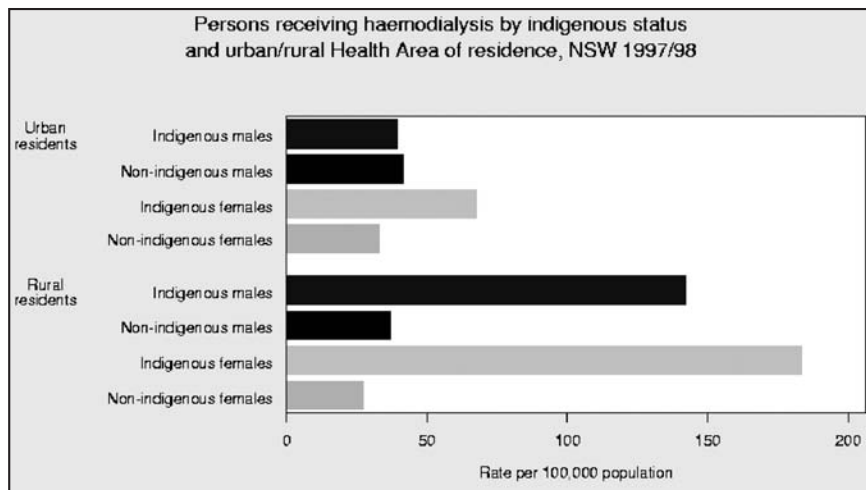


Figure 1
Source: NSW Inpatient Statistics Collection linked data and ABS population data. Epidemiology and Surveillance Branch, NSW Health Department.

Figure 2
Source: Linked NSW Midwives Data Collection and NSW Registry of Births, Deaths and Marriages birth registration data. Epidemiology and Surveillance Branch, NSW Health Department.

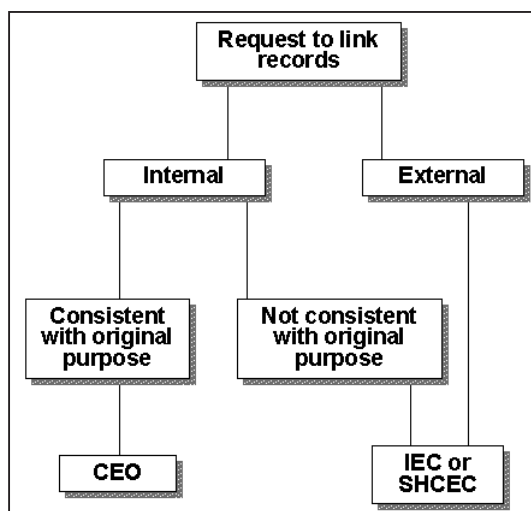
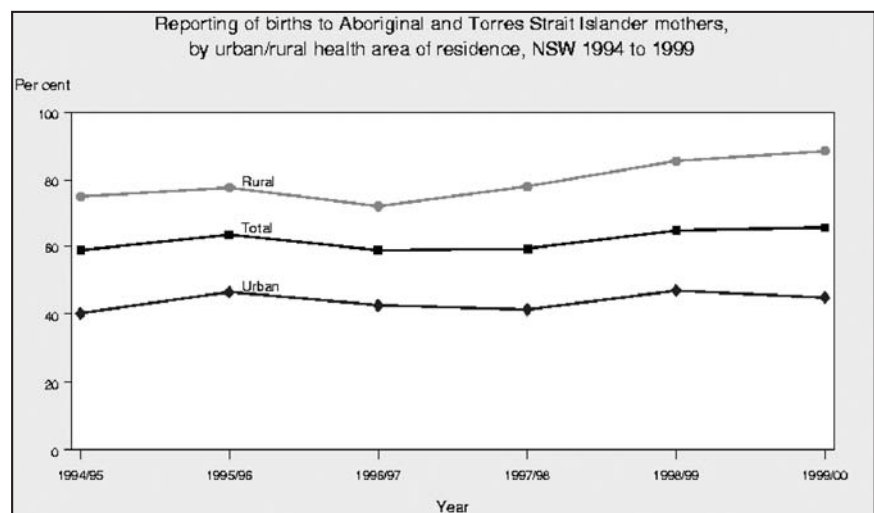


Figure 3 Decision path for requests to link records from two or more databases
Source: NSW Health Information Privacy Code of Practice – Second Edition. Sydney: NSW Health Department, 1998.

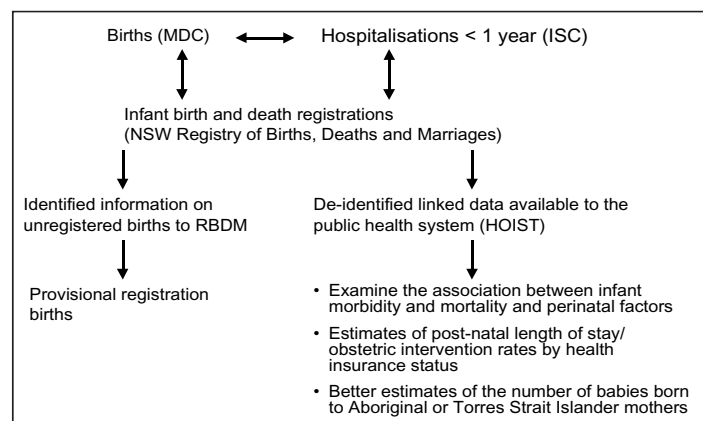


Figure 4 Linkage of the NSW Midwives Data Collection, the Inpatient Statistics Collection and the Registry of Births, Deaths and Marriages Birth and Death registration data for children up to one year of age