

The use of record linkage in epidemiological research into maternal and child health: Introduction and overview of the maternal child health research database (MCHRDB)

Gail Reading, Peter Cosgrove and Hoan Nguyen (Presenters)

Telethon Institute for Child Health Research, Perth, Western Australia

The Maternal and Child Health Research Database (MCHRDB) was established in the mid 1980's. The database is a population-based set of data of all births registered in WA from 1980 – 1999. The midwives' notifications of births are linked to Registrar General birth registrations creating a composite birth record. These records are then linked to hospital morbidity data, mortality data and census data. Birth records are also linked to the cerebral palsy register, the birth defects registry, the disability services commission database and other study cohorts as required. In addition, siblings have been linked producing a valuable data resource for research into family health and future intergenerational studies.

Over the last 15 years the database has proved to be an effective resource for maternal and child health research. In the area of policy and planning, the Department of Health of WA have recently funded a project at the Telethon Institute for Child Health Research called the Collaboration for Applied Research and Evaluation (CARE). The aim of this project is to facilitate partnerships between researchers and policy makers undertaking research and evaluation projects and to assist agencies to translate research of various forms into system and program level policy. The MCHRDB is one of the main data sources for this new project.

The presentation will start with an overview of the Maternal and Child Health database and the CARE initiative mentioned above. A 'snapshot' of a number of different research projects that have used data from the database will be presented including examples of mapping to illustrate the geographic distribution of study cohorts and/or populations.

The Maternal and Child Health Research database is located at the Telethon Institute for Child Health Research in Western Australia (WA). The history of the database starts back in 1980 with the establishment of the Midwives Register and Birth Defects Registry in WA. Following that, in 1984 and 1988 data from the midwives register, registrar general births and deaths and morbidity data were used in two major studies. The database has then evolved into an ongoing concern funded primarily by NHMRC project and program grants. It remains an important resource for research both within the Institute and in the wider research community.

The purpose of the database is to provide linked child and maternal health data to support health research, health planning and policy development. The database is used by researchers working under the current program grant and other researchers and students both inside and outside the Institute.

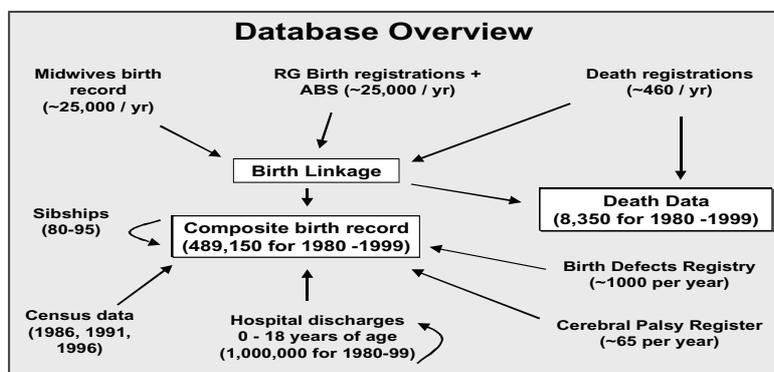


Figure 1

This is an overview of the database. The core of the database is the births table, which holds all WA births from 1980 to 1999. This table is generated by linking midwives data to registrar general births and death data together with some additional demographic data from the ABS. This linkage is done on an annual basis and the births data is then linked to census data, morbidity data and the Birth Defects and Cerebral Palsy registries. Birth records are also linked to themselves to form sibships. Death data is linked on an ongoing basis for all deaths up to the age of 18 for children born after 1980.

The data is then also linked to a variety of other study cohorts as "once-off" linkages to allow researchers working on these other studies to tap into the data in the database.

Finally, I was going to say something more about the "Collaboration for Applied Research and Evaluation" (CARE) but in the end there wasn't enough time to talk about it in detail so I'll just briefly say that CARE is an initiative that has recently started at the Institute for the purpose of facilitating partnerships between researchers and health research and the policy development and service delivery arms of our health system. For anyone who wants to know more details about this initia-

tive they can speak to me afterwards. I did bring across some printed material about it.

Examples of use

I am going to talk briefly about three different areas of research being performed at our Institute on data acquired from our Maternal and Child Health Research Database (MCHRDB).

The first one is 'The Study Of Birth Defects Diagnosed in Assisted Conception infants by 1 year of age' by Michele Hansen, Dr Jenny Kurinzczuk, Dr Carol Bower and Sandra Webb.

A paper on this research was published on 7 March 2002 in the New England Journal of Medicine.

The study involved record linkage between four Western Australian (WA) population registers for births occurring between 1993 and 1997. The Reproductive Technology Register (RTR) is a statutory data collection of all treatment cycle data for women undergoing assisted conception treatment in WA. The Midwives Notification System (MNS) collects information on all infants delivered in WA at 20 weeks of gestation or later. Our Maternal and Child Health Research Database (MCHRDB) and The Birth Defects Registry (BDR). The Birth Defects Registry collects information on birth defects occurring in live births and stillbirths in WA and in pregnancies terminated under 20 weeks because of foetal malformation.

The aim of the record linkage study was to obtain birth outcome information for all assisted conception infants and a sample of naturally conceived infants and to identify any of these infants who were diagnosed with a major birth defect by one year of age.

Children born following the use of two different assisted conception techniques were sourced from the Reproductive Technology Register: There were 301 children conceived by Intracytoplasmic Sperm Injection (ICSI) and 837 children conceived by standard In Vitro Fertilisation (IVF). Following exclusion of all assisted conception births it was possible to obtain a random sample of 4000 naturally conceived births from the Midwives Notification of Birth System.

The prevalence of major birth defects diagnosed by one year of age was compared between all three population groups.

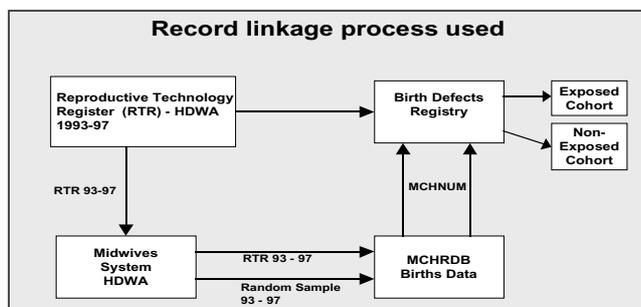


Figure 2

Figure 2 shows a diagram of the record linkage performed.

Data from the RTR was used to identify all pregnancies of at least 20 weeks gestation and all terminations of pregnancy because of foetal abnormality (regardless of length of gestation) resulting from ICSI or standard IVF treatment.

A random sample of 4000 infants born in WA was selected, after exclusion of the infants conceived with assisted reproductive technology, from the MNS.

We then linked our MCHRDB to the 4000 records from the MNS to supply the MCHNUM, which acted as the link identifier to the BDR.

When linkage was complete, birth records were available for all infants in the study and records of birth defects were available for those for whom a link was found within the BDR.

The results were as follows:

- 8.6% of ICSI infants, 9% of IVF infants, and 4.2% of naturally conceived infants were diagnosed with a major birth defect by one year of age.
- Therefore, infants conceived with ICSI or IVF have twice the risk of a major birth defect as naturally conceived infants. ($P < 0.001$)

The second research example is that of Maxine Croft's PhD studies which have been focused on setting up a database of Sibships to act as a tool to aid research to be done on WA families rather than an individual child. Maxine used probabilistic record linkage between all birth records from 1980 to 1992 to construct a unique maternal identifier (the SIBNUM) and thus, built the Sibships database. Each sibship record holds this unique SIBNUM together with the child's unique MCHNUM.

The linked records provide a longitudinal record of each woman's reproductive life. There were a total of 181,133 women. We have since done further linkage up to and including children born in 1995.

Future research teams may go on to use this database as a sampling frame for the study of "healthy" (or "unhealthy") families. It should be emphasized here that this database is limited to information from statutory health collections and registries. In particular, it does not include any information relating to: general practitioner, outpatient or emergency department visits, pharmaceutical or allied health care usage.

Two parts of Maxine's PhD research has been her interest in families which have had a child die and in particular a SIDS death and also any children having a birth defect.

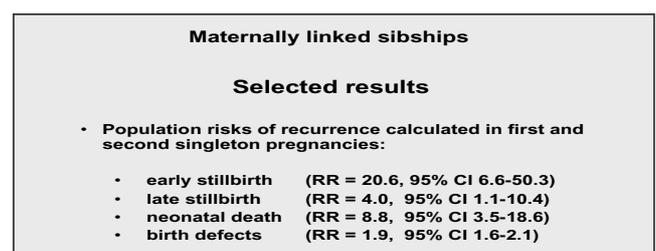


Figure 3

Figure 3 shows results of Maxine’s research on the risk of a recurring childhood death and also the risk of a second child being born with a birth defect between two reference groups of WA women:

- The first group were women whose first singleton birth resulted in one of these outcomes. The second group were women whose first singleton infant was alive at 1 year of age.
- An early stillbirth is one, which occurs between 20 and 27 weeks gestation; a late stillbirth is one which occurs after 27 weeks gestation.
- If a woman had experienced an early stillbirth in her 1st pregnancy she was 20.6 times more likely to have another early stillbirth. (Absolute Rate 3.7%)
- Women whose 1st pregnancy resulted in late stillbirth were 4 times more likely to have another late stillbirth in their 2nd pregnancy (Absolute Rate 1.3%)
- Women whose 1st pregnancy resulted in a Neonatal death – death in first 28 days of life – were 8.8 times more likely to have a 2nd pregnancy resulting in another Neonatal death. (Absolute Rate 2.8%)
- Women whose 1st pregnancy resulted in the baby having a birth defect were 1.9 times more likely to have their second pregnancy result in a birth defect.

The third and last research example is that of another PhD student, Jane Freemantle.

Jane’s primary aim is to determine indicators that better predict the health of Indigenous Western Australian infants and children.

Jane has performed extensive interrogation of her data generated through record linkage of births, morbidity and mortality data from our MCHRDB.

The aims of this part of her PhD research are:

- To determine trends in infant and child mortality for birth cohorts from 1980 to 1997
- To compare Indigenous and non-Indigenous populations
- To determine all-cause and cause specific infant childhood mortality rates.

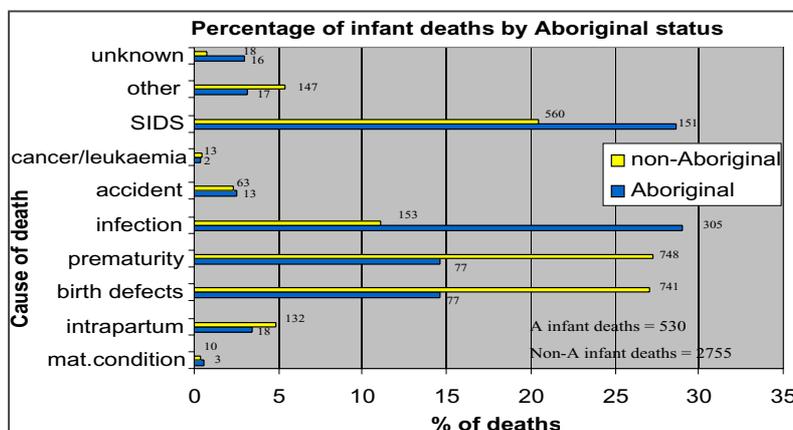


Figure 4

Figure 4 shows the causes of death for infants born in WA between 1980–1997 and who died between 1980–1998. The chart identifies the percentage of total deaths attributable to the various causes of death.

Note that for Aboriginal infants the main causes of deaths are SIDS and infections, whereas for non-Aboriginal infants the main causes are birth defects and prematurity (or sequelae thereof).

The numbers represent the number of deaths.

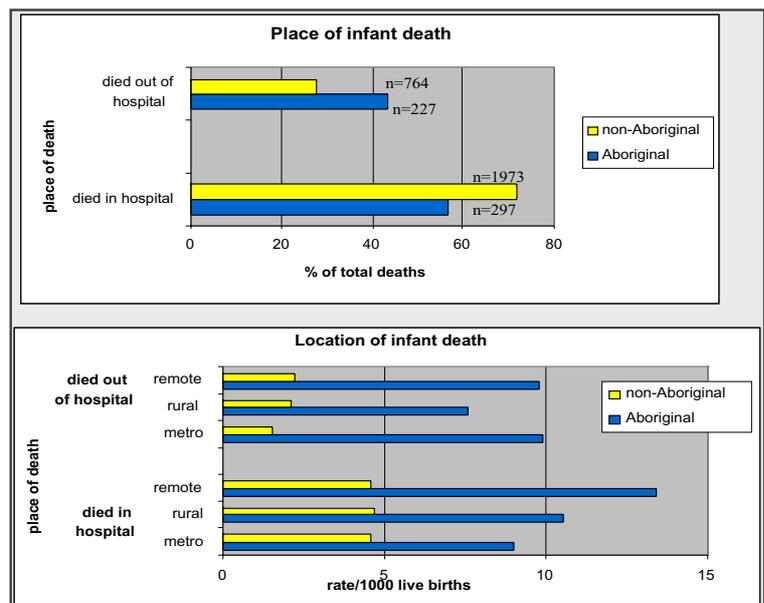


Figure 5

The top chart in figure 5 identifies the place of death as a % of total deaths (in/out of hospital) and location of death (by postcode and geographic area) for both Aboriginal and non-Aboriginal children.

57% of Aboriginal infants died in hospital, with 43% dying out of hospital. This is compared with 72% of non-Aboriginal infants who died in hospital, and 28% died out of hospital. This information can then be analysed to determine the specific causes of death associated with the place of death. These data have been further interrogated to identify the location of death for children who died out of hospital and the residence at time of death for those infants who died in hospital.

The lower chart shows the infant mortality rate per 1000 live births for the place of death by geographic location of death. For infants dying in hospital the geographic residence at time of death is identified. For those who die out of hospital the location of death is identified.

Note the rate of infant mortality of Aboriginal children who were living in the remote area who

died out of hospital was 4.5 times the infant mortality rate of non-Aboriginal infants living in the same area. The infant mortality rate for those who died in hospital was 3 times greater in Aboriginal infants living in the remote area than their non-Aboriginal peers.

For deaths that occurred out of hospital, the infant mortality rate of Aboriginal infants was similar for those who lived in the metro and remote areas. For non-Aboriginal infants, deaths occurred 1.5 times more frequently in infants living in the remote area when compared with the metro area.

These data have been further analysed to identify the cause of death and location of death by Aboriginal status.

Address matching and mapping

I'd like to take the last 5 minutes of our talk to discuss address matching and mapping that has been done at the Institute.

We have location variables for each study record sampled in our database. Enabling us to define the geosocial background of a mother and child.

With the product called CDATE from ABS, we can access the detailed demographic data such as population, education, accommodation, transport and income which were included on the Australian Census forms. While these characteristics do not describe the exact household income of an individual, they can display an accurate estimate of these characteristics by the collection district (CD), the smallest geographic unit for which census data are released.

For this reason we decided to link the addresses to the ABS CD.

Furthermore, with the CDATE add-on datapaks (also from ABS), we can access much more population Census information such as Socio Economic Indexes For Areas, Indigenous Profile and Usual Residents Profile.

How do we do address matching?

We create The Address Range File (ARF) and applications. ARF is a file that holds all streets segments covering metro, rural and remote areas of WA. Its records have nine fields. First four fields are assigned for street numbers that show from low to high numbers on odd side and even side of the street. The last five fields contain street name, street suffix, suburb, postcode and CD number.

We select tools for manually checking unlinked records.

And then we cross check with other linkage projects.

With permission from the Valuer General's Office of Western Australia, the Institute have used their Geography File.

What is the geography file?

- From the Valuer General's Office (VGO)
- A master file of approximately 1,000,000 Western Australian land parcels.

- Each parcel is identified by an address or lot number.
- Each parcel has an associated latitude and longitude.

We added the CD numbers from CDATE for every record in the VGO file and then created our Address Range File, which covers the whole state.

There are a number of software programs available for address matching, however, the Institute has built its own application programs. They are designed to perform matching. They match an address in the target table with an address in the Address Range File. If the matching is successful, a data relationship between the two addresses is established and a CD code is transferred from the Address Range File to the target table.

These programs can run either in automatic mode or interactive mode. In the interactive mode, the users have a chance to change records, to pick the best match or to force a record to match. In most cases, the automatic mode is set for the first run then the program is re-run in interactive mode to check and correct the unmatched records.

At the end of each run, the unmatched records are given the error codes that indicate which problem occurred during the matching. These unmatched records are then manually checked.

After the address matching is completed, we can use the data for mapping.

To conclude my talk I'd like to show you a couple of examples of maps to support research projects of Jane Freemantle, one of our PhD candidates.

This project is to develop indices that better predict indigenous infant & child health, to analyse cause of deaths and be able to geographically describe the rates of all-cause mortality that can be pictorially displayed using Geographic Information Systems (GIS). These data can inform policy makers and service providers of where mortality is occurring by cause of death, location of death and age of death.

These are maps of "All Cause Rate of Death for Children born in WA from 1985 to 1998" in the remote and rural areas and in the Perth metropolitan area.

The maps display two variables:

Percentage of Children Population – using the colour gradient, dark red indicates high percentage and grey means low. The map shows more children live in metro than rural and remote areas.

Rates of Death divided into 2 groups: less than one year old and from 1 to 18 years of age. Each group has 2 sub-groups. One for non-Aboriginal and one for Aboriginal.

The blue colour shows the rate for less than 1-year-old non-Aboriginal children and yellow shows the rate for less than 1-year-old Aboriginal children.

Wherever we look, we'll see that the rate of death for the infant group is easily visible and in this group the rate for the Aboriginal children is much higher than non-Aboriginal.

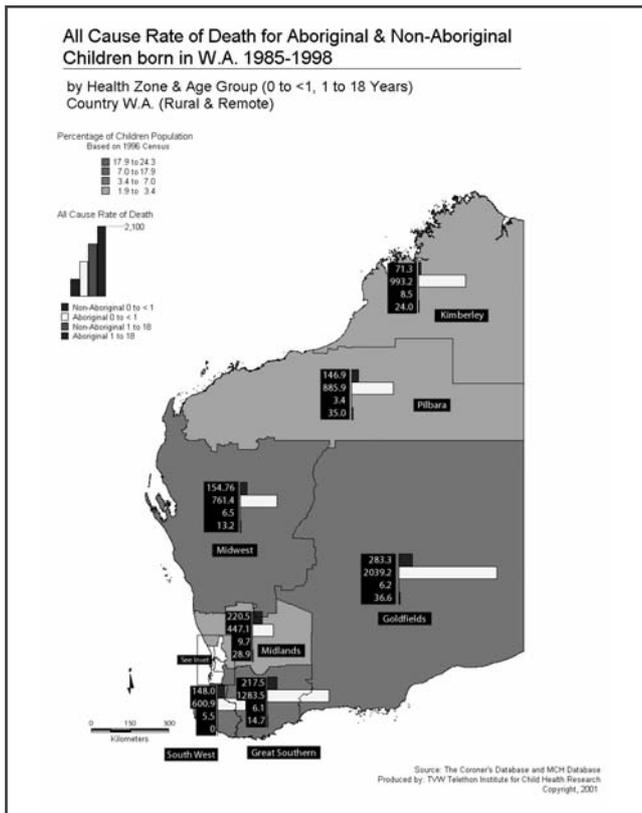


Figure 6

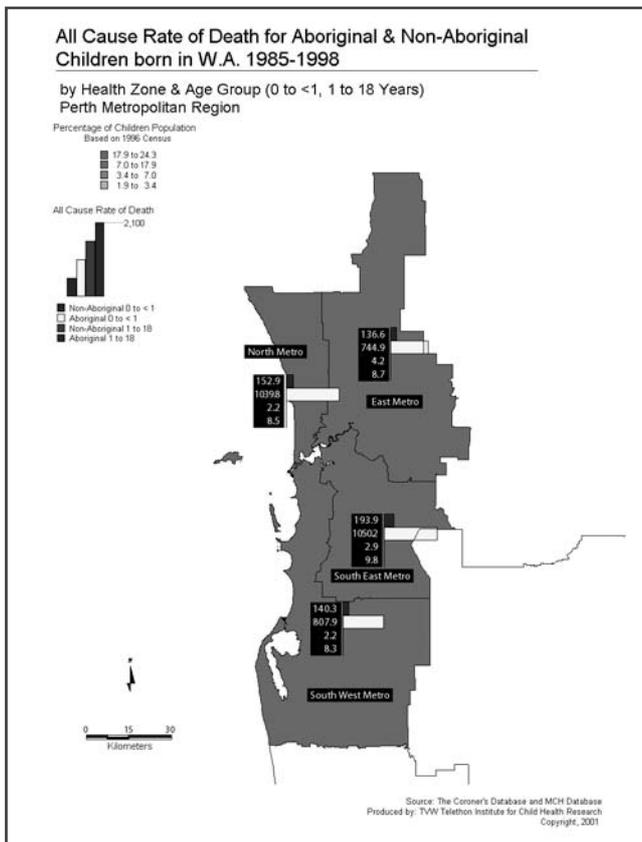


Figure 7

Questions

Dr Eshpeter, Population Health Division, DoHA: Have you conducted validation of the database race variable against coroner's reports?

Ms Reading: We haven't conducted a validation of the whole database mother's race variable against coroner's reports however validation of the race variable on the database has been performed over the years by individual researchers. In 1992 the indigenous status of the mother, by which we classify the race of the baby, was verified in a validation study involving comparing the data for a sample of cases from the Midwives' Notification System with the information recorded on the hospital medical record. The mother's race variable was found to have been recorded incorrectly in less than 2% of the Midwives' notification of case attended forms. Jane Freemantle did extensive validation of all her deaths in her PhD studies by viewing coroner's reports. The child's race was not always recorded on coroner's reports, however, whatever information on the child was reported which eluded to the indigenous/non-indigenous status of the child was used as confirmation of race. Jane found the race variable recorded on our MCHRDB at birth for her population to be 99% accurate.