Report on the use of linked data relating to Aboriginal and Torres Strait Islander people

Report on the use of linked data relating to Aboriginal and Torres Strait Islander people

A complete list of the Institute's publications is available from the Institute's website <www.aihw.gov.au>.

**Please note that there is the potential for minor revisions of data in this report.
Please check the online version at <www.aihw.gov.au> for any amendments.**

Contents

Acknowledgments

Abbreviations

ABC	Aboriginal Birth Cohort study
ABS	Australian Bureau of Statistics
ACCHO	Aboriginal Community Controlled Health Organisation
ACT	Australian Capital Territory
AHLO	Aboriginal Health Liaison Officers
AIHW	Australian Institute of Health and Welfare
ALRI	acute lower respiratory infection
AMI	acute myocardial infarction
ANZDATA	Australian and New Zealand Dialysis and Transplant Registry
APDC	Admitted Patients Data Collection
ARC	Australian Research Council
ASR	Age–standardised rates
ASV	statistical Aboriginal variable
BBV	blood borne virus
CF	case fatality
CHC	Child Health Check
CHCI	Child Health Check Initiative
CHD	coronary heart disease
CHeReL	Centre for Health Record Linkage
CMI	Client Master Index
COAG	Council of Australian Governments
CRIAH	Coalition for Research to Improve Aboriginal Health
DALYs	Disability Adjusted Life Years
DSC	Western Australian Disabilities Services Commission
DoHA	Australian Department of Health and Ageing
DoHWA	Department of Health Western Australia
HMDC	Hospital Morbidity Data Collection

HRN	Hospital Registration Number
ID	intellectual disability
IDEA	Intellectual Disability Exploring Answers Dataset
IHD	ischaemic heart disease
MCCD	Medical Certificate of Causes of Death
MCHRDB	Maternal and Child Health Research Database
MDC	Midwives Data Collection
MHIS	Mental Health Information System
NDI	National Death Index
NDD	New South Wales Notifiable Diseases Database
NIDD	National Intellectual Disability Database
NIRA	National Indigenous Reform Agreement
NMD	ABS and AIHW National Mortality Database
NOCS	Queensland Health Notifiable Conditions database
NTER	Northern Territory Emergency Response
OIMS	Offender Integrated Management Services
PAARS	Perth Aboriginal Atherosclerosis Risk Study
PAF	population attributable fraction
PDC	Perinatal Data Collection
PHDAS	Pharmaceutical Drugs of Addiction System
PHRN	Population Health Research Network
PI	performance indicators
QHAPDC	Queensland Hospital Administration Patient Data Collection
RBDM	Registry of Births, Deaths and Marriages
SA-NT DataLink	South Australia-Northern Territory Data Linkage Unit
SEARCH	Study of Environment on Aboriginal Resilience and Child Health
STI	sexually transmitted infection
TICHR	Telethon Institute for Child Health Research
VIVAS	Vaccination Information/Vaccination Administration System

WAACHS   Western Australian Aboriginal Child Health Survey

WADLS   West Australian Data Linkage System

Summary

This review presents various Australian studies and analyses based on linked data relating to Aboriginal and Torres Strait Islander people. We reviewed whether the purpose of each study was:

- to enhance the value of Indigenous status information across datasets, or
- to enhance data other than Indigenous status for the purpose of undertaking research that cannot be undertaken using data from only one source.

The review also examined:

- the core themes of the studies
- the datasets used
- the data linkage methodology used, or intended to be used, in the studies
- data quality issues encountered or anticipated
- problems with the quality of the Indigenous status variable on the various datasets
- the method of analysis
- what algorithms or methods were used or are planned to be used in deriving Indigenous status if Indigenous status was missing or was inconsistent across input datasets
- whether researchers explored, or intend to explore, the impact of various algorithms or methods for deriving Indigenous status on the estimated outcome measures.

We found that an emphasis on the use of data linkage to improve accuracy of Indigenous status in datasets has only recently emerged. Moreover:

- the use of survey data as a 'gold standard' for Indigenous status information when linking with other datasets has great potential
- to date, data linkage has rarely been used for exploring topics related to Council of Australian Government's (COAG) Closing the Gap Performance Indicators
- there has been little consistency in how Indigenous data linkage is done or evaluated
- there is a small evidence base for the relative merits of different approaches to identifying Aboriginal and Torres Strait Islander people through data linkage
- the existing evidence base may not be universally applicable across all Australian jurisdictions and datasets.

The amount of work using linked data about Aboriginal and Torres Strait Islander people varies greatly between different states and territories. This partly reflects how long data linkage nodes have existed in each jurisdiction, but differences in legislative regimes between states and territories may also be partly responsible. Only the AIHW and ABS are currently linking data related to Aboriginal and Torres Strait Islander people on a national scale. This report should be read in conjunction with the *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People* and its online attachment, *Thematic list of projects using linked data relating to Aboriginal and Torres Strait Islander people* (henceforth referred to as the *Thematic List*).

1   Introduction

1.1 About this review

Schedule F of the National Indigenous Reform Agreement (NIRA) calls for: 'an examination of current and planned data linkage work (Commonwealth and State and Territory) relevant to Indigenous identification' as part of the development of the *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People* (AIHW & ABS 2012; COAG 2008). Schedule F of the NIRA further states:

> 'This project will need to take into account existing and planned data linkage work across all jurisdictions (including the methods currently being used, and the results from any completed work). This will help inform the development of national guidelines' (COAG 2008).

This report is the first of two documents:

1. *Report on the use of linked data relating to Aboriginal and Torres Strait Islander people*
2. *Thematic list of projects using linked data relating to Aboriginal and Torres Strait Islander people*.

These reports are intended to fulfil the AIHW's agreed data quality improvement obligation described in the NIRA Schedule F. In describing and comparing data linkage practices to date, these documents provide an evidence base for the *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People* (henceforth referred to as the *Guidelines*) (AIHW & ABS 2012). Readers of the *Guidelines* can draw upon this resource to find data linkage activities that may have used similar datasets to those they wish to use, or encountered similar problems that they may face. In reviewing data linkage work related to Aboriginal and Torres Strait Islander people, we have highlighted gaps in the evidence base that future data linkage work must fill. We also draw attention to problems with linking data about Aboriginal and Torres Strait Islander people that future data linkage work must take into account, and attempt to quantify and resolve.

This first report describes the static body of data linkage activities related to Aboriginal and Torres Strait Islander people, along with current and planned activities as of 2012. It is followed by the second report that will assist readers by presenting a list of projects using data linkage, organised by subject matter, and described by title, years of data covered, datasets used, and linkage and Indigenous identification algorithms used where available.

1.2 Terminology

The Australian Institute of Health and Welfare refers to the Indigenous people of Australia as Aboriginal and Torres Strait Islander people. However, Aboriginal and Torres Strait Islander people are referred to in various ways in the data linkage literature reviewed in this report. Throughout this report, the practice of each study or report in using the terms 'Indigenous', 'Aboriginal' or 'Aboriginal and Torres Strait Islander' has been retained when describing each study, as imposing AIHW practice may risk misrepresenting the reviewed studies.

Other terms used in these reports, unless explicitly defined, are defined in the glossary and abbreviations section of the *Guidelines* (AIHW & ABS 2012).

1.3 About this report

The literature review found that data relating to Aboriginal and Torres Strait Islander people have been linked for two main purposes:

- Indigenous status data in a Dataset have been improved or evaluated by linking a Dataset to other datasets as additional sources of Indigenous status data, or linking it to another Dataset that is regarded as having Indigenous status data of higher quality, or

- Data about different variables have been linked across different datasets to explore relationships that could not be investigated using one Dataset alone. In this way, linking adds value, since new variables can be considered which were not originally available in the first Dataset. Indigenous status may be one of the new variables if it was not present in the first Dataset. However, new insights into the health and wellbeing of Aboriginal and Torres Strait Islander people can be gained by linking to other variables as well.

Uses of linked data relating to Aboriginal and Torres Strait Islander people can be broadly segregated into three kinds:

- Some examples assess a whole population, or a subset of it, and are used to compare people with different Indigenous status in the selected population for particular variables

- Other studies or reports are principally interested in a variable other than Indigenous status, but will compare people of differing Indigenous status in pursuit of information about the variable of interest

- Other data linkage projects have exclusively selected Aboriginal and Torres Strait Islander people to study a particular condition, and generally use data linkage to add extra information about Aboriginal and Torres Strait Islander people from extra datasets. Often these additional datasets contain valuable clinical information.

Data linkage related to Aboriginal and Torres Strait Islander people is done for different purposes, and is done in a variety of ways to achieve these purposes. This report of data linkage activity is structured accordingly to assist researchers and analysts to incorporate elements of best practice project design into their own data linkage activities. Different data linkage activities are summarised in relation to:

- their objectives and design

- their subject matter or key themes

- the data sources used

- the methods used in linking the data

- problems encountered in the data linkage and their management

- how Indigenous status was determined if it was missing or was inconsistently reported across datasets.

This report concludes with a discussion of several aspects of the state of data linkage activities related to Aboriginal and Torres Strait Islander people as practised in Australia.

This report aims to flag problems with linking data in regard to Aboriginal and Torres Strait Islander people that data linkers and analysts should be aware of, and indicate areas where the evidence base needs to be developed. In drawing attention to these issues, readers will be encouraged to analyse and quantify these issues, and in doing so, contribute to the growing evidence base for best practice in linking data about Aboriginal and Torres Strait Islander people.

2 Linkage for improving Indigenous identification

Data linkage has been used to assess or enhance the quality of Indigenous identification in administrative datasets. A relatively large amount of detail about the linkage process tends to be reported when data linkage is used in this way. In the projects described below, improved Indigenous identification through data linkage is treated as a complete end in itself, rather than a means to an end. The ultimate aim of enhancing the Indigenous status variable in datasets in these activities is to either facilitate their use to more accurately compare the characteristics of Indigenous and non-Indigenous people, or else compare other characteristics using improved Indigenous status as one of a number of secondary variables.

2.1 Nationally based linkage projects

Indigenous identification on deaths datasets can be improved by linking alternative datasets that contain information about Indigenous deaths and Indigenous identification. The AIHW Indigenous Mortality Data Enhancement Project (AIHW 2012) and the ABS Indigenous Mortality Quality Study (ABS 2008a) attempt to assess the level of under-reporting of Indigenous deaths by linking deaths data to either health datasets or to Census data respectively. These national projects aim to inform the improvement of reporting of Indigenous status of deceased people. Each study found that Indigenous deaths data can be improved by linkage, but that considerable obstacles exist to the ongoing use of data linkage for this purpose.

The AIHW Indigenous Mortality Data Enhancement Project

The AIHW has conducted a feasibility study for its Indigenous Mortality Data Enhancement Project (AIHW 2012), which investigated whether it is possible to enhance Indigenous death data through the joint use of several additional alternative data sources that contain information on Indigenous deaths and Indigenous status (see *Thematic List*, page 18). By linking these data sources, deaths of Aboriginal and Torres Strait Islander people can be detected when they are not already identified as Indigenous in the ABS and AIHW National Mortality Database (NMD). The project linked the following additional datasets held by the AIHW to deaths registered from 2001–2006 in the NMD and the National Death Index (NDI): the Residential Aged Care Dataset, National Hospital Morbidity data, and the National Perinatal Data Collection. Small numbers of deaths can cause erratic death rates over time, so death rates were smoothed using partially monotonic penalised regression splines. The enhanced data are expected to enable more accurate estimates of Indigenous mortality and life expectancy to be made.

The linkage was conducted in three stages. Records for all deaths (including non-Indigenous deaths) in 2001–2006 from the NMD, which does not have names, and from the NDI were linked to create a mortality extract dataset. Pass structures for linking the mortality extract dataset to the other datasets were then constructed, consisting of a series of linkages using different linking variables within subsets of records defined by different blocking variables. The mortality extract dataset and Indigenous death records from the additional datasets were then linked. Each additional dataset was linked in separately, since they did not contain

the same linkage variables (e.g. the aged care dataset contained names, while the hospital and perinatal datasets did not).

By linking each dataset separately in this way, it was possible to use different linkage methods. Probabilistic linkage was used for the aged care dataset (with weights partly based on name similarity), so that records with names with slightly different spelling could be linked. Deterministic linkage was then used for the perinatal and hospital datasets (which lacked names), to provide greater specificity. In total, five separate pass structures were used to link the data, with different linkage variables used at each step. The final linked dataset was then used to calculate estimates of life expectancy for Aboriginal and Torres Strait Islander Australians for 2001–2006.

Linkage quality for the perinatal–mortality extract linkage and hospital–mortality extract linkage was assessed by validating each pass structure once they were used for linking the candidate datasets. This was done by using the same pass structures to link aged care and mortality extract records. The linkage of these additional datasets was then used as the gold standard against which the perinatal and hospital linkage processes were validated. Specificity and sensitivity quality measures were calculated and then plotted to a Receiver Operating Characteristic curve.

Rules were developed to distinguish whether two records were similar enough or too dissimilar to be considered a match. Links were then further divided into excellent, good quality or acceptable links, with most classified as either excellent or good quality matches. Links that were delineated this way were then quantified, as was the proportion of records that linked. Where there were proportions of records that did not link, the characteristics of the linking variables were examined for possible causes. This examination revealed a lack of precision in names in some datasets and a frequency of dummy 'date of birth' fields (e.g. proportionally high numbers of records with 1 July as date of birth, presumably where year of birth may have been surmised but precise day of birth was unknown).

The 'ever-Indigenous' algorithm was used to derive Indigenous status. While the potential for over-count due to single instances of people being misidentified as Indigenous was acknowledged, the authors did not feel that this would be substantial, since only records of deaths and not of other interactions recorded by the other datasets were used. In this project, it is relatively unlikely that multiple entries for individuals would bias some individuals towards being misidentified as Indigenous, since death, unlike hospital visits, can only happen once.

The linking of these datasets to the NMD added 1,081 or 10% more death records to the number of Indigenous deaths registered by the state and territory registrars of births, deaths and marriages for the period 2001–2006. As expected, the additions affected deaths of those aged under 5 or 65 and over, more than deaths in other age ranges. They also affected the south-eastern States more than the Northern Territory, South Australia and Western Australia, which was attributed to better Indigenous identification in the latter three jurisdictions. Life expectancy estimates from the linked dataset were lower than those derived from unlinked data by 1.7 years for males and 2.0 years for females respectively.

The authors stressed that not all deaths in the mortality extract dataset would be covered by the other datasets. They also emphasise that, if it can be assumed that additional Indigenous deaths lowers life expectancy estimates, the resulting under-count of Indigenous deaths would mean that their revised life tables would represent the upper limits of Indigenous life expectancy. They particularly note that deaths occurring in the homes of Aboriginal and

Torres Strait Islander people are not recorded as Indigenous in Registry of Births, Deaths and Marriages (RBDM) records. They also point out that deaths of Indigenous babies with non-Indigenous mothers in perinatal data would not be covered if they were not elsewhere recorded as Indigenous. Nonetheless, the new dataset created from the data linkage produced plausible life expectancy estimates, which are comparable (though slightly lower) to those produced by the ABS for the 2005–2007 period. The ABS reported these estimates in an appendix to their life tables publication in May 2009 for comparison purposes (ABS 2009). Single–year life expectancy estimates were also made for the study period, but these fluctuated greatly between years for both sexes, even for Australia–wide estimates. The authors therefore recommend using rolling 3–year estimates instead of single–year estimates.

Ethics approval has been granted to the AIHW to extend the Indigenous Mortality Data Enhancement Project from 2006 onwards, as a continuation of the feasibility study conducted on enhanced deaths data from 2001–2006 (AIHW 2012). As with the feasibility study, the NDI and NMD will be linked, followed by a further linkage to admitted patients data, residential aged care data and perinatal data. As for the 2001–2006 linkage, the intent of this linkage is to use the latter datasets to identify Indigenous deaths that are not recorded as Indigenous in the NDI or NMD. The same linkage variables and methodology will be used for the post-2006 linkage as was used for the 2001–2006 linkage. This enhanced mortality database will be used to calculate life expectancy for Aboriginal and Torres Strait Islander Australians.

ABS Census Data Enhancement Indigenous Mortality Quality Study

The Census Data Enhancement Indigenous Mortality Quality Study (ABS 2008a) linked 2006 Census records with death registration records to examine differences in the reporting of Indigenous status across the two datasets (see *Thematic List*, page 18). The aims of the project were to: assess the under-coverage of Indigenous deaths in death registration records; identify factors that may be contributing to under-coverage of Indigenous deaths in death registrations; and assess the feasibility of calculating and applying adjustment factors to improve estimates of Indigenous mortality.

Names and addresses are temporarily held by the ABS after a Census while the Census is being processed and then destroyed. The period before destruction provides a window for the ABS to link Census data while personal information is still available. The timing of the study is dependent on the availability of name and address variables. Records from the Census conducted on 8 August 2006, were linked with records of deaths registered between 9 August 2006 and 30 June 2007, except for Victoria where death registrations were only available up to mid-March 2007 (Deaths registered in Victoria were available only to mid-March 2007 (ABS 2010)). Not all deaths that occurred between 9 August 2006 and 30 June 2007 were registered by the time of the study (due to a lag between occurrence and registration of a death), i.e. the study did not have full coverage of the deaths within scope (ABS 2010c). Results from the Indigenous Mortality Quality Study were used to create factors to adjust death registrations and improve estimates of Aboriginal and Torres Strait Islander life expectancy for the study period.

The data linking methodology used for the ABS Indigenous Mortality Quality Study, and the assessments of data quality of the variables and quality of the linking that accompanied the study, have been reported in detail (ABS 2010c). They are summarised here because they provide a good indication of how quality of linkage can be systematically evaluated and

compared. Many of the linkage measures used are quite comprehensive and would be very good candidates for universal application where nationally consistent measures are practical.

As no unique identifier was available for linkage, the best available method was probabilistic linkage, as it accommodates variability in the quality of some linking variables for Aboriginal and Torres Strait Islander people. The ABS modified the data linkage program Febrl 0.3 to allow for the size of the Dataset, notably speed of access, and to make clerical review more feasible. Functional separation between data linkers and analysts was enforced.

One–to–one assignment of record pairs was incorporated into the algorithms used for linkage, that is, the algorithms were designed and run in such a way that a record from one file could not be linked to more than one record in the other file.

Names on the Census Dataset were of poorer quality than those on the deaths Dataset. Automated name repair of the Census data still left 4 million names unrepaired. Specific sub-groups of these (with particular attention paid to Aboriginal and Torres Strait Islander names) were selected for clerical repair.

The number and proportion of missing responses for each linking variable disaggregated by Indigenous status was calculated (excluding invalid and incomplete responses). Proportions of missing responses for linking variables were higher for most variables of people with unstated Indigenous status than Aboriginal and Torres Strait Islander people, and were consistently higher for Aboriginal and Torres Strait Islander people than for non-Indigenous people, across most linking variables, in both datasets. Mesh block, a variable created from address or place of usual residence information, was found to be the worst performing variable. First name, surname, and date of birth in the deaths Dataset were found to have very few missing responses. Missing response levels were higher for Aboriginal and Torres Strait Islander people than for non-Indigenous people for the address component variables in the deaths Dataset than in the Census Dataset. These differences in the quality of address components made it harder to successfully link Aboriginal and Torres Strait Islander people than non-Indigenous people between these two datasets.

Addresses on the deaths Dataset were thought to be affected by a tendency to specify a nursing home and its postcode, rather than street number and name. Parsing (breaking up) the address string into components could also have resulted in a loss of information. Using individual components of these strings, however, meant that the linkage process was not disrupted by an error in one of the components.

Of the total registered deaths, 98,898 (92%) were linked to a Census record. However, a much lower linkage rate was achieved for Indigenous deaths. More than one quarter of all Indigenous death registrations (26% or 473) could not be linked to a Census record.

While only 1.42% of non-Indigenous Census records had missing values for two or more linking variables, 7.33% of the Indigenous records and 7.99% of records where Indigenous status was not stated had missing values for two or more linking variables.

Linkage quality was assessed by:

- examining population characteristics of linked and unlinked death records
- finding out the reasons for unlinked death records
- estimating the number of false links
- estimating match-link rate and link accuracy
- observing the effect of false links on analysis.

Aboriginal and Torres Strait Islander people, males and people under the age of 60 were over-represented in the unlinked death records. As such, the linked file was not representative of all death records.

Several possible reasons for a death record not linking to the Census data were explored, but only two were surmised to make a substantial contribution to the numbers of unlinked death records: Census under-count, and missing responses in one or more of the linking variables. Census under-count rates had been estimated during the Post Enumeration Survey: when these were applied to the unlinked death records, it was estimated that 3,747 out of 8,047 unlinked death records must not have linked because these individuals were missed in the Census count, that is, these individuals never completed a Census form. Missing responses for each linking variable had been summed prior to linkage, and this number was deemed large enough to be able to explain the remainder of failed links. It would be useful to estimate the number of these unlinked records which had missing linking variables, to assist with determining the reasons they did not link.

The Census under-count rates applied to the unlinked death records were derived from estimates of the proportion of all people (by 5-year age increments and by sex) not counted in the 2006 Census. These proportions were based on matching respondents to the 2006 Census Post Enumeration Survey with the Census during the Census processing period. From this application of the under-count rates to unlinked death records, it was estimated that slightly less than half of the unlinked death records could be for people not counted in the Census. Indigenous undercount had elsewhere been estimated at 11.5% nationally, and had been proportionally greater than for non-Indigenous people (ABS 2008b).

Wrong or out of date data in the linking variables for the Indigenous Mortality Quality Study was nominated as a possible explanation of failure to link (ABS 2010). This was not quantified, but data quality issues were surmised by a process of elimination to account for failure to link for slightly more than half the unlinked records. It is possible that name ambiguity and different addresses due to increased mobility may have contributed to failure to link, and these may be greater issues with data about Aboriginal and Torres Strait Islander people than non-Indigenous people.

The ABS repeated the Indigenous Mortality Quality Study as a statistical study for the 2011 Census using a similar methodology (ABS 2010b, see *Thematic list*, page 20–1). The Indigenous mortality project was conducted during the Census processing period following the 2011 Census, from 9 August 2011 for approximately 11 months, before variables in the Census that would be used for linkage (i.e. name and address) are destroyed. A number of other studies linking non-ABS and ABS data to the Census are being undertaken as part of the 2011 Census Data Enhancement Project. Details are available in *Census Data Enhancement Project: an update, Oct 2010* (ABS 2010b).

The 2006 Indigenous Mortality Quality Study found that more than one-quarter of Indigenous death records could not be linked to the Census. The predominant reasons proposed for not linking death records were failure to count people in the Census and the poor quality of linkage variables such as address (ABS 2008a). Moreover, the total number of unlinked death records of all people was much greater than the total combined number of linked Indigenous death records and unlinked Indigenous death records. It was also estimated from the 2006 Census Post Enumeration Survey that 11.5% of all Aboriginal and Torres Strait Islander people were not counted in the 2006 Census. The Census methodology of using a 'usual address' has been criticised as inappropriate given evidence suggesting that

some Aboriginal people are highly mobile and may not have a single address they regard as a usual place of residence (Memmott 2011; Memmott et al. 2004; Memmott et al. 2006).

The ABS has implemented procedures tailored to the enumeration of Aboriginal and Torres Strait Islander people living in discrete communities since the 1976 Census. The 2011 Census procedures build on this experience with the 2011 Census Indigenous Enumeration Strategy. Procedures have been tailored in response to the requirements of each Indigenous community, and are described elsewhere (ABS 2011b). These procedures have been designed to reduce the under-coverage of Indigenous people in the 2011 Census. The ABS believes these procedures should therefore produce better linkage results for the 2011 Indigenous Mortality Study.

For more information about the Census Indigenous Enumeration Strategy see <http://www.abs.gov.au/ausstats/abs@.nsf/lookup/2903.0Main%20Features82011> (ABS 2011a).

2.2   State–based linkage projects

Western Australia

Resolving missing Indigenous status in deaths

Draper and others (2009) sought to resolve the Indigenous status for ABS-validated deaths data for WA from 1997–2002 with missing Indigenous status (see *Thematic List*, pages 27–8). They did this by linking these death data to linked data (hospital morbidity, mental health and midwives notification data), accessed through the Western Australian Data Linkage System (WADLS). Two alternate algorithms for deriving Indigenous status were compared. In the 'majority' algorithm, an individual was classified as Indigenous if the number of times a record was coded as Indigenous was greater than the total number of times it was coded as non-Indigenous or unknown; otherwise, they were classified as non-Indigenous. In the 'ever' algorithm, an individual was classified as Indigenous if any record for them was coded as Indigenous; otherwise they were coded as non-Indigenous. Linkage was not used to assess Indigenous identification quality in death records with recorded Indigenous status, but this was only because records with missing Indigenous status were the original target of the investigation, and not because of any inherent extra difficulty with linking or analysing all of the deaths data (Draper 2010).

Indigenous status from each Dataset was treated as a separate variable to facilitate algorithm application. The 2.8% of registered deaths in the study period with no stated Indigenous status were linked with three other datasets so that the Indigenous status in the latter could be assigned to them. 4.3% of these could not be linked to any dataset. Indigenous status was resolved for all linked records or 95% depending on algorithm. Hospital morbidity data provided a link to the largest number of deaths with missing Indigenous identification. Draper and others (2009) found that missing Indigenous status could be resolved by linking with one or more of these datasets for 95% or more of these deaths, depending on the algorithm used to determine Indigenous status.

Abridged life tables for 5–year age groups were constructed to calculate life expectancies using records resolved by linkage combined with those previously identified in the deaths data. They found that the original mortality data had underestimated Indigenous mortality by 6–7.5%. There was no significant change in life expectancy for non-Indigenous people, as

the change in the proportion of non-Indigenous deaths was small, even though 88.2% (using the 'ever' algorithm) and 89.1% (using the 'majority' algorithm) of those with unstated Indigenous status were resolved as non-Indigenous. The increases in life expectancy estimates for Indigenous males and females over the study period shrank slightly when the improved Indigenous data were used to calculate life expectancy. Because of under-identification of Aboriginal and Torres Strait Islander people in the hospital morbidity data, they concluded that even the 'ever-Indigenous' algorithm may not over-count Aboriginal and Torres Strait Islander people.

As an extension of this work, the ABS, the Department of Health Western Australia (DoHWA) and the Telethon Institute for Child Health Research (TICHR) are undertaking a collaborative project that aims to explore and develop different methods for deriving Indigenous status from multiple data sources (see *Thematic List*, page 21, Getting Our Story Right). By utilising the WADLS, the project is examining the impact of these methods for deriving Indigenous status on a sample of health and educational outcomes among the Aboriginal and Torres Strait Islander population.

The project aims to compare various algorithms used for deriving Indigenous status across multiple data sources. Recommendations regarding the optimal use of existing information resources as they pertain to measuring the gap in Aboriginal and Torres Strait Islander disadvantage will be made.

Various methods of deriving Indigenous status from linked data sources will be explored and the impact of these methods examined against a selection of performance indicators from the National Indigenous Reform Agreement and National Healthcare Agreement. These will include health and educational outcomes such as mortality rates, hospitalisation rates, and school–based reading and writing scores from standardised tests.

Linked data have been drawn from 11 collections held respectively by the WA Department of Education, the WA Registrar General of Births, Deaths and Marriages, the Telethon Institute of Child Health Research and the Department of Health Western Australia.

Project outputs will include measurement of the impact of different methods/business rules for deriving Indigenous status on both the number of Aboriginal and Torres Strait Islander and non-Indigenous people on each dataset and on various outcome measures/performance indicators. In addition, the original Indigenous status flag will be compared to the various enhanced Indigenous status flags as a means of validating the reliability of the Indigenous status flag on each dataset.

Resolving missing Indigenous status in notifiable diseases data

Mak & Watkins (2008) linked records for 2004 in the Western Australian Notifiable Infectious Diseases Database that were missing Indigenous status information to WADLS linked data (see *Thematic List*, page 22). The linkage was for only the subset of the infectious diseases data for which Indigenous status was missing. However, the linkage process and its evaluation were unusually thorough.

This study was the first to use data linkage to improve Australian notifiable disease surveillance. They evaluated the use of data linkage to improve the accuracy of estimated notification rates for sexually transmitted infections (STIs) and blood borne viruses (BBVs) in Aboriginal and non-Aboriginal groups in Western Australia. Probabilistic methods were used to link with core population health datasets from WADLS. The proportion of STI and BBV notifications with missing Aboriginality data was reduced by 74%. They concluded that

data linkage can be effectively used to improve the accuracy of estimated disease notification rates.

Two alternate definitions were trialled to determine Indigenous status: a 'specific' definition and a 'sensitive' definition. In the specific approach, a patient was classified as Indigenous if they were always identified as Indigenous across all WADLS linked datasets, and as non-Indigenous if they were always identified as non-Indigenous across these datasets. In the 'sensitive' approach, a patient was classified as Indigenous if they were classified as such in one or more WADLS Dataset (i.e. 'ever Indigenous'), and as non-Indigenous if they identified as such in one or more WADLS Dataset (i.e. 'ever non-Indigenous'). Under the sensitive definition, a person could be classified as both Indigenous or non-Indigenous. By using these two approaches, the numbers of participants with inconsistent Indigenous status could be easily calculated as the difference between the sensitive and specific categories. This linkage reduced the proportion of infectious diseases records with missing Indigenous status from 26.0% to 6.8%.

As with Draper and others (2009), new rates for the overall population disaggregated by Indigenous status were calculated by adding those identified by linkage to those already recorded as Indigenous in the primary Dataset prior to linkage, and in some instances, the new rates were significantly different from the old ones. Age standardised disease notification rates were calculated for unlinked records and for alternate data using different Indigenous status derivation. 99% of the National Intellectual Disability Database (NIDD) unknowns were linked with morbidity records, while 0.02% were linked with deaths. The capacity to resolve by linkage differed by disease. Unknowns were more likely to be resolved as non-Indigenous. Only individuals whose Indigenous status was missing in NIDD were linked; linkage was not used to compare Indigenous status between NIDD and WADLS data to, for example, assess accuracy of the NIDD.

Most people who had missing Indigenous status in their infectious diseases data were reclassified as non-Indigenous. The disease rates between sensitive— and specific—defined categories differed by no more than 1 or 2 percentage points. This is to be expected; of the records successfully linked to a WADLS Dataset with Indigenous status information, 61% linked to only 1 of the 4 datasets, so the specific and sensitive definitions were identical for at least this number of patients. Most linked notifications had more than one record in the hospital morbidity Dataset (1,091), but only 49 of these had inconsistent Indigenous status.

Under-ascertainment of Aboriginality in records of cardiovascular disease in hospital morbidity and mortality data in Western Australia

Under-ascertainment of Aboriginal status in records of people who had been admitted to Western Australian public hospitals for cardiovascular disease was estimated after linking their records in the Hospital Morbidity Data Collection (HMDC) and, where applicable, Western Australia deaths data (Briffa et al. 2010, see *Thematic List*, page 5). All Western Australia residents aged 25–74 who were admitted with a principal diagnosis of cardiovascular disease during 2000–05 were identified, along with any death record for them in either the HMDC or Western Australia deaths data, and any other hospital admission record they had back to 1980. Hospital records for each episode of care in the relevant period were first linked to form person–records. From these linked person–records, the most recent separation could be identified. Then Western Australian deaths data were linked to these records (death records that could not be linked to these person–records were not analysed, even if the principal cause of death was cardiovascular disease). Hospital data

going back to 1980 were linked to them to find all their prior admissions. Four alternate algorithms for resolving Aboriginal status were compared:

i. 'Most recent admission/current admission': the person was identified as Aboriginal in their most recent admission during 2000–05 for which cardiovascular disease was the principal diagnosis. This was also called the index admission (because it was used as the baseline comparator for the other algorithms) or most conservative estimate.

ii. The person was identified as Aboriginal in either their most recent admission or their subsequent death record (i.e. their record in the deaths data).

iii. 'Majority': the person was identified as Aboriginal on at least 50% of their HMDC records in the previous 20 years or their subsequent death record.

iv. 'Ever-Indigenous': the person was identified as Aboriginal on any HMDC record or their subsequent death record. This was also called the least conservative estimate.

Indigenous status as recorded in the most recent hospital separation and in the prior admissions were compared by region, age, sex and year of admission, as was Indigenous status in the deaths data. Differences in the performance of the 'most recent admission' and 'ever-Indigenous' algorithms were compared for several socio–demographic factors.

The extent to which Aboriginal and Torres Strait Islander people had been under-counted in these HMDC records was calculated. The 'most recent admission' algorithm underestimated Aboriginal status compared to each of the other three algorithms: by 2.6% for algorithm (ii); 3.5% for algorithm (iii); and 17.2% for algorithm (iv).

Algorithm (i) resolved 4.9% of the person–records as belonging to Aboriginal people. Using algorithm (ii) increased this number by 83 (2.7%) records, while algorithms (iii) and (iv) added another 78 (3.7%) and 60 (20.8%) records to the numbers of Aboriginal people respectively. Indigenous deaths in the analysed deaths data in the linked Dataset underestimated Indigenous deaths by 26.8%, based on being identified as Aboriginal in any of their records in the HMDC in 2000–05. Death records with unknown Indigenous status were mostly resolved as non-Indigenous, but were still nearly twice as likely to be Aboriginal as death records with known Indigenous status.

Under-ascertainment (as measured by comparing algorithms (iv) and (i) only) occurred disproportionately more in Aboriginal people who were over 64 or lived in the least socioeconomically disadvantaged areas, and less in Aboriginal people who lived in remote or very remote areas. The results were largely consistent with previous estimates of under-identification in WA hospitals data made by the WA Department of Health and the AIHW.

Researchers at the School of Population Health, at the University of Western Australia, are now building on this previous experience with Indigenous data linkage (Bradshaw et al. 2009, 2010a), by looking at cardiovascular disease and Indigenous identification in WA hospital morbidity data (for further details, see *Thematic List*, pages 4–5). Recently, hospital morbidity data have been linked to deaths data for all coronary heart disease admissions back to 1980. Three different algorithms have been compared to assess Indigenous status at each admission: 'ever-Indigenous', 'majority', and whether a patient was found to be Indigenous after linkage with deaths data. Results from this linkage project had not yet been published at time of writing.

An optimal method for linking data for investigating Indigenous child health

Researchers at TICHR are using linked administrative health datasets to develop an optimal method for identifying Aboriginal and Torres Strait Islander people across datasets (see *Thematic List*, pages 21–2). The use of coding rules is being considered as part of this process. The combination of data from multiple datasets will allow for Indigenous status of the child to be extrapolated, not only from the Indigenous status of their mother or father, but also from the Indigenous status of their siblings. This optimal method will initially be used to analyse disparities in Indigenous child health using data linkage, but may also be more widely applicable (Shepherd 2011).

Queensland

In Queensland, Bryan Kennedy and others have specifically investigated two major issues for Indigenous data linkage in Queensland datasets (Kennedy et al. 2009a; Kennedy et al. 2009b). These issues were:

- the impact on numbers of Indigenous children of including Indigenous children with non-Indigenous mothers and Indigenous fathers recorded in perinatal data (Kennedy et al. 2009a), and

- the impact of using different algorithms to determine Indigenous status when linking with the objective of enhancing Indigenous identification (Kennedy et al. 2009b).

Using data linkage to find Indigenous children with non-Indigenous mothers in perinatal data

Kennedy and others (2009a) explored whether the observed patterns of perinatal outcomes, historically drawn from the Perinatal Data Collection (PDC, i.e. midwives notifications data) and based on the Indigenous status of the mother, would change if babies born with an Indigenous father and a non-Indigenous mother were included (see *Thematic List*, page 29). They analysed births of Indigenous babies registered with Queensland RBDM from 2003–2006, disaggregated by Indigenous status of the parents. This study found that if these babies were included, then 29.1% of Indigenous babies were registered as having an Indigenous father and a non-Indigenous mother. The total number of births recorded in the PDC was greater than in the birth registrations (Kennedy et al. 2009a).

While Indigenous status of the father is recorded in births registrations, it is not recorded in the PDC. At the time of this study, the Queensland Health Statistics Centre did not have birth registrations incorporated into their linkage system. Indigenous status of both mothers and babies was also often missing from Queensland Hospital Admitted Patient Data Collection records (Kennedy et al. 2009a). To identify babies with Indigenous fathers and non-Indigenous mothers in hospitals data, Kennedy and others (2009a) linked all live born singleton babies admitted to a Queensland hospital from the Queensland Hospital Administration Patient Data Collection (QHAPDC to their equivalent PDC record for the period 2003–2006. Other data recorded in the PDC could then be associated with the Indigenous status of the baby (as independently recorded in the QHAPDC). In this way, Kennedy and others (2009a) found that of the 11,302 babies recorded as Indigenous in QHAPDC data, 2,037 (18.0%) had mothers recorded as non-Indigenous in PDC data. Of the babies whose mothers were identified as Indigenous in the PDC, 15.4% were classified as non-Indigenous in the QHAPDC and a further 2.8% of babies had no Indigenous status recorded in the QHAPDC. That is to say, the indigeneity of many babies that were identified

as Indigenous when using mothers' status in perinatal data was still missed when their hospital data was recorded.

Kennedy and others (2009a) then used the above data to calculate the effect of including Indigenous babies with non-Indigenous mothers with other Indigenous babies on their calculations of the measures of disparities between Indigenous and non-Indigenous babies. For a number of variables, including proportion of babies born of low birth weight (a ratio of interest for measuring whether Closing the Gap targets are being reached), Indigenous babies with non-Indigenous mothers have better health as measured with variables in the perinatal data than either Indigenous babies with Indigenous mothers or all babies born to Indigenous mothers. While rates tended to be significantly higher than non-Indigenous babies born to non-Indigenous mothers, Kennedy and others (2009a) found that including Indigenous babies with non-Indigenous mothers made it appear as if the gap between Indigenous babies and non-Indigenous babies had narrowed, and falsely suggested an improvement in outcomes. Kennedy and others (2009a) concluded that the long–standing practice of determining Indigenous status of babies in perinatal data from only the Indigenous status of their mothers should continue for monitoring disparities between Indigenous and non-Indigenous babies.

Using different algorithms to improve Indigenous identification in hospitals data

Kennedy and others (2009b) extracted data from the Queensland Health Public Hospital Client Directory Project in order to compare different algorithms for linking as a method to enhance Indigenous identification in the QHAPDC (see *Thematic List*, page 17). The Client Directory Project provides the ability to link records for an individual client across time and facility within coverage of the data available to it, that is to internally link events in the data to clients. The Client Directory Project is an ongoing linkage run separately to the Queensland Population Heath Research Network (PHRN) node branches, which links data from emergency departments, morgues, pathology labs, and morbidity data. While 94% of episodes of care are covered by the Client Directory, Mater hospitals as well as 25 smaller public hospitals were not included. Two different algorithms were compared:

- use of the most recent record as the source of Indigenous status ('Current'), which has been the usual practice to date, and
- classifying a person as Indigenous if any single record for them recorded them as Indigenous ('ever–Indigenous').

Other algorithms were tested, but only 'Current' and 'ever' were described in Kennedy and others (2009b), as these captured the most extreme figures for proportions of people who were Indigenous (O'Brien 2010). They found that using the 'ever' algorithm resulted in age standardised hospital separation rates 21% higher for all-cause admissions relating to Aboriginal and Torres Strait Islander people than if the 'Current' algorithm was used. However, using either algorithm resulted in much higher hospitalisation rates for Aboriginal and Torres Strait Islander people compared to rates calculated from pre-linked data.

Kennedy and others (2009b) is one of the few attempts to quantify the extent to which the use of an 'ever' algorithm might over-count Aboriginal and Torres Strait Islander people. The 'ever' algorithm is often used without comparing it to other algorithms. While it is generally conceded that as the most sensitive approach, it is most likely to falsely classify some people as Indigenous, this over-count is usually assumed without evidence to either be slight, or else to probably not compensate for overall under-identification in the Dataset in

question. However, Kennedy and others (2009b) attempt to quantify the actual over-count by taking the national figure for coverage by hospitals data of Aboriginal and Torres Strait Islander people of 89% (AIHW 2010), adding the Queensland figure of 21% derived in Kennedy and others (2009b), and concluding: 'A 21% increase in all cause admissions using the 'ever' identified over the 'Current' (that is unadjusted for under-count) results in an estimated public hospital coverage of Indigenous identification of 110%' (Kennedy et al. 2009b).

An over-count that is somewhat greater than the documented level of under-identification may seem contentious. However, measuring over-count in this way should be regarded with caution. Kennedy and others (2009b) obtained the figure of 89% from the AIHW Indigenous identification audit project for the years 2007 and 2008, which calculated a weighted estimate for completeness for all of Australia, excluding the ACT, to be 89% (AIHW 2010). The weighted estimate for completeness for hospitals separations for Queensland alone for the same period is in fact 86% (95%CI: 82–89%). For further information see Table 4.4, page 28, AIHW 2010. Moreover, while the Indigenous identification audit was conducted from admitted patient records made from February to April 2005, the hospital separation records used by Kennedy and others (2009b) were for the financial years 2005–06 and 2006–07. Additionally, the Queensland component of the Indigenous identification audit surveyed 14 hospitals chosen for their likelihood to service as much of the overall Indigenous population as possible. In comparison, the estimates of Kennedy and others (2009b) were based on the total number of Queensland public hospitals in the Client Directory, that is excluding smaller public hospitals and Mater hospitals.

Aside from the use of the national figure, the approach of Kennedy and others (2009b) is defensible. The differences between the national and Queensland figures are small. In the absence of better sources of data, these figures may be the best available means to estimate over-count. Furthermore, it may be that the flaws in this approach are just as likely to underestimate the over-count as overestimate it. Nonetheless, combining these incompatible figures is probably not appropriate, and highlights the need to develop a more rigorous way to measure the relative performance of algorithms for enhancing Indigenous identification.

Kennedy and others (2009b) argue against the use of the 'ever-Indigenous' algorithm from three premises:

- that a single instance of wrongly identifying someone as Indigenous can never be corrected
- that the growth between Censuses in people identifying as Indigenous is primarily due to people changing their Indigenous status, so that use of the 'ever-Indigenous' algorithm for the numerator will cause even greater discordance with a Census–derived denominator
- that since identification is based on self-identification, how that person is identified 'may well change over an individual's lifetime' (Kennedy et al. 2009b).

The first premise maintains that the 'ever-Indigenous' approach is an unreliable method for ascertaining Indigenous status, because a single error in recording a person as Indigenous will fix that person's Indigenous status as Indigenous across the linked data. In order to overcome the limitations that Kennedy and others (2009b) noted with respect to the 'ever' algorithm, they propose that a person's most recent identification should be used instead.

This method presupposes that the most recent identification is the most accurate, and is less likely to be a mistake in the data, but there is little available evidence for this. Indeed, a

linkage project by Bradshaw and others (2009) demonstrates how the 'current' algorithm can be more inaccurate than the 'ever-Indigenous' algorithm (see *Thematic List*, pages 4–5). Bradshaw and others (2009) linked WA hospital morbidity data with a longitudinal survey of Aboriginal people for data from 1980–2006 to see what proportion of the surveyed Aboriginal people were identified as Indigenous in hospital data over this period. While they found identification consistently improved after 2000, there were dramatic fluctuations in the proportion of correct identifications before then. For much of the 1990s, the proportion of correct identifications actually declined; for males, only 57% were identified as Indigenous during hospital admissions in 1995–99. (The reasons for the fluctuation, including this prominent drop, are not known and were not explored as part of this study (Bradshaw 2010). The first large drop coincides with the 1996 federal election). If the 'Current' approach as used by Kennedy and others (2009b) had been used to quantify Aboriginal encounters with the WA hospital system during 1995–99, or just after, it may have severely under-counted the number of Aboriginal and Torres Strait Islander people who had previously been recorded as Indigenous in a prior hospital encounter. While there was a significant and consistent improvement in WA data after 2000 (Bradshaw et al. 2009, 2010a), a similar trend would have to be demonstrated in Queensland data before it is assumed that the most recent identification is correct. At the very least, it should not be assumed that constant steady improvement in the quality of Indigenous identification will always continue, given the changes observed in the WA data prior to 2000.

There is little available evidence to support the second premise (that the increase in Aboriginal and Torres Strait Islander people counted in the Census is mostly due to people consciously changing status) asserted by Kennedy and others (2009b). The ABS Indigenous Mortality Quality Study 2006, found that the major contributors to Indigenous under-count were that many Aboriginal and Torres Strait Islander people were missing from 2006 Census data altogether, and that personal information variables for Aboriginal and Torres Strait Islander people in the Census were of poor quality (ABS 2008a). It may be that there is no single predominant reason for the unexpected change in Indigenous numbers in the Census, but more evidence would have to be gathered before the contribution of all possible factors could be known for certain.

In contrast to the call of Draper and others (2009) for a registry–type system, where a person's Indigenous status is permanently fixed and recorded on a central list, Kennedy and others (2009b) argue that individuals may consciously choose to change their Indigenous status over time, and that they are entitled to do so. Kennedy and others (2009b) therefore conclude that Indigenous status should be observed from as many data points as possible, so that a person's most recent choice about their identity can be inferred, and that data collection practices be improved to maximise the opportunity of individuals to record their Indigenous status independently at each data point.

Indigenous health: towards development of evidence-based policies using data linkage techniques

In an unrelated project, ethics approval has been obtained by Dr Srivanas Kondolsamy-Chennakesavan and others of the University of Queensland School of Medicine, to link hospital separations from 1996–2006 with deaths, perinatal data and cancer registrations to explore reasons for hospitalisations of Aboriginal and Torres Strait Islander people (see *Thematic List* page 15). Linking data from these datasets allows multiple sources of Indigenous status to be used so that under-identification can be partly ameliorated. Data can also be brought together about various health conditions and outcomes experienced by

hospitalised Aboriginal and Torres Strait Islander people. This project takes advantage of the data linkage capabilities of the Health Statistics Centre in the Queensland Department of Health and Health LinQ at the University of Queensland, which comprise the Queensland node of the PHRN. Using these nodes, privacy can be protected by having the linkage conducted onsite within Queensland Health, which would then release the de-identified linked data to the researchers for analysis. Proof of concept for this project has already been established (Coory et al. 2002; Coory & Walsh 2005, see *Thematic List*, page 6). They initially used membership of discrete Indigenous communities as a gold standard against which to assess quality of Indigenous identification.

New South Wales

Enhanced reporting of deaths among New South Wales Aboriginal and Torres Strait Islander peoples using linkage between New South Wales administrative datasets and ABS mortality data

Neville and others (2011) investigated the use of record linkage between mortality data and New South Wales hospital data to improve reporting of deaths among Aboriginal and Torres Strait Islander peoples. This relatively comprehensive study compared Indigenous status in ABS-coded mortality data with five different algorithms for resolving Indigenous status using linked New South Wales Admitted Patients Data Collection (APDC) records. Algorithms and published details of the linkage process are listed in the *Thematic List* attachment that accompanies this report (see *Thematic List*, pages 18–9). These were applied using APDC data from varying periods, presumably to assess the bias created by using a greater number of APDC records made over a longer period of time. Relatively complex algorithms that take into account the number of records and the number of facilities that report a person as Aboriginal or Torres Strait Islander on the APDC were regarded as being most methodologically sound. The authors stress that data linkage is not a replacement for improved Indigenous identification during data collection; they did not assess the accuracy of identification in these datasets using linkage, but rather enhanced the number of Indigenous deaths in the mortality data using Indigenous identification in the APDC. Enhancement was most likely in females, older people, and residents in major cities. The linkage process, carried out by the Centre for Health Record linkage (CHeReL), was quite well reported in this study, though aspects of linkage quality such as numbers of unlinked records were not available to the authors.

Inclusion of other datasets was felt to be warranted to eliminate the bias of using the APDC only, so these researchers subsequently linked these two datasets with births, emergency department admissions and perinatal data. Whereas Neville and others (2011) tested Indigenous status resolved using five different algorithms against a baseline of Indigenous status in the ABS mortality records, this more extensive linkage tested this baseline against an 'ever-Indigenous' algorithm and an algorithm that takes into account the weight of evidence of a person being Aboriginal or Torres Strait Islander from multiple independent sources. This subsequent research had been submitted for publication at the time of writing.

Taylor and others in the New South Wales Department of Health are extending their investigations into how Indigenous status in administrative health datasets can be enhanced by linking seven different New South Wales datasets using up to eight different algorithms (CHeReL 2006, 2008, see *Thematic List*, pages 19–20). This work builds upon previous research described in Neville and others (2011).

Enhanced reporting of other data regarding New South Wales Aboriginal and Torres Strait Islander peoples using linkage

In a related project, the same researchers have gained ethics approval to link the New South Wales Notifiable Diseases Database (NDD) with other databases (admitted patients data from 1 July 2000–31 June 2006, and midwives (i.e. perinatal) data, births data and deaths data from 1 January 1994–31 December 2006), to improve the completeness of epidemiological data about Aboriginal and Torres Strait Islander people (CHeReL 2006, 2008, see *Thematic List*, page 20). One of the priorities of this linkage project is to assess the effectiveness of interventions against notifiable diseases informed by these data (CHeReL 2006, 2008). These databases, apart from the NDD, are used to construct the CHeReL Master Linkage Key, and so compatible personal identifier variables from the NDD for 1 January 1991–31 December 2007 will be used by CHeReL to link with these databases.

The linkage will be used to create a new de-identified Dataset using CHeReL's standard processes. This Dataset will consist of all non-personal variables from the NDD plus Indigenous status from all five datasets (the midwives' variable being from the mother rather than the baby), date of separation from the admitted patients data, date of delivery from the mother's record in the midwives' data, date of birth from the births data, and date of death from the deaths data. The four Master Linkage Key datasets have been chosen because they routinely collect Indigenous status data in a nationally approved format. The compiled Indigenous status variables in the new Dataset will then be used to amend the data extract from the NDD, using an 'ever-Indigenous' algorithm. An Indigenous or non-Indigenous response in any other Indigenous variable will override an 'unknown' response in the NDD-sourced Indigenous variable.

The quality of the linkage will be measured by calculating the proportion of NDD records linked with a record in another Dataset, and by calculating the time between 'contact' with the NDD and 'contact' with the other datasets in linked records compared by age, sex and disease. The performance of the algorithm will be assessed by comparing age, sex, and disease in the linked records based on the percentage of consistent Indigenous status responses between datasets, the proportion of NDD records with previously unknown Indigenous status resolved by linking, and the proportion of NDD records previously recorded as non-Indigenous that are resolved to be Indigenous by linking. If the proportion of NDD records that link proves to be high, and if a large proportion of NDD records with unknown Indigenous status are resolved by linkage, then a de-identified 'epidemiology analysis Dataset' will be loaded onto the New South Wales Health data warehouse as an enhanced de-identified NDD.

If the time between 'contacts' with the linked datasets is short enough, the researchers will seek to maintain and update the linked Dataset on an ongoing basis. This linked Dataset would have great potential for use to reduce notifiable diseases in Aboriginal and Torres Strait Islander people, since the NDD contains data about risk factors such as type of residence and previous workplaces as well as disease data.

Enhancing births data about Aboriginal and Torres Strait Islander people in New South Wales

Other researchers in New South Wales are attempting to improve the accuracy of counts of births to Aboriginal and Torres Strait Islander women by linking the New South Wales Midwives Data Collection (MDC) with birth registrations by the New South Wales Registry of Births, Deaths and Marriages (Xu et al. 2012). Personal identifiers from the MDC and

births have been used to construct the CHeReL Master Linkage Key, so it is straightforward to link and extract a de-identified Dataset including the Indigenous status field from each of the original datasets. The linked Dataset contains New South Wales births from January 2001–December 2005. A new variable, a statistical Aboriginal variable (ASV), was derived from the two Indigenous status variables using an 'ever-Indigenous' algorithm. If the variable from one Dataset was missing, the variable in the other Dataset was entered into the new Indigenous variable. The linkage was assessed by calculating the sensitivity and specificity of the linkage with respect to each Dataset, as well as the completeness of the records and the consistency of linkage (the proportion of linked records that had the same Indigenous status in both datasets). The new Indigenous variable was assessed using capture–recapture analysis, whereby the population of births with Aboriginal mothers is calculated using how many there were in each Dataset and how many were identified in both datasets using linkage. It was then determined if this population of births with Aboriginal mothers was significantly different from the population of births with Aboriginal mothers derived from the ASV. It is hoped that the ASV, by virtue of being derived through linkage from two alternate data sources, will be more complete than either Dataset on its own. The results of this project have recently been published (Xu et al. 2012).

2.3 Studies of Aboriginal and Torres Strait Islander people exclusively

Registries of Aboriginal and Torres Strait Islander people or surveys that target Aboriginal and Torres Strait Islander people can be used as a gold standard as defined in pages 46 & 48 of the *Guidelines* (AIHW & ABS 2012) to assess the quality of Indigenous identification in another Dataset that covers a broader subset of the population and is independent of the register or survey. In this instance, the survey of Aboriginal and Torres Strait Islander people effectively acts like a register of Aboriginal and Torres Strait Islander people, and fluctuations in the quality of Indigenous identification over time in a Dataset linked to the survey can be measured against it.

Evaluating Indigenous identification in Western Australian hospital morbidity data

Bradshaw and others (2009) linked data from the Perth Aboriginal Atherosclerosis Risk Study (PAARS) with WA hospital morbidity data from January 1980–December 2006, and used the former to calculate the accuracy of Indigenous identification in the latter (see 2.2 and *Thematic List*, pages 4–5). The PAARS was used as a gold standard to calculate accuracy of Indigenous identification. 90% of PAARS participants were correctly identified as Indigenous at least once, yet less than 50% were correctly identified at every admission. While Indigenous identification improved consistently after 2000, the proportion of people correctly identified as Indigenous fluctuated wildly before then, with a massive drop, especially for males, around the mid-1990s, before gradually recovering.

Improving Indigenous identification in Western Australian hospital data from four health regions

While surveys conducted for other purposes have been used for linkage, surveys of Aboriginal and Torres Strait Islander people have also been conducted for the specific purpose of using the survey data to improve or measure the level of Indigenous identification in another Dataset. Young (2001) conducted face-to-face interviews in four WA

Health Regions with WA hospital patients and then linked the interview data with WA hospital data (see *Thematic List*, page 17). Sensitivity and specificity was calculated using the survey data as the gold standard. Accuracy of linkage was used to calculate a correction factor for each Health Region and the whole state. A pilot project was used to calculate the standard error for the correction factor. Only 85.5% of interviewed Aboriginal and Torres Strait Islander people were recorded as Indigenous in their hospital records.

Using record linkage to survey data to adjust for under-identification of Aboriginal births on the WA birth register

The Western Australian Aboriginal Child Health Survey (WAACHS) has been extensively used in data linkage studies that bring together different characteristics of Aboriginal and Torres Strait Islander people (see section 3.3 below). The survey's possible use as a gold standard for enhancing Indigenous identification in the Western Australian Births Registry is now being explored (see *Thematic List*, page 21). Where consent was given, each child in the WAACHS was linked to their birth record in WA birth registrations and/or midwives notifications. Indigenous status of the child's mother from these two data sources was then compared. WAACHS data were used to explore whether children whose Indigenous status was recorded differently between the two data sources were on average different from those whose status matched between the two sources. The linked records were used to develop a model for correcting time series based on birth records only. WA birth register data consistently underestimated Indigenous identification by approximately 300 births a year (around 25%).

Auditing Indigenous identification in Australian Capital Territory hospital data: the ACT Hospital Data Linkage Project

Data from Aboriginal medical services, collected by Aboriginal health workers who are known to and trusted by patients, can be a reliable gold standard for data linkage and can protect the privacy of Aboriginal and Torres Strait Islander people using these and other services. The ACT Hospital Data Linkage project aimed to assess the quality of Indigenous identification in ACT hospital data (AIHW 2010, see *Thematic List*, page 17). The quality of Indigenous identification quality in hospital data in other states and the NT was assessed by audit, but given the small population of the ACT, it may not have been cost-effective to conduct an audit over the full time frame of the project (February–April 2005). Data linkage was used for the ACT project to follow-up patients for the 2002–03 collection year over the designated time frame, through collaboration between the AIHW, ACT Health and Winnunga Nimmityjah, the Aboriginal medical service in the ACT. In this project, the AIHW linked hospital data from ACT Health with Indigenous status data from Winnunga Nimmityjah patients, using the latter as the gold standard to determine the accuracy of Indigenous identification in the hospital data. As an Aboriginal–controlled health service, Winnunga Nimmityjah is likely to be trusted highly by Aboriginal and Torres Strait Islander people, so Indigenous status data from Winnunga Nimmityjah are assumed to be more accurate and therefore to be a reliable gold standard. A distinct advantage of using data linkage for this quality assessment was that data linkage avoids the need for face-to-face interviews, which risk being intrusive and culturally insensitive.

ACT whole-of-government Aboriginal and Torres Strait Islander data project

The Australian Capital Territory Government is organising a whole-of-government Aboriginal and Torres Strait Islander data project. This project aims to map data across

various data collections held in the ACT to create a central repository of data regarding Aboriginal and Torres Strait Islander people in the ACT. Linkage may be used as one of several tools to help create this repository. The project is being co-ordinated by the ACT Department of Disability, Housing and Community Services. For further information, see *Thematic List*, page 17.

Improving Indigenous identification in Victorian hospital data: the Victorian Aboriginal Hospital Liaison Officer Collection

The contribution to greater data quality made by data collectors who are trusted by Aboriginal and Torres Strait Islander people and who deliver services to Aboriginal and Torres Strait Islander people in a culturally appropriate manner is difficult to quantify, but is widely recognised and supported by data linkage exercises. In Victoria, for example, it had been the practice since at least 1998 to match data collected by Aboriginal Hospital Liaison Officers to the Victorian Admitted Episodes Dataset to identify and count Indigenous babies (Sullivan 2010, see *Thematic List*, page 30). This has proved important, since Indigenous status of the baby was often not recorded in Victorian hospital records, even though there was the capacity in the records to do so. Moreover, up to 40% of Indigenous babies in Victoria have a non-Indigenous mother and an Indigenous father (ABS 2007), and information about these babies would often be missed in perinatal data, given that, until January 2009, Indigenous status of mothers only and not babies themselves was collected in Victorian perinatal data. Aboriginal Hospital Liaison Officers reported on all Indigenous babies admitted to their hospitals, and the data they collected was used to estimate how many babies were not recorded in other datasets (Sullivan 2010).

2.4 Improving identification to improve the measuring of gaps

When the main aim of a study or report is to compare non-identifying characteristics of Aboriginal and Torres Strait Islander people with those of non-Indigenous people as recorded in a Dataset, data linkage could first be used to improve Indigenous identification in the Dataset before analysis of the characteristics of interest. Comparisons between Indigenous and non-Indigenous people are commonly done with the aim of identifying risk factors other than Indigenous status that might explain disparities between Aboriginal and Torres Strait Islander people and non-Indigenous people, and this would require accurate Indigenous status information for as many study participants as possible.

Despite the potential benefits, there are only a few relatively recent examples where data linkage has been used for this purpose. It is possible that analysts have found other methods to be more effective or more straightforward for improving the quality of their Indigenous status data than data linkage. There may also be concerns that data linkage done badly will lead to false conclusions about Aboriginal and Torres Strait Islander people.

Linking data for this purpose would appear to be an undeveloped practice. Available evidence suggests that the distribution of people with missing Indigenous status is not uniform within populations as a whole, or within subsets based on Indigenous status, after subsequent attempts are made to resolve missing status. Mak & Watkins (2008), for example, found that missing Indigenous status could be resolved through linkage more often for females than males, and more often for some sexually transmissible infections than for others.

Researchers at the School of Population Health at the University of Western Australia, have produced some notable examples that utilised data linkage to pre-emptively improve Indigenous status information. Bradshaw and others (2009) linked WADLS data to a survey of Aboriginal people as a method of deriving a validation gold standard to assess the quality of Indigenous identification before comparing cardiovascular disease characteristics in Aboriginal and non-Aboriginal patients in a cohort study, see sections 2.2 and 2.3 3.1 and *Thematic List*, pages 4–5 (Bradshaw et al. 2010a, 2010b)

The approach of Bradshaw (2009, 2010a, 2010b) is quite distinct from the more conventional use of data linkage described in Briffa and others (2010) (see section 2.2 and *Thematic List*, page 5) and Katzenellenbogen et al. (2010) (see 3.1 and *Thematic List*, page 5). These latter studies test the impact of using different Indigenous status algorithms upon their rates of interest, but rely solely upon sources of Indigenous status data internal to the datasets used for their clinical analyses instead of an external gold standard.

3   Linkage for improving data other than Indigenous identity

3.1   Comparisons of Indigenous and non-Indigenous people

In these studies, multiple datasets are linked, and then people with differing Indigenous status are compared within the linked data. This approach is used for government reporting about service provision or population information, and as such includes deaths data originating from jurisdictional RBDMs. Publicly available reports that use linked data include *WA health measures 2005* (Draper et al. 2005, see *Thematic List*, page 3), and *New South Wales mothers and babies 2006* (Taylor & Bejuk 2009, see *Thematic List*, page 29). The Western Australian Department of Health also regularly uses the WADLS to assist with annual reporting of Aboriginal deaths (WADLS 2010). However, such data are sometimes not publicly reported at all, and even when they are, details on the quality of the linkage and how the linkage was done may not be available.

The examples detailed below illustrate the type and distribution of comparative studies using linked data about Aboriginal and Torres Strait Islander people.

End stage renal failure

In Western Australia, linkage keys in the WADLS are frequently used to extract records from the linked datasets that the keys connect. WADLS-linked data have been linked to compare rates of renal failure, for example, between Indigenous and non-Indigenous Australians (Brameld et al. 1999, see *Thematic List*, page 25). Data were extracted through WADLS for all records with particular renal diagnosis codes. Rates calculated included race-specific incidence rates and age–standardised prevalence estimates. Aboriginal population estimates from the Health Information Centre, DOHWA, were used as the Aboriginal population denominator. Incidence rates were 15 times that of the non-Indigenous population, and higher in Indigenous women than men, unlike non-Indigenous population. Reasons for high rates are discussed. Transplantation waiting times were calculated by Indigenous status.

Coronary procedures in heart disease patients

Following evaluation of Indigenous identification in WA hospital data (Bradshaw et al. 2009), Aboriginal and Torres Strait Islander people with ischaemic heart disease (IHD) recruited from PAARS and a matched non-Indigenous group recruited from the electoral roll were compared for use of coronary revascularisation (Bradshaw et al. 2010b) (see *Thematic List*, pages 4–5). To recruit the latter group, the matched non-Indigenous group were linked to the hospital morbidity database to identify people with IHD for the study; if their hospital records listed them as Indigenous, they were not included in the control group. The only significant difference found between the groups was that Indigenous patients were more likely to have a coronary artery bypass graft than a percutaneous coronary intervention. Similarly, PAARS participants and a matched non-Indigenous group who had not had a coronary heart disease (CHD) event were followed up using data linkage to hospital morbidity data from 1998–2006 for CHD events (Bradshaw et al. 2010a, see

Thematic List, pages 4–5). They were also linked to deaths data to identify patients who died during the follow-up period. These researchers expressed concern that the PAARS participants may not be perfectly representative of the Indigenous community; PAARS participants had higher employment rates than the general Indigenous community, and controlling for socio-economic status suggested the true gap between Indigenous and non-Indigenous patients would be worse than observed. Nonetheless, they found that CHD event rates were 6 times higher in the Indigenous group compared to the non-Indigenous group, and that CHD events occurred at a younger age in the Indigenous group than the non-Indigenous group.

Case fatality following acute myocardial infarction in Aboriginal and non-Aboriginal Western Australians

Using similar data to Briffa and others (2010), Katzenellenbogen and others (2010) compared incidence of acute myocardial infarction (AMI) and case fatality (CF) within 28 days of AMI between Aboriginal and non-Aboriginal Western Australians aged 25–74 from 2000–04 (for further details, see *Thematic List*, page 5). Katzenellenbogen and others (2010) then determined the contribution of hospital admitted cases and non-admitted fatal cases to total incidence of AMI. Sensitivity of incidence rates and case fatality numbers to two different algorithms for resolving Indigenous status was evaluated. Incident AMI events were defined as first-ever hospital admissions for AMI in 2000–04 (i.e. no prior admission for AMI in 1985–2004) or else death due to ischaemic heart disease in 2000–04 with no prior admission for AMI in 1985–04.

Data from records for each individual admitted for AMI to any WA hospital from 1985–2004 were extracted from the HMDC and, if relevant, from death data. A separate WADLS extract consisted of all WA death records for 2000–04 where cause of death was ischaemic heart disease, regardless of previous hospital admissions. The HMDC data extract and the deaths data extract were linked.

Two alternate algorithms were used to derive Indigenous status:

- Inclusive definition ('ever-Indigenous'): A person was counted as Aboriginal if they had ever been identified as Aboriginal in any hospital admission record since 1980 or on their death record

- Restricted definition: A person was counted as Aboriginal if they had been identified as Aboriginal on either their incident AMI hospital record or their death record only.

Age–standardised rates (ASR) for Aboriginal and non-Aboriginal patients, and Aboriginal to non-Aboriginal age–standardised rate ratios were calculated using both Indigenous status algorithms and for 25–54 and 55–74 age groups. Denominators for rates were derived from mid-year estimates for each year from 2000–04 made by the ABS. Sex-specific rates were calculated using Australian 2001 population estimates. Aboriginal patients represented 7.4% of AMI cases, and were much younger than non-Aboriginal cases. Rate ratios were much higher in the younger age group than the older ones, but the proportions of Aboriginal cases were much greater than their share of the population in both age groups. Fatal non-admitted AMI cases contributed a slightly greater percentage of the total ASR in Aboriginal than in non-Aboriginal males but a lesser percentage in Aboriginal than non-Aboriginal females, but the disparity was much smaller than for AMI incidences.

Age-specific total incidence rates for Aboriginal patients were lower when the restricted Indigenous status algorithm was used in comparison with the inclusive algorithm for both

age groups. However, using a different algorithm made no consequential difference to non-Aboriginal rates. Therefore, rate ratios were slightly lower under the restricted algorithm in comparison to the inclusive algorithm. There was a greater effect in the older age group, which was attributed to more previous hospitalisations in older patients and therefore more records in which patients could be identified as Aboriginal. However, case fatalities for Aboriginal records were greater using the restricted algorithm. This was attributed to the reduction in Aboriginal deaths under the restricted algorithm being less than the corresponding reduction in incident cases. Katzenellenbogen and others (2010) did not conclude that the inclusive definition resulted in net over-count of Aboriginal cases, because the differences between algorithms were less than contemporary estimates of under-identification in WA hospital records.

Burden of stroke in Indigenous Western Australians

The impact of stroke as a specific subset of cardiovascular events has rarely been examined in Aboriginal and Torres Strait Islander people. Katzenellenbogen and others (2011) used linked hospital and mortality data to estimate the burden of stroke as Disability Adjusted Life Years (DALYs) (see *Thematic List*, pages 5–6). First-ever hospital admissions for acute stroke in the study period were found by internally linking the hospital Dataset, with an 8–year clearance period to exclude existing/prevalent cases. Incident cases who survived to 28 days were then identified by merging these hospital records with their mortality records (hospital and mortality records would already have been linked when incorporated into the WADLS). Out-of-hospital strokes were found using the Perth Community Stroke Study and used to adjust counts estimated from the linked datasets (i.e. hospital and mortality records). Stroke burden was found to be considerably greater for Aboriginal and Torres Strait Islander people than for non-Indigenous people in WA, and especially so in younger age groups, with Indigenous to non-Indigenous age–standardised stroke incidence rate ratios for patients aged 15 and older 2.6 in males (95% CI, 2.3–3.0) and 3.0 in females (95% CI, 2.6–3.5). To resolve Indigenous status, the 'ever-Indigenous' algorithm was used from records within the study from the hospital and mortality records.

Causes of hospitalisation in Western Australian children

Carville and others (2007) investigated reasons for hospitalisation of children before the age of 2 (see *Thematic List*, page 14). They extracted a cohort of singleton live births for 1990–2000 from Western Australian registered births data and admitted patients data through the WADLS, and then linked these with Western Australian Midwives Notification System data. Indigenous status information was used from the midwives data, rather than the births or patients data, on the premise that this information would be of greater quality in the midwives Dataset than in the other datasets. However, in that time period, a child was classified as Indigenous only if their mother was identified as Indigenous in the midwives data; children with a non-Indigenous mother and an Indigenous father, or whose mother had missing Indigenous status information, would have been classified as non-Indigenous. They found that while infections were the leading cause of hospitalisation for both Aboriginal and non-Aboriginal children, for most reasons for admission, rates of admission were higher for Aboriginal than for non-Aboriginal children. Admission rates were falling for Aboriginal children even as they were rising for their non-Aboriginal counterparts; they noted that a similar pattern had been found in rates in the United States between Indigenous and non-Indigenous children (Holman et al. 2003).

The same data used for Carville and others (2007) were extracted through WADLS for all singleton live births that were recorded occurring between 1 January 1990 and 31 December 2000 and associated hospital admissions for children up to 2-years, whose admission records were coded for particular acute lower respiratory infections (ALRIs) (Moore et al. 2007, see *Thematic List*, page 14). Again, a child was deemed Aboriginal if their mother was listed as Aboriginal on Midwives' Notification System. Relative rates of admission between Aboriginal and non-Aboriginal children and between regions were calculated. Aboriginal total ALRI rates and rates for some specific ALRIs were much higher than for non-Aboriginal children. A decline in pneumonia rates in Aboriginal children was attributed to vaccine campaigns, but rates for ALRIs increased in children aged 12–23 months.

Risk factors for acute lower respiratory infections

While many cohort studies using data linkage seek to assess whether particular risk factors are associated with particular conditions in Aboriginal people, relatively few seek to quantify the relative contribution of each risk factor to those conditions. Moore and others (2010) calculated population attributable fractions (PAFs) for different risk factors for ALRIs in Aboriginal and non-Aboriginal children born in WA from 1996–2005 (see *Thematic List*, pages 23–4). They extracted data through WADLS from birth and death registrations, the Midwives' Notifications System and the Hospital Morbidity Database System for all singleton births in WA in that period, and calculated incidence rates for admissions for ALRIs separately for Aboriginal and non-Aboriginal children for several risk factors. PAFs were then modelled for each cohort to assess the proportional risk of each factor if the other factors were eliminated. The proportional contribution was found to be greatest for males, births in autumn, and multiple previous pregnancies. Low gestational age, in contrast, had a relatively small PAF of 3%. For this study, a child was considered Aboriginal if they were recorded as being Aboriginal at least once in any of the used records from any of the source datasets (i.e. 'ever-Indigenous').

The authors alluded to a particular restriction on their study which did not affect their results but which may have consequences for other data linkage studies involving Aboriginal and Torres Strait Islander Australians. While their analysis by socio-economic status and remoteness was based on location at birth as recorded in birth registrations, current address of study subjects is not available to researchers through the WADLS, so it was not possible to analyse their results by latest or most recent region of residence (from which socio-economic status and remoteness are extrapolated) of study subjects. This restriction would affect data available through other PHRN nodes as well. Study subjects may have changed residence and/or socio-economic status since birth. Aboriginal and Torres Strait Islander Australians are known to be more mobile than other Australians (Memmott et al. 2004), and differential mobility between ethnic groups has been found to differentially affect the capacity of data linkage to correctly collate data about individuals (Blakely et al. 2000). Moore and others (2010) do not believe their data linkage itself was affected by this, since variables in addition to address components were used for linkage, and the restrictions on data available to them meant they did not analyse by most recent location anyway. However, data linkage projects that aim to examine characteristics of Aboriginal and Torres Strait Islander Australians longitudinally may be affected by this problem, and would have to assess whether this would adversely affect their project in the context of their project's aims.

Pneumonia hospitalisations

Moore (2011) examined population trends in hospitalisation rates for pneumonia in Indigenous and non-Indigenous children aged 5 or less for a period before and after the introduction in 2001 of the seven-valent pneumococcal conjugate vaccine (see *Thematic List*, pages 23–4). A retrospective cohort study spanning 1996–2005 was designed based on singleton live births and diagnoses in WA for acute lower respiratory infections for various causes. Linked data from the Midwives' Notifications System, birth registrations, death registrations, and hospital morbidity data was used. Aboriginal children were identified from their birth register records, hospital morbidity records, and from their mothers' identification in midwives data. Annual age-specific incidence rates per 1000 children per year were calculated, and incidence rate ratios of Aboriginal versus non-Aboriginal rates for different diagnoses and age groups were compared. Year-to-year trends were analysed by log-linear modelling. All-cause hospitalisation rates declined significantly between 1996–2000 and 2001–05 for Aboriginal children in all examined age groups (i.e. under 5). In contrast, all-cause hospitalisation rates for non-Indigenous children only declined significantly in the same period for children who were under 1-month old. Incidence rate ratios (i.e. the disparity between Aboriginal and non-Aboriginal children) fell significantly for children older than 6 months, though they still remained substantially greater than unity (one). The annual incidence of all-cause pneumonia declined in Aboriginal and non-Aboriginal children in all age groups with Aboriginal children experiencing the largest decline. Pneumococcal pneumonia declined in both Aboriginal and non-Aboriginal children, but this was only significant in some age groups. Incidence rates of bronchiolitis did not change significantly. Declines and reduction in disparities were attributed to a number of factors, in addition to the vaccination program.

Cancer in Aboriginal and Torres Strait Islander Australians in Queensland

Data linkage can enable the study of interactions between multiple conditions by uniting information about different diseases kept in different datasets. In this way, data linkage can be used to identify complex problems that disproportionately affect Aboriginal and Torres Strait Islander people. Martin and others (2009) assessed survival, stratified by diabetes, among Indigenous cancer patients and a randomly selected cohort of equal size consisting of non-Indigenous Australian cancer patients who had been frequency–matched to the Indigenous patients for several variables. Aboriginal and Torres Strait Islander Australians with diabetes had poorer survival than those without diabetes. This was not the case for non-Indigenous cancer patients, though some distortion in rates may have occurred because Indigenous patients were younger than non-Indigenous patients in the study sample. First diagnoses from 1997–2002 on the Queensland Cancer Registry were linked to the National Death Index and Queensland hospital admissions data, as well as to discharge abstracts to find out the stage of diagnosis and co–morbidities. Patients were classified as Indigenous if they were recorded as Indigenous in any instance in admitted patients' hospital data (Valery 2010). Hospital data was regarded as being more trustworthy with respect to Indigenous identification because medical charts were consulted to confirm Indigenous status. Information about Indigenous status is often recorded in the nurses' notes (e.g. comments about the patient being seen by the hospital Aboriginal Liaison Officer), though medical charts are not primarily an administrative tool, don't contain a formal Indigenous status question, and doctors do not to systematically note Indigenous status when recording clinical information. Hospital patients with missing Indigenous status were excluded from

the analysis. A similar analysis was previously conducted on survival rate differences between Indigenous and non-Indigenous people (without using diabetes as a cross-variable) (Valery et al. 2006). For further details, see *Thematic List*, page 13.

The Western Australian Maternal and Child Health Research Database

For a time, several distinct datasets had been derived from WADLS linked datasets using the WADLS, and were continuously updated. Notably, the Western Australian Maternal and Child Health Research Database (MCHRDB) was created through progressive linking of midwives' Birth Notifications, ABS-validated death data, birth registrations, inpatient hospital morbidity data for children aged up to 15, the WA Birth Defects Registry and WA Cerebral Palsy Register (Stanley et al. 1994, Stanley et al. 1997, see *Thematic List*, page 6). This Database was used to compare rates of genitourinary tract infections (Schultz et al. 1991, see *Thematic List*, page 7), and infant mortality (Freemantle et al. 2006; Freemantle et al. 2004; Kliewer & Stanley 1993, see *Thematic List*, page 7 & 8–9) between Aboriginal and non-Aboriginal infants.

The MCHRDB could either be used on its own, or linked with other datasets in ad hoc linkages, to either further enhance MCHRDB data or allow for analysis of additional variables. Before linking to datasets other than perinatal data, for example, hospital morbidity data were linked to the MCHRDB to compare infant hospital admission rates in general (Read et al. 1994, see *Thematic List*, page 7) and specifically for lower respiratory tract infections (Read et al. 1996, see *Thematic List*, page 8) between Aboriginal and non-Aboriginal infants under the age of 2. Datasets like the MCHRDB could be extremely useful for developing measures for Closing the Gap Indicators relating to children. A limitation of the MCHRDB has been that Indigenous status is only recorded for the mother of a child (Leonard et al. 2003), so it had not been possible, without additional linkage, to distinguish children with an Indigenous father and a non-Indigenous mother from children for whom both parents are non-Indigenous.

TICHR researchers are continuing to link education and health data to look at the relationship between literacy and numeracy in years 1 and 3 (relevant to Closing the Gap education Performance Indicators) and health events such as hospitalisation. They are also using pathology data to enrich cancer registries using linkage. Mortality has recently been investigated by linking deaths, the Birth Defects Registry and midwives data, with Indigenous status being derived from the mother's status in the midwives data.

Mortality in intellectually disabled people

Linking datasets in an ad hoc way to WADLS linked data has allowed useful comparisons of people by Indigenous status. Glasson and others (2005) linked data from the Western Australian Disabilities Services Commission (DSC) to WADLS data to assess prevalence, survival patterns and causes of death in Aboriginal and Torres Strait Islander people with intellectual disabilities. To find deaths and Indigenous status, DSC data at 31 December 2000 for all DSC–listed people diagnosed with an intellectual disability (ID) was linked with hospital morbidity data, WA deaths data held by the WA RBDM, and the NDI. Where cause of death was not recorded, permission was obtained from the WA Registrar–General to review the text of original Medical Certificates of Causes of Death (MCCDs). The NDI was used to track deaths of WA residents who moved interstate. The number of intellectually disabled people who were classified as Indigenous was increased from 330 to 734 as a result

of linking the DSC data to other datasets. Pneumonia and accidents were found to be most common causes of death. For further details, see *Thematic List*, page 25.

Child maltreatment

Researchers at the TICHR have investigated child maltreatment using retrospective cohort studies of children born in WA constructed from several linked WA datasets (O'Donnell et al. 2010a; O'Donnell et al. 2010b). In the initial study, child and parental factors associated with increased vulnerability to substantiated child maltreatment were examined, using a retrospective cohort study of children born in WA from 1990–2005 (O'Donnell et al. 2010a). Cases were identified from child protection data derived from substantiated reports made to the WA Department of Child Protection from 1990–2005. The Midwives Notifications System, Birth Registrations, Hospital Morbidity Database and Mental Health Database were linked. Disabled children were ascertained using the Cerebral Palsy Register, Births Defects Register and Intellectual Disability Exploring Answers (IDEA) Dataset (Petterson et al. 2005). Indigenous status was not derived in a systematic fashion. Rather, Indigenous status was initially taken from information about the child and parents from Birth Registrations and Midwives Notifications; unknown or inconsistent status was then resolved using records from hospital morbidity data or child protection data where they were available (O'Donnell 2011).

Western Australia operates according to an Aboriginal Child Protection Placement principle, which ensures that Aboriginality is specifically investigated and identified in order to ensure appropriate placement of children. Child protection data were considered to have reliable Indigenous status data because of the enforcement of this principle.

Aboriginal and non-Aboriginal children were analysed separately, with socioeconomic disadvantage analysed differently in Aboriginal families due to small numbers in the least disadvantaged categories. Intellectual disability in the child, parental socioeconomic status, parental age and parental hospital admissions related to mental health, substance use or assault were the factors most strongly associated with child maltreatment (O'Donnell et al. 2010a). Aboriginal children and their parents had higher levels of confounding factors such as living in a socioeconomically disadvantaged area. However, for Aboriginal families, children with an intellectual disability had almost double the risk of substantiated maltreatment and children with birth defects had almost 1.5 times the risk. Bias in reporting of child maltreatment was flagged by the authors as a possible reason for the results, but their results were consistent with other studies. For further details, see *Thematic List*, page 10.

Heart conditions in Queensland admitted patients

Western Australia's data linkage infrastructure is considerably more developed than that in other jurisdictions. In other states, researchers have relied upon ad hoc linkages conducted for specific research projects. Coory & Walsh (2005) compared rates of surgical interventions and time to surgery after acute myocardial infarctions (AMI) between Indigenous and non-Indigenous people by linking Queensland Hospital Administration Patient Data Collection (QHAPDC) records to each other based on individuals (see *Thematic List*, page 6). In this way, they were able to follow up patients with AMI to see how they were treated. Described as the first Australian study to look at racial differences in coronary procedure rates, this cohort study sought to avoid numerator–denominator incompatibility by comparing outcomes in Indigenous versus non-Indigenous patients directly. Public sector patients with their first admission for AMI between 1998 and 2002 were followed up using linkage over a

1-year period. They were also linked to death registration data. The rate of percutaneous coronary interventions and adjusted rate of any coronary procedure was lower in Indigenous patients than in non-Indigenous patients. Further, readmitted Indigenous patients were more likely to die during the follow-up period than non-Indigenous patients. Coory & Walsh (2005) were constrained in their analysis because of the small number of Indigenous patients in the cohort (3.8%). Probabilistic linking of computerised discharge abstracts was conducted by Queensland Health's Health Information Centre using date of birth, sex, country of birth and address of usual residence, among others, as linking variables.

Incidence and survival after Acute Myocardial Infarction in the Northern Territory

Incidence and survival rates for AMI were estimated and compared for Northern Territory Indigenous and non-Indigenous populations (You et al. 2009). This was done through a retrospective cohort study for all new AMI cases recorded in hospital admissions data and/or registered deaths due to ischaemic heart disease (IHD) between 1992 and 2004.

Cohorts were constructed through two separate data linkages. Records for the first admission of a patient for AMI for 1992–2004 were extracted from Northern Territory hospital inpatients data. These records were then linked to the National Death Index using name, date of birth, and sex as linkage variables. The successfully linked records were then linked to ABS-coded deaths data for individuals who died between 1992 and 2004 with IHD as underlying cause of death, using the death registration number as the linkage variable. The unlinked death records from this second linkage formed the first cohort (i.e. deaths from IHD of people never admitted to hospital for AMI). The linked records from the second linkage formed the second cohort (i.e. deaths from IHD of patients who had been previously admitted to hospital for AMI). The third cohort consisted of the remaining extracted hospital records minus the records identified for the second cohort (namely, AMI first admissions who either didn't die during the study period, or who died from causes other than IHD). For further details, see *Thematic List*, page 4.

These three cohorts were used to calculate AMI incidence, the proportion of IHD deaths not treated for AMI prior to death, and survival time from first AMI admission to either death or end of the study period. Indigenous and non-Indigenous people were compared using their Indigenous status as indicated on their hospital records or, when not available, their death record. Indigenous status was among the risk factors used for regression modelling. AMI incidence increased for Aboriginal and Torres Strait Islander people during the study period, while decreasing for non-Indigenous people. All-case survival improved for both groups, though more so for Aboriginal and Torres Strait Islander people, but IHD mortality increased for Aboriginal and Torres Strait Islander people due to the increased AMI incidence. Pre-hospital death rates were similar for urban and remote Aboriginal and Torres Strait Islander people.

Mortality in Indigenous and non-Indigenous children in Victoria

Researchers at the University of Melbourne received an Australian Research Council (ARC) Discovery Grant in 2008 for a project that would link data to examine mortality in Indigenous and non-Indigenous children in Victoria. Various hospital, births and perinatal datasets will be used. The researchers intend to produce similar linkage and analysis to that

conducted previously in Western Australia (Freemantle et al. 2004). For further details, see *Thematic List*, page 11.

3.2 Indigenous status as a secondary variable

In these data linkage studies, the principal aim of the study has been to examine a variable other than Indigenous status, such as risk factors for stroke (Lee et al. 2004b) or outcomes for women with drug or alcohol related hospital admissions during pregnancy (Burns & Mattick 2007; Burns et al. 2006a, 2006b, see *Thematic List*, pages 33–4). Analysts in these studies are interested in the Indigenous status of people recorded in data, but examine Indigenous status as one of several possible risk factors for the condition of interest. The characteristics of Aboriginal and Torres Strait Islander people in these studies are normally not analysed in detail, since this is not their principal focus. Linkage used in this way is mostly for research focussed on a specific disease, with little or no published information about how the linkage was performed or the quality of the linkage. In many instances, it is not stated which Datasets the Indigenous status information was taken from. This makes it difficult to compare the use of these datasets for Indigenous data linkage, and to repeat the linkage.

Data linkage in these circumstances is used to either improve the quality of Indigenous identification in a subset of people in a particular health Dataset, or else bring together more clinical information about Aboriginal and Torres Strait Islander people by linking multiple datasets. Data linked in this way are useful for determining if Aboriginal and Torres Strait Islander people are at a greater, or lesser, risk of a specific condition. As subsets of the general population, the numbers of Aboriginal and Torres Strait Islander people can be quite low, even if they represent a disproportionately higher number of people in a given Dataset. This is an issue for research, but can still be a useful tool for reporting purposes. For example, internal linkage of admitted patients data is routinely used in Victoria to track transfers of Aboriginal patients between hospitals, using age, sex, residence, date and separation or admission code as linking variables (Sullivan 2010), that is different records for the same individual within the same Dataset are linked, so that every contact with the hospital system for that individual can be found (see *Thematic List*, page 15). midwives data have also been routinely linked for reporting purposes to admitted patients data, births and deaths data, and perinatal data, though limited information has been reported about the linkage process itself, (AIHW 2009; AIHW: Laws et al. 2007; Taylor & Bejuk 2009). For further information regarding Midwife data see *Thematic List*, page 29

The examples of data linkage referred to in this section are not necessarily of immediate interest as far as developing measures for the Closing the Gap targets are concerned. However, they do give an indication as to what is possible with data linkage, and the benefits and complications of linking with particular datasets.

Nationally based projects

Discordance in cause of death information between deaths data and the Australian and New Zealand Dialysis and Transplant Registry (ANZDATA)

The extent to which there is agreement in the causes of death that are recorded in deaths data and in the ANZDATA for patients with end–stage renal disease was assessed (Li et al. 2003, see *Thematic List*, page 25). To achieve this, ANZDATA was linked to deaths data held by state and territory Registries of Births, Deaths and Marriages (RBDMs) which had cause of death codes added by the ABS using causes of death sourced from Medical Certificates of

Causes of Death (MCCDs). Data linkage was conducted by the Health Gains Planning Branch of the Northern Territory Department of Health and Community Services. Sex and exact dates of birth and death were used as linkage variables; however, deaths data from Victoria, Queensland and Tasmania had to be excluded since dates of birth are not recorded in these states' deaths data. In the absence of names, the combined use of sex and exact date of birth as linkage variables essentially acted as a unique identifier and minimised the risk of false links with discordant causes of death. The proportion of unlinked records and a comparison of some characteristics of the linked and unlinked records, including Indigenous status as recorded in the deaths data, were published. Indigenous status was mostly taken from the deaths data, as this was regarded as a good quality source partly because of contemporary audits, but where Indigenous status was missing from deaths data, it was taken from the ANZDATA data or else assumed to be non-Indigenous. Agreement was assessed using the kappa statistic, which can incorporate both positive and negative agreement between two records (i.e. whether they agree on what cause of death is, and whether they agree on what cause of death isn't). In addition, deaths for which chronic renal failure and some related causes were the underlying cause of death but which failed to link to ANZDATA records were then extracted from the deaths data of the analysed jurisdictions. These unlinked deaths were then compared with the linked deaths.

ANZDATA patients were significantly less likely to link to an ABS record if they were Indigenous. For all patients, the level of agreement on cause of death between the two datasets was generally poor, and Indigenous patients' records were significantly less likely to agree than non-Indigenous patients (28% vs. 38%). Distribution of deaths by cause, age, sex and Indigenous status were compared between the analysed jurisdictions and the non-analysed jurisdictions and found to be broadly similar, suggesting the level of discordance might have been similar in the latter states if linkage had been possible. Patients in the included states were more likely to be Indigenous than in the excluded states. The poor agreement was attributed to difference in coding practice. While the researchers could not distinguish this possibility from false links, they regarded the former as more likely since only two linked records linked to more than one death record. They suggested that a non-significant decrease in matching Indigenous deaths over time may be due to late registrations, since Indigenous deaths have been found to be registered late 3 times more often than non-Indigenous deaths (Cunningham & Paradies 2000).

Western Australia

The WADLS allows for a great number of variables from various data sources to be extracted, and has made possible the analysis of a vast array of risk factors for particular conditions. Comparisons for various risk factors of people for a range of indicators including mortality have been routinely done for reporting purposes by WA Government departments using WADLS data (e.g. Draper et al. 2005; see *Thematic List*, page 3) and compared with other states and territories. Data about people with specific conditions can also be extracted to study specific conditions, such as risk factors for people who have had strokes, including Indigenous status (Lee et al. 2003, 2004b, see *Thematic List*, page 3). Crucially, in these examples, the deaths of patients outside of hospitals who had been admitted at some previous point for stroke or ischaemic attack could be identified.

Recurrent Gastroenteritis among Infants in WA

Data from a subset of the WADLS data records can also be isolated using the WADLS links. Lee and others (2004a) extracted Hospital Morbidity data and deaths data to examine

recurrent gastroenteritis among WA infants, and, of various risk factors, found Aboriginality to be the only significant risk factor for re-admission (see *Thematic List*, pages 13–4). Records were extracted for infants born in 1995 who had a gastroenteritis admission in their first year of life, and were followed up retrospectively by linkage for gastroenteritis re-admissions until 31 May 2002 using their patient number. The number of admissions was compared using a negative binomial regression model to look at incidence across different hospitals. Aboriginal and non-Aboriginal patients were compared, among other factors. Aboriginality was the only significant risk factor for re-admission. Aboriginal patients were also less likely to be dehydrated, but no relationship between gastroenteritis and dehydration was found.

The Western Australian Quality of Surgical Care Project

A particularly powerful use of WADLS has been demonstrated through the Quality of Surgical Care Project (Semmens et al. 1998). Data about people who have undergone specific surgical procedures have been brought together from the various WADLS Datasets using the WADLS keys. These data have then been analysed for an array of demographic characteristics, including Indigenous status, for outcomes on procedures as diverse as vasectomy reversal (Holman et al. 2000, see *Thematic List*, page 34) and breast cancer reconstructive surgery (Hall & Holman 2003, see *Thematic List*, page 1).

Outcomes for mentally ill people

Lawrence and others (2000) calculated mortality rates for psychiatric patients by linking WADLS data with the Mental Health Information System (MHIS) to track deaths within 2.5 years of first contact with the mental health system. They concluded that while Indigenous patients had greater mortality than non-Indigenous patients, the gap was no greater than that between Indigenous and non-Indigenous people in the general population. For further details, see *Thematic List*, page 26.

The MHIS and deaths data have been linked with morbidity data, through WADLS, to study interpersonal violence and its consequences. Meuleners and others (2008) quantified hospitalisations due to injury by others and analysed them for several risk factors, including Indigenous status (see *Thematic List*, page 32). Indigenous status was derived from the hospital morbidity data. In the period and age groups they examined, 47% of hospitalisations, and 74% of repeat hospitalisations were for Aboriginal and Torres Strait Islander people. Most Indigenous repeat hospitalisations were for females, whereas as the majority of repeat hospitalisations for non-Indigenous people were for males. Where more than one record existed, the Indigenous status option that was recorded on a majority of records for that individual (i.e. 'majority') was used. If there was an unreconciled conflict in the Indigenous status field, Indigenous status was treated as missing. Records with missing Indigenous status were quantified, and excluded from analysis rather than combined with non-Indigenous records. Despite the large numbers of Indigenous records in these studies, the numbers of people of Torres Strait Islander descent were too small to be considered separately, so Aboriginal and Torres Strait Islander people were still combined. A similar analysis was done for victims of interpersonal violence admitted for mental illness (Meuleners et al. 2009, see *Thematic List*, page 32). This work is being continued with more recent data (see *Thematic List*, pages 26 and 33).

Palliative care for dying cancer patients

Linkage of the WADLS datasets with other datasets has allowed a greater range of variables to be analysed. A relatively recent example of this demonstrates how useful these secondary

linkages can be, especially when the derivation of the analysed Indigenous variable is explained. Rosenwax & McNamara (2006) compared specialist palliative care for people dying with cancer as opposed to other conditions by linking WADLS morbidity and deaths data with another WA data system, the Silver Chain Nursing Association system, which had more extensive information about palliative care. When comparing by Indigenous status within these groups, they used a kind of 'ever-Indigenous' approach, based on the morbidity and deaths data, while combining people with missing Indigenous status in both these Datasets with non-Indigenous people ('Others'). By publishing this approach, the extent of possible over-count or under-count can be surmised. Using the 'ever' approach might be expected to lead to over-count and therefore disguise any differences between 'Indigenous' and 'Other' people. This would obscure observed differences between Aboriginal and Torres Strait Islander people and Others, and yet the authors found Aboriginal and Torres Strait Islander people were the worst serviced amongst the non-cancer group, and that a greater proportion of Indigenous patients were in hospital rather than in other palliative care services. For further details, see *Thematic List*, page 29.

Cumulative mortality in Western Australian children

Western Australia's derived linked datasets, such as the MCHRDB (see above), have proven to be a rich resource with data from a great number of variables. Alessandri et al. (1999) recruited cohorts for what they describe as the first study of cumulative mortality in birth cohorts of children in a total population (see *Thematic List*, page 8). Data for Western Australian children born between 1980 and 1989 were extracted from the MCHRDB and followed via internal linkage until their seventh birthdays to find deaths between ages of 1 and 6. The cause of death recorded in the MCHRDB, ultimately from ABS-coded and validated deaths data held by the WA RBDM, was then compared with that from post-mortem records from various WA sources. The Births Defects Registry had been linked to the MCHRDB by the time of this study, so cause of death as recorded in the Births Defects Registry was used where no other cause was available. Indigenous status data from the MCHRDB were used, though the ultimate data source and how conflicts were resolved was not stated. They found that the mortality rate for Indigenous children was 4 times higher than for non-Indigenous children, and that after accidents, the next most common cause of death differed between Indigenous (infections) and non-Indigenous children (birth defects).

Intellectually disabled children

Linkage of the MCHRDB with Western Australian Disabilities Services Commission data and educational services data allowed intellectually disabled children to be identified and to have several characteristics of this group defined and rates of intellectually disability among Western Australian children estimated (Leonard et al. 2003). Numerator–denominator incompatibility was avoided by also using the MCHRDB as the source of the denominator for calculating rates. They found that intellectual disability was more common in children of Indigenous mothers than non-Indigenous mothers. A subset of children with mild to severe intellectual disability, who had an unspecified cause of ID or autism, were selected from this linkage and were analysed for several demographic characteristics (Leonard et al. 2005). By doing so, Leonard and others (2005) found that while mild to severe intellectual disability was more common in Indigenous than non-Indigenous children, autism was less common. This linkage was further developed into a permanent linked Dataset, called the Intellectual Disability Exploring Answers (IDEA) (Petterson et al. 2005), which has allowed for more nuanced analyses. For more information about IDEA see section 3.1. For example, Leonard and others (2008) found through using IDEA that low birthweight was associated with

intellectual disability in both caucasian and Indigenous children, though they warn that the numbers of Indigenous children in IDEA were small. No significant differences were found between caucasian and Indigenous children. For further details about these projects, see *Thematic List*, page 24.

Hospital admissions as indicator of child abuse

In conjunction with O'Donnell and others (2010a), which analysed Aboriginal and non-Aboriginal cases separately (see section 3.1), TICHR researchers investigated prevalence and other characteristics of child maltreatment using hospital admission and death records, and identified common injuries and conditions associated with maltreated children's hospital admissions (O'Donnell et al. 2010b). Two cohorts were constructed for this purpose. The cohort for investigating prevalence rates and age-specific rates of relevant admissions was constructed from linked data about all children aged 0–17 from 1980–2005 included in the Hospital Morbidity Data Collection, death registrations, midwives notifications, and birth registrations. The cohort for assessing whether various risk factors were associated with assault or maltreatment-related admission consisted of children born in Western Australia from 1980–2005. Annual trends in prevalence of these admissions were calculated and child characteristics investigated using logistic regression. ABS census data were used to derive denominators for rates. Aboriginal children were over-represented in assault and maltreatment admissions compared to non-Aboriginal children. 19% of the identified assault-related deaths were of Aboriginal children. Aboriginal children in this study were more likely to have infectious and parasitic diseases, and of those reported with maltreatment, a higher proportion had nutritional diseases, respiratory problems and skin disorders. Possible reasons for these findings are discussed, including reporting bias. The possibility of reporting bias led O'Donnell and others (2010b) to suggest that their findings with respect to Aboriginal children should be treated with caution. For further details, see *Thematic List*, page 10.

Inequalities in preterm births

Langridge and others (2010) used linked data to investigate factors that contribute to inequalities between social and racial groups in preterm births in Western Australia (see *Thematic List*, page 30). Data for all live singleton births in Western Australia between 1984 and 2006 were extracted from the Midwives Notification System and birth registrations, and linked for each birth. Data were organised into 5–year batches by birth year around each census year. Data were then analysed for preterm birth frequency by maternal age group, parity, marital status, maternal skill level, maternal residential location at birth, pre-existing medical conditions and pregnancy complications, father's skill and age, remoteness, socioeconomic status of residential area, and maternal smoking during pregnancy. Aboriginal status of infants was defined as having at least one parent being identified as Aboriginal in any of the data sources (a variation on 'ever-Indigenous'). The Western Australia Family Connections Genealogical Project (Glasson et al. 2008) was used to obtain linked information about who the parents of the infants in the study were and to then link to the parents' data, along with the Midwives Notification System data. The mother's country of birth from the child's births data was checked against maternal ethnic origin in the midwives' data to exclude overseas–born women who had been misidentified in the midwives data as Aboriginal (Langridge 2010). Logistic regression was used to compare Aboriginal and non-Aboriginal infants against three models with different collections of risk factors. Unlike other studies, Langridge and others (2010) found that the social gradient at the area level disappeared among Aboriginal and non-Aboriginal infants after parental,

biological, demographic and socio-economic factors, and smoking during pregnancy were adjusted for. However, Aboriginal infants were almost 50% more likely to be born pre-term than non-Aboriginal infants, even after other examined factors were adjusted for. The disparity in rates of pre-term births between Aboriginal and non-Aboriginal infants was found to have increased over the study period, so that by 2004–06, the disparity was similar to inequalities observed 20 years previously.

Characteristics of road-related arrests

Outside of the WADLS, the Western Australian data linkage infrastructure has been used to facilitate linkage of the Road Injury Database to the offenders database to compare the characteristics of people arrested for drink–driving and at road crashes (Rosman et al. 2001, see *Thematic List*, pages 12–3). Records of all Western Australia drink–driving arrests from 1987–95 were linked to records of crash–involved drivers from the WA Road Injury Database and arrest records from the INOIS offender tracking database. Non-crash related drink–driving arrests were assumed to be from routine police enforcement. Indigenous status was not recorded on the Road Injury Database, but was taken from arrest records, where Indigenous status is assessed by the police (the cited study claims 94% success by police in this). Aboriginal people were over-represented in routine police enforcement alcohol-related arrests (12.4%) compared to crash arrests, though they were still 7.2% of the latter. They postulate that this may be because crashes involving Aboriginal people were less likely to be reported, or else because of over-policing of Aboriginal people.

The linkage for this study was relatively straightforward, since it could make use of driver's licence number as one of the linking variables between the two Datasets. However, when analysing by Indigenous status, the standard definition of Indigenous identification (and in particular, self-identification) was not used. Rather police identification of offenders as recorded in the offenders' database was utilised to identify Indigenous status. This source of Indigenous status information is not recommended.

New South Wales

The New South Wales data linkage node CHeReL is much newer than the WADLS. However, New South Wales draws upon a larger population base and has already produced some useful linked research about the health of Aboriginal and Torres Strait Islander Australians. Additionally, the New South Wales Inpatient Statistics Collection is regularly linked with RBDM death registration data and ABS mortality data to allow outcomes of hospitalisations to be monitored, principally for a more comprehensive recording of deaths (CHeReL 2006, 2008).

Alcohol and drug use during pregnancy

This linkage project examined alcohol and drug use during pregnancy across a range of variables including Indigenous status (Burns et al. 2006a, 2006b, see *Thematic List*, pages 33–4). This project linked the New South Wales Inpatient Statistics Collection to births in the New South Wales Midwives Data Collection. The linkage process is well described, but Indigenous status was used as one of the linking variables. As such, babies with conflicting Indigenous status (or who had mothers with conflicting Indigenous status), including where Indigenous status was missing in some records, could not have been linked and would have been excluded from the final Dataset. The characteristics of these babies and mothers excluded by the linkage process are not reported, so the effect on the comparison of Indigenous and non-Indigenous babies cannot be assessed. A further linkage project linked

these two datasets with the NSW Pharmaceutical Drugs of Addiction System (PHDAS), which records approvals to dispense methadone to individual patients, to examine prevalence and risk factors for neonatal abstinence syndrome (Burns & Mattick 2007, see *Thematic List*, pages 33–4). A full list of linking variables is not described in this study. However, the linkage process is otherwise similarly described in Burns et al. (2006a) and Burns et al. (2006b), so Indigenous status may also have been used as a linking variable with similar consequences. This example demonstrates why Indigenous status should not be used as a linking variable.

Mortality in New South Wales prisoners and former prisoners

Mortality in New South Wales prisoners has been studied by linking the New South Wales Corrective Services' Offender Integrated Management Services (OIMS) with the NDI warehoused by the AIHW (Kariminia et al. 2005, Kariminia et al. 2007a, Kariminia et al. 2007b, see *Thematic List*, page 32). Indigenous status was taken from self-identification records in the OIMS for analysis purposes, and was not enhanced by comparison with Indigenous status in the NDI. However, the level of reportage of Indigenous status for the linked records was quite high (97%). Documentation of the linkage was relatively detailed in this study, with sensitivity and specificity quantified. In particular, as a first step, the quality of the NDI with respect to its capacity to identify deceased former prisoners was assessed and quantified using data linkage, with the offenders database used as the 'gold standard' (Kariminia et al. 2005). However, the links and non-links were not disaggregated by Indigenous status, so it was not reported whether deceased Indigenous former inmates were more or less likely to be missed or incorrectly linked by the linkage process.

Northern Territory

To a certain extent, linkage of data related to Aboriginal and Torres Strait Islander Australians does not face the same issues in the Northern Territory as in other states and territories. This is because each person who comes into contact with the NT health system (for Aboriginal and Torres Strait Islander people, usually at their birth) is given a unique identifier, the Hospital Registration Number (HRN), which is then used to identify the health records of patients and to record new data about them. The HRN is used not only on hospital records, but on a number of other datasets held by the NT Government as well. NT public hospitals have had a single information system with the HRN as a NT-wide unique client identifier since 1991; in the late 1990s NT-wide information systems were introduced for several other services (such as mental health and child protection) and the client identification module from the hospital system became the basis of a Client Master Index (CMI) serving multiple information systems. Consequently, the HRN effectively acts as a master linkage key to join data about individuals across different datasets. Finding data about individuals recorded in different health datasets is a straightforward matter of searching for records with specific HRNs, rather than requiring an ad hoc data linkage process using personal information variables like name or date of birth. A person's Indigenous status is recorded upon the creation of their HRN and stored in the CMI (Condon 2011). While the recording of Indigenous status is not fixed, observations of staff practice during hospital data validation (Condon et al. 1998, see *Thematic List*, page 14) suggest that Indigenous status is not likely to be changed at subsequent contacts with the health system.

Aboriginal Territorians may be quite mobile within the NT, but Aboriginal Territorians are relatively unlikely to migrate out of, or into, the NT (Carson 1994; Condon et al. 2004).

Aboriginal communities in the NT are quite cohesive, and the membership of these communities is very stable. The use of traditional names among Aboriginal Territorians is commonplace, as are common surnames within the Aboriginal community, and these practices are often relied upon to ensure people are identified as Aboriginal in datasets from the outset. As such, the first recording of people as Aboriginal other than at, or soon after, birth is regarded as rare.

However, there has been an increase in the use of data linkage to link data with HRNs with datasets that do not have HRNs (e.g. deaths data) in the past decade, and this is proving useful for both reporting and research purposes. Since these linkages have had to rely on personal information for linkage variables, and link datasets on which Indigenous status has been recorded independently, many of the issues with linking data about Aboriginal and Torres Strait Islander people found in other parts of Australia apply in these instances. The relatively high proportion of Aboriginal people in the NT may also allow for the testing of algorithms in circumstances less affected by small sample size.

While HRNs are now used regularly to link data for NT Health reporting purposes, external linkages from data carrying the HRN are only possible for research, and on an ad hoc basis, and require appropriate ethics approval.

Growing up in the Territory

The Health Gains Planning Branch of the NT Department of Health and Community Services reports on health in the NT. As part of these duties, surveys of parents and schools have been conducted on children's health and wellbeing. In 2004, the results of these surveys were linked to broader health and education department information relating to the children who were the subject of these surveys using the CMI (Li et al. 2006). This linkage allowed the survey data to be analysed and reported on in the context of the children's family, overall health, and social environment. The Parent Survey asked parents to disclose the Indigenous status of both themselves and their children who were the subject of the survey (Carson et al. 2006). A second round of these surveys is planned for 2012, so linkage with general NT health and education data is likely to remain an ongoing fixture of *Growing up in the Territory* reporting. For further details, see *Thematic List*, page 10.

Other Northern Territory activities

Ethics approval has been obtained for a project investigating immunisation of newborns against hepatitis B, led by Bette Liu of the University of New South Wales. A cohort of people born from the late 1980s will be examined by linking immunisation and perinatal data. Participants will be compared by Indigenous status in conjunction with other factors. Data linkage for this project has been conducted through the University of New South Wales.

Child protection data are being examined to determine the characteristics of children recorded in these data. To facilitate this, child protection data will be linked to hospital and perinatal datasets. This project is led by Steve Guthridge, through the University of Adelaide.

The pathway of mental health in children is being analysed by bringing together various datasets through data linkage. The aim of this project is to examine the impact of health promotion on children's health. As a minor aim, the mobility of children between data collection points will be examined.

The NT Health Gains Planning branch is in the early stages of building a child health universe, a data storage area that brings together all possible information relevant to a child's health that can be linked using the HRN. Universes already exist for perinatal, growth assessment and hospital data.

Karen Dempsey of the NT Health's Health Gains Planning branch is examining the characteristics of road accidents in the NT by linking police crash records with NT health data, including an Aboriginal cohort. Probabilistic linkage is being used, using name, age at the time of crash, sex and Indigenous status among other linkage variables. Results may be published in 2012. Indigenous status information that is recorded by police on police crash records is being used as the source of Indigenous status for participants in this project. For further details, see *Thematic List*, page 13.

Other States

Rotavirus infection in Queensland children

Linkage of datasets has not been used a great deal to examine infectious diseases in Aboriginal and Torres Strait Islander Australians. This is partly because of the lack of personal information in national communicable diseases data, but also because of the low level of Indigenous identification in infectious disease data. For example, Campbell and others (2009), found that Indigenous status is missing from 42% of records in the Queensland Health Notifiable Conditions database (NOCS). Campbell and others (2009) analysed Queensland hospital records for children under 5 admitted for rotavirus infection or acute gastroenteritis after linking them to a sample of pathology laboratory test requests (see *Thematic List*, page 22). Each individual within the hospital data had a unique identifier, so no other additional linkage variables had to be used. Using this unique identifier, they were able to link the de-identified admitted patients' data with the sample of de-identified laboratory test requests, as well as to link records within the admitted patients' data for the same individuals, without the need for names. Campbell and others (2009) also used age group, gender, month of notification and facility to assist with the linkage. Campbell and others (2009) found that 89% of rotavirus diagnoses were confirmed by laboratory tests, and there was a relatively greater burden in Indigenous children, concentrated in children less than 1-year old, in rotavirus-related hospitalisations prior to the commencement of a vaccination program. Rates in the NOCS Dataset were also calculated, and were consistent with the hospital data, but could not be linked in the absence of names from the hospital data and sufficient other variables to link the two datasets. Researchers at the Queensland Children's Medical Research Institute and Menzies School of Health Research intend to extend this kind of research to assess the effectiveness of a complete course of a rotavirus vaccine at preventing rotavirus and non-rotavirus acute gastroenteritis hospitalisations in Indigenous and non-Indigenous children in Queensland and the Northern Territory (see *Thematic list*, page 24).

Literacy and numeracy outcomes in Queensland secondary school students

The Queensland Studies Authority routinely links data about literacy and numeracy outcomes, though the results are not published. They are currently following cohorts of state school students after Year 12. Students are surveyed 4 months after finishing school and followed up 3-years later (the follow-up rate has been around 80%). The Queensland Studies Authority also followed up students by phoning those who left school before Year 12 (at 12 months since leaving or, for early leavers, 6 months after they left school). For all students,

Indigenous status was recorded. The study used data provided to schools from the students' family, rather than data gained by interview. Other datasets employed in the undertaking of this study included disabilities data. The cohort will eventually be extended to private schools as well. Queensland is currently the only state running this type of study.

Victorian hospital data

Deaths in hospitals in Victoria are routinely linked to death registrations, and are disaggregated by Indigenous status for reporting purposes (Sullivan 2010, see *Thematic list*, page 28). Data collected by Victorian Aboriginal Health Liaison Officers and maintained in the Aboriginal Health Liaison Officers (AHLO) Data Collection have been provided to the ABS for cross-matching with registered deaths to help identify Indigenous deaths. Where Indigenous deaths are concerned, Aboriginal Hospital Liaison Officers report on any deaths known in the Aboriginal community. These data are cross-matched with the registered deaths. This provides an estimate of how many Aboriginal deaths are not registered as Aboriginal (Sullivan 2010). The AHLO data collection has now been discontinued as a separate health Dataset; in future, data collected by Aboriginal Health Liaison Officers will be incorporated directly into Victorian hospital datasets.

Risk factors for incomplete immunisation in Victorian children

Haynes and Stone (2004) linked Victorian birth records for 1998 to records from the Australian Childhood Immunisation Register, using probabilistic linkage, to determine the predictors of incomplete immunisation in Victorian children (see *Thematic list*, page 16). After multivariate analysis, some maternal factors were found to be significant predictors of incomplete immunisation in children aged 12 months, but being Indigenous was only one of these predictors.

3.3 Studies of Aboriginal and Torres Strait Islander people exclusively

In this final type of study, data about Aboriginal and Torres Strait Islander people from a Dataset with an Indigenous focus are linked with data from other datasets. The former is often a Dataset specifically developed about Aboriginal and Torres Strait Islander people or a group of Aboriginal and Torres Strait Islander people, such as survey data, and effectively acts as a register of Aboriginal and Torres Strait Islander people to use as a gold standard to measure the quality of Indigenous status recording in other datasets.

Data are linked in this way for different reasons. In some instances, extra information about surveyed Aboriginal and Torres Strait Islander people that was not collected in the original survey can be obtained by linking with another Dataset.

Evaluation of the Northern Territory Emergency Response Child Health Check Initiative

The Northern Territory Emergency Response (NTER) was announced by the former Australian Government on 21 June 2007 in response to the *Little children are sacred* report by the NT Board of Inquiry into the Protection of Aboriginal Children from Sexual Abuse (Anderson & Wild 2007). As part of the NTER, the Department of Health and Ageing conducted health checks of eligible children based on MBS item 708 under the Child Health Check Initiative (CHCI). The AIHW collected and maintained data about the CHCI, reported

on its progress (DoHA & AIHW 2008), and assisted evaluation of the CHCI. As part of this evaluation, data from the AIHW CHCI database were linked with NT health data via the CMI by the NT Health Gains Planning Branch in the NT Department of Health and Families, and analysed and reported on by Allen and Clarke Policy and Regulatory Specialists (Allen & Clarke 2011, see *Thematic list*, page 11).

Children who were eligible for a Child Health Check (CHC) (residents of prescribed areas aged under 16) were identified in the CMI (14,647 children). In the course of dispensing child health checks, DoHA collected patient Hospital Registration Numbers (HRNs), date of birth, sex and community identifiers from patients, the patients' parents or guardians and local clinics. These variables were used as linkage variables to match CHCI records to records in the CMI. In this way, the number of eligible children who had a CHCI health check could be calculated (8,415), as could the number of children who received a CHC but could not be linked to the data extract from the CMI (2,190). Of the latter, 49.9% were residents of non-eligible communities (presumably visiting eligible communities), while the remaining 50.1% included records with missing HRNs, children recorded as non-Indigenous, or children with unknown or interstate residential status (Allen & Clarke 2011).

Data from the NT Hospital Morbidity dataset, which carries the HRN, were matched to the eligible data extract from the CMI. In this way, the proportion of eligible children who had been admitted to hospital could be calculated, and the reasons for their admission quantified. Similarly, the CMI data extract was linked to the NT midwives dataset, so that the birth weights of eligible children could be analysed. The CMI extract was also linked to the Growth Assessment and Action program Dataset (a program that monitors growth and health of children aged 0–5 in remote NT communities and implements action if growth falters) but little extra unit record data useful for the purpose of analysis was made available to the analysts. Nonetheless, accessing extra data items through linkage from these three datasets made it possible to compare hospitalisation rates between eligible children who received a child health check and those who did not, and to determine whether eligible children with low birthweight had had a child health check.

While not quantified, the report identified problems for data linkage with the CMI Dataset such as children with multiple HRNs, incorrect or unrecorded Indigenous status, and out–dated residential status. On this basis, Allen and Clarke (2011) advised that the comparisons between the CHCI Dataset and the linked data described above be treated with caution.

Notably, Allen and Clarke (2011) reported that protracted negotiations and limited access to some datasets restricted the amount of useful information they could access for data linkage and analysis. Ethics approval by both Northern Territory human research ethics committees was not enough to assure these data custodians that the privacy of individuals whose records are held on the affected datasets would be sufficiently protected.

Lifestyle factors in coronary heart disease

Burke and others (2007) linked data from the Kimberley Aboriginal lifestyle survey for 514 participants, with WA hospital, cancer and deaths data, constructing a variable from the latter datasets for coronary heart disease (see *Thematic list*, page 27). Lifestyle, including elements such as diet, which were included in the Kimberley survey, could then be compared with coronary heart disease (CHD). Indigenous status was already known or assumed from the survey. Survival endpoints over the time interval of linked datasets were calculated. Two–step cluster analysis and endpoints for CHD and all-cause mortality calculated from linked data were compared. Lifestyle was found to predict CHD and all-

cause mortality, though the number of deaths from CHD was small. This was the first study of lifestyle in Aboriginal and Torres Strait Islander Australians as a predictor of long–term outcome using longitudinal data. Importantly, the survey contained dietary data for individuals, rather than aggregate data, which made it possible to directly examine the relationship between diet and the manifestation of CHD.

Evaluating a health assessment

Calver and others (2005) used a random sample of Aboriginal and Torres Strait Islander people who had previously completed questionnaires about alcohol use and blood pressure prior to a trial of a multi-component health assessment (see *Thematic list*, page 27). The interviewees had been selected for the questionnaires using the WA Community Health and Leprosy Survey Registers. All other Aboriginal and Torres Strait Islander people in the registers were then selected as a control group for this study. The data from these registers were linked to hospital, cancer and death data for a 13–year period to assess the impact on mortality and morbidity of the multi-component health assessment. Few differences in risk of death were found. No difference was found in cancer rates. While treatment increased during the study period, no effect was found on longevity.

Causes of hospitalisation of Aboriginal and Torres Strait Islander people in Queensland

The existence of a Community Health Register in Western Australia made it possible to construct random samples of people to survey for the PAARS who could then be asked for permission to use data about them from other datasets in the course of surveying them (see sections 2.3 and 3.1). In other states and territories, more creative solutions have been required. Coory and Johnston (2006) used probabilistic matching of computerised discharge abstracts to identify all the admissions to any hospital in Queensland of Aboriginal and Torres Strait Islander people whose usual residence was one of a particular set of remote Indigenous communities (see *Thematic list*, page 15). Living in one of these communities was taken to indicate that the person was Indigenous, thus avoiding the problem of inconsistent identification as Indigenous over time. While useful for this particular exercise, this is not an approach that is applicable to most Aboriginal and Torres Strait Islander people, notably urban dwellers, either in Queensland or nationally.

Uses of the WA Aboriginal Child Health Survey for linkage

The Western Australian Aboriginal Child Health Survey (WAACHS, see *Thematic list*, pages 9–10) was originally constructed partly by linking surveyed people to health service and education system data (Silburn et al. 2006; Silburn et al. 1996; Zubrick et al. 2004; Zubrick et al. 2006). The original survey consisted of a random representative sample with usable data for 5,289 Aboriginal children aged 0–17, including 1,480 adolescents aged 12–17 (Silburn et al. 2007). WAACHS data have been linked to hospital morbidity data and the MCHRDB to examine feeding associations with parent–reported infections and hospitalisations in WA Indigenous infants and children (Oddy et al. 2008). Of particular interest is that low birth weight (relevant for Closing the Gap Performance Indicators) was found to be among the risk factors for parent-reported chest infections in children and hospitalisation for upper and wheezing lower respiratory infections. This study clearly demonstrates the potential of the WAACHS for linkage.

In a more recent example, Zubrick and others (2011) used WAACHS data to examine the impact of maternal death or loss on health and education characteristics of Aboriginal children and young people. Most of the analysis variables for this study were already present in the WAACHS Dataset, but maternal death was detected by linking the mother's details to death registrations. Permission to link data was sought from individuals when data was collected from them during the survey process, so the numbers of individuals for whom data linkage was not attempted due to lack of consent was also reported.

The WAACHS and other survey enterprises are made possible by the advanced data infrastructure that has developed in WA over several decades, and are excellent resources for linkage. Ironically, the lack of such sophisticated resources in other states and territories may render the results of linkage projects that use these resources incompatible with these other jurisdictions and, therefore, these methods or results might not be applicable nationally.

Pathways to healthy outcomes for WA Aboriginal mothers and babies

Research that has been made possible by the development through data linkage of the MCHRDB is being extended at TICHR to link birth data with a wide range of data about the health of the mother. This further research will investigate pathways into healthy as well as unhealthy outcomes for Aboriginal people.

Some studies suggest that smoking and other stressors can adversely affect offspring for several generations after the stress through epigenetic and other environmental mechanisms. Fetal growth is being examined over three generations by linking data across several datasets (see *Thematic list*, page 11) for women who had given birth after 1980, their daughters who had given birth, and their offspring. The current project will focus on Aboriginal and Torres Strait Islander populations. Percentage of optimal birth weight is used as a measure of fetal growth for an individual fetus accounting for non-pathological determinants including, maternal parity and height in addition to fetal sex and gestational age. Optimal birth weight is calculated from reference values from a population free of known pathological determinants, as a means of providing a measure of predicted growth in 'ideal' conditions. Initially, the perinatal data will be used to identify mother–offspring pairs as this information is regarded as the most accurate in relation to Indigenous status of women residing in WA. Further optimisation of the identification of Indigenous participants utilising sibships, other family connections and data sources within the linked dataset is currently being explored. 'Sibship' is a term used in epidemiology and public health to refer to the group or number of children produced by a pair of parents. The study will exclude Aboriginal children born to non-Aboriginal mothers until better data for identifying Aboriginal fathers in these datasets is obtained. Preliminary results presented at the 2011 Coalition for Research to Improve Aboriginal Health (CRIAH) Conference indicated a high frequency of sub-optimal birth weights (approximately 1 in 5 births) were recorded for Aboriginal women in each generation at the jurisdictional level. How this relates to the later risks of chronic disease in both mothers and their infants will be investigated through additional linkages to the mothers' hospitalisation data and to births in the next generation. Comparisons with non-Aboriginal mothers will not inform the best preventive strategies for Aboriginal populations, so Indigenous and non-Indigenous mothers will not be compared in this phase of their research. For further details, see *Thematic list*, page 11.

Impact of Northern Territory Pneumococcal Vaccine Coverage Program

In 2001, the NT Childhood Immunisation schedule began to provide 7-valent conjugate vaccine doses against pneumonia with a 23-valent booster to a large number of NT children, including all NT Aboriginal infants (Selvey 2001). The impact of this program on pneumonia and acute lower respiratory tract infections (ALRIs) in Aboriginal children was recently assessed by linking the NT Immunisation Register and NT Hospital Discharge Dataset, with the latter used as the source of information about hospital admissions for ALRIs and various manifestations of pneumonia (O'Grady et al. 2010a; O'Grady et al. 2010b, see *Thematic list*, page 23). Linking these datasets is straightforward, as each person who comes into contact with the NT hospital system is assigned a unique hospital record number (HRN), which is used in both these datasets, and so data about each individual could be linked anonymously without the use of any identifying variable. Indeed, age, sex and region of residence were the only personal variables available to the researchers for co-variate analysis (O'Grady et al. 2010b). The Dataset that the Indigenous status information was taken from is not indicated. Selvey (2001) describes some Indigenous identification issues with the NT Childhood Immunisation Database within the time frame of these evaluations (1998–2005), so it may not be the case that the use of the HRN means that other variables such as Indigenous status will also have the same source. Both studies found limited evidence that the universal vaccination program was reducing hospitalisations for pneumonia (O'Grady et al. 2010a; O'Grady et al. 2010b), and there is some evidence that partly and fully vaccinated Aboriginal children were at increased risk of hospitalisation for ALRIs and pneumonia (O'Grady et al. 2010a). O'Grady and others (2010a) describes some possible confounding factors causing this result, and attempted to control for these, notably possible differential access to hospitals, for which they also compared the relationship between vaccine dose and hospitalisation for gastroenteritis (an unrelated condition, and for which increased risk was not found). While startling, the authors were able to point to overseas studies that had also found more respiratory events following vaccination programs with these particular vaccines. This evaluation was only possible because information about doses of vaccines in the immunisation Register was linked with information about principal diagnosis at hospital admission from the NT hospitals data.

The Aboriginal Birth Cohort (ABC) Study, Northern Territory

The development of an Aboriginal birth cohort since 1987 by researchers at the Menzies School of Health Research offers a good example of how the use of alternative names for the same individual can be recorded and monitored reliably (Sayers et al. 2003) (see section 4.4 below). Funding is being sought at the time of writing to link information recorded for the cohort study to Northern Territory datasets that include social variables about education achievement and justice system contact, so that the antecedents to positive outcomes in these areas can be determined in a cost effective way.

Study of Environment on Aboriginal Resilience and Child Health

The Study of Environment on Aboriginal Resilience and Child Health (SEARCH) is a cohort study of Aboriginal and Torres Strait Islander people aged under 18 in urban and regional centres (Williamson et al. 2010, see *Thematic list*, pages 34–5). Participants will be recruited from four different Aboriginal Community Controlled Health Organisations (ACCHOs) in New South Wales. Participants' permission will be sought to link data about them in health

databases such as admitted patients' data. Participants will be followed up through this linkage over a 5–year period to monitor health outcomes other than those being surveyed and to access variables being documented during the study such as height, weight, and blood pressure. Furthermore, linkage may also be a useful tool for following up participants over the duration of the study if they move house or pass away.

4 Other considerations

4.1 Reporting of linkage details

It is useful for details to be published about how a particular linkage was performed (e.g. how many passes, and what linking and blocking variables were used), about the quality of the linkage, and how quality was measured. Unfortunately, the extent of reporting on these aspects is generally poor. It is only when data linkage is being used to improve Indigenous identification that these details tend to be reported at all. If the primary focus of research is something other than Indigenous health, the linkage is treated as a means to an end, so little or no detail about the linkage is provided. Even when details and quality of the linkage are documented, there is no consistency about how this is done or how quality of linkage is measured. This makes it difficult to compare the efficiency and quality of linkage across studies, and the relative usefulness of particular datasets for Indigenous data linkage.

The problem of poor reporting of linkage details is not confined to uses of data linkage related to Aboriginal and Torres Strait Islander people, but rather affects uses of data linkage generally. However, it presents a specific challenge to developing consistent procedures for linking data about Aboriginal and Torres Strait Islander people, particularly if it is to be done consistently between jurisdictions. Bohensky and others (2011) recently developed a set of reporting guidelines for studies using data linkage and built these into a table (Table 1 in) as a useful tool for researchers. They identified 14 aspects of data linkage that could be reported more explicitly, including those referred to above. Analysts could consider making use of this tool when reporting on data linkage studies, and report against the guidelines in this table in addition to the information about Indigenous status data suggested in the *National best practice guidelines for data linkage activities relating to Aboriginal and Torres Strait Islander People*.

4.2 Research into COAG targets and performance indicators for Aboriginal and Torres Strait Islander people

Two of the 6 COAG Closing the Gap targets are directly related to Indigenous health. Seven performance indicators (PIs) each were initially proposed for these two targets (COAG 2008). Indigenous data linkage may assist with enhancing measures for these PIs, so it will be useful to know what experience and knowledge base currently exists that can be tapped into for this purpose. A lack of research in particular areas using Indigenous data linkage may, conversely, suggest that there are difficulties with using Indigenous data linkage for the relevant PIs.

Mortality rate and hospitalisation rate by principal diagnosis were the foundation COAG Closing the Gap performance indicators for which we found the most studies using data linkage that were related to them. Most of the studies relevant to mortality rate were conducted in WA, while more than half of those examining hospitalisation rates were conducted outside of WA. This may indicate that WA deaths data are more suitable for Indigenous data linkage than data from other jurisdictions. Hospital admissions data were used more widely, including for the relatively few studies of lifestyle-related factors.

While the mortality rate for children aged under 5 and mortality rates by leading causes for different age classes under 5, were sometimes addressed separately in some form by studies assessing overall mortality rate, there were no data linkage studies that focussed exclusively on these topics in Indigenous children. This seems surprising, as linking death, perinatal and hospital admissions data may provide valuable insight into which causes of mortality differ between Indigenous and non-Indigenous children. Only one study was found that attempted this; it used New South Wales data only, and did not disaggregate by Indigenous status (Gordon & Jeffery 2008).

Studies at a national level have been related to life expectancy and mortality rate; the only state-specific study of life expectancy has been in Western Australia (Draper et al. 2009). Four Indigenous data linkage studies have examined hospitalisation rates by principal diagnosis exclusively for children; two each from Queensland (Campbell et al. 2009; Kennedy et al. 2009a) and Western Australia (Moore et al. 2010; Moore et al. 2011; Oddy et al. 2008). There are fewer than four examples each of research using Indigenous data linkage that are related to topics covered by the remaining PIs under the two Closing the Gap targets related to health. These examples mostly linked surveys about lifestyle with administrative datasets.

4.3 Distribution of Indigenous data linkage work

The bulk of Indigenous data linkage studies use WA data. This is partly because the Western Australian data linkage infrastructure has been established for much longer than in other jurisdictions and is more comprehensive. However, it is also clear that other components of the advanced Western Australian health information infrastructure facilitate the application of Indigenous data linkage in that state in ways that would be difficult or impossible in other jurisdictions. Notable among these is the WAACHS, and a number of health surveys of Aboriginal and Torres Strait Islander people developed using the WA Community Health Register.

Of the 258 project applications accepted by WADLS from 1995–2003 (WADLS 2010), 12 (4.7%) had a specifically Indigenous focus. While encouraging, this number may not reflect the disproportionate health burden borne by Aboriginal and Torres Strait Islander people, and suggests that more could be done to harness this resource to bring about health equality for Aboriginal and Torres Strait Islander Australians.

The presence of HRNs on all health records for individuals listed in the Northern Territory's Client Master Index (Robbins 1999) allows for relatively straightforward matching of data across different health datasets in the Northern Territory. The Client Master Index has been tapped into for matching records for reporting and some research, but not as much as might be expected given this considerable advantage over other jurisdictions. The Northern Territory Information Act (2002) strictly limits the use of data related to the Client Master Index. Some other constraints on the use of the Client Master Index for reporting or research are described elsewhere (Cunningham et al. 2006, see *Thematic list*, page 12).

Most Queensland studies predate the Queensland PHRN node, which is very new, and hence did not make use of the Queensland PHRN node linked data. Similarly, it is only in WA and New South Wales that their respective PHRN nodes have been the main site of linkage, since these are the only nodes that are well established.

Relatively few Indigenous data linkage studies have been conducted at a national level, with almost all taking place within one specific state or territory. There are some barriers to linking data at a national level. Variables are not used or recorded consistently across all

jurisdictions, not only for Indigenous status but also for linking and analysis variables as well. The use of data from multiple jurisdictions requires permission to be gained from each jurisdiction; delay or denial of data from just one jurisdiction will prevent a genuinely national study from being conducted. Additionally, there are often restrictions on the provision of personal information in these datasets that would be required for linking with other datasets, notably in the case of the perinatal data that would be required to measure COAG PIs about child mortality. If personal information is not provided by just one jurisdiction for one of the datasets, it is not possible to create a truly national linked Dataset.

4.4 What's different about linking data about Aboriginal and Torres Strait Islander Australians

Generally, the challenges faced when linking data about Aboriginal and Torres Strait Islander Australians are no different to those for anyone else. However, there are a few aspects that may, on the basis of the current evidence, affect data about Aboriginal and Torres Strait Islander peoples more so than other Australians.

People born overseas misclassified as Indigenous

This is a particular issue in datasets where the data are recorded by someone other than the person the data are about, especially if the recorder does not know the participant well. Among datasets useful for COAG PIs, the most important Dataset affected may be midwives' notification data (i.e. perinatal data), where data are recorded by the midwife. Researchers at the Telethon Institute of Child Health Research have identified some of these issues in the course of using linked data to examining perinatal health and wellbeing in immigrant women in WA (von Katterfeld 2010) and in preparing linked data about Aboriginal people for analysis (for example, Langridge et al. 2010, Langridge 2010). However, the nature and extent of this problem has yet to be systematically explored in relevant datasets. It may be useful for analysts to quantify the impact of this issue if the datasets they are using are likely to be affected.

A mother's country of birth in midwives' data could, for example, be recorded as 'unknown' or a specified overseas country, but then mistakenly recorded as 'Indigenous' in the midwives' data or in another Dataset. Birth registration data records father's country of birth as well as mother's country of birth, and could also be used to exclude people mistakenly recorded as Indigenous in other datasets when both their parents are born overseas. However, there may well be instances where children are born overseas to Aboriginal and Torres Strait Islander people who have either left Australia or been removed previously through adoption, or where an Indigenous person has one Indigenous parent and one immigrant parent.

Related to this issue is the misclassification as 'Indigenous' or 'Aboriginal' of people who are indigenous or aboriginal to countries other than Australia, or as 'Torres Strait Islander' of people who identify as being of islands other than the Torres Strait (e.g. Pacific Islanders). Use of the standard Indigenous identification question may have reduced the instance of misidentification of this kind, but cannot guarantee its complete absence. How often this misclassification occurs does not appear to have been quantified in datasets. Some of these

records may also be excluded from analysis by excluding records of people born overseas who are recorded as being Indigenous.

Different addresses between two datasets as a result of greater mobility

Address or components of a person's address are often used as linkage variables. If a person has changed address during a period in which data is being collected about them, either in the same Dataset or in two different datasets, then using address as a linkage variable will fail to link their records.

Memmott and others (2004) found that Aboriginal and Torres Strait Islander people were more likely than non-Indigenous people to engage in repeated movements at all ages. The repercussions of this for linking data about Aboriginal and Torres Strait Islander people are yet to be investigated systematically. At the very least, if Aboriginal and Torres Strait Islander people are more likely to change address between contacts with datasets than non-Indigenous people, then linking using addresses would be more likely to fail to link Aboriginal and Torres Strait Islander people than non-Indigenous people.

The problem for data linkage may not be so straightforward. Memmott and others (2004) described and investigated a number of aspects of Indigenous mobility; they found that Indigenous mobility was complex, with great differences in the pattern of mobility between different areas, cultural groups, and age groups. Additionally, Memmott and others (2004) state that 'ABS (Census) methodology assumes that households occupy one place of residence whereas there is strong evidence in remote Aboriginal communities of linked households or clustered households that are characterised by an extended family group dispersed across a number of places of residence'. In these circumstances, people may move between several residences and not regard themselves as having a single usual place of residence (Memmott 2011). The use of a single address as a personal identifier is the norm in many datasets in Australia, including the Census, but this may be a poor fit for the cultural practice of clustered households.

The proportion of Aboriginal and Torres Strait Islander people who change address between time points in a given Dataset is rarely quantified. It is interesting to note that when evaluating the Aboriginal Birth Cohort Study in the Northern Territory, Sayers and others (2003) found that 18% of children in the cohort had changed address by the age of 4. This rate of changed address over only 4 years suggests that, for at least the area covered by the Aboriginal Birth Cohort Study, this issue is not minor.

Users of linked data should be aware of these mobility issues for data linkage, and should try to quantify them. The report of Memmott and others (2004), *Between places: Indigenous mobility in remote and rural Australia* may be a useful resource for this. Importantly, Memmott and others (2004) used 2001 Census data for their investigation, and reviewed research into Indigenous mobility using the 1986, 1991, and 1996 Censuses. The ABS Indigenous Mortality Quality Study (ABS 2008a) identified poor quality of personal information as a major contributor to under-counting of Aboriginal and Torres Strait Islander people by the 2006 Census; poor quality of address variables due to mobility may also be a contributing factor to this problem. While the effect of Indigenous mobility on data linkage has not been assessed in Australian data, Blakeley and others (2000) found mobility to be an important contributor to poor linkage of data about Maori peoples in New Zealand data. The approach of Blakeley

and others (2000) may be a useful model for assessing the impact of Indigenous mobility on Australian linked data.

Population estimates derived from the Census are intended as the denominator for many COAG Closing the Gap indicator measures. The impact of this mobility issue on the Census data, and on its use as the denominator for these measures, while not yet quantified, is likely to be profound for some specific subgroups within the Aboriginal and Torres Islander population.

The 2006 Indigenous Mortality Quality Study, which found that more than one quarter of Indigenous death records could not be linked to the 2006 Census, determined that the predominant reasons for not linking death records were failure to count people in the Census and poor quality of the linkage variables such as address (ABS 2008a). Failure to count 'visitors' not at their 'usual' address, and address variables different from those at other data collection points, may in part be manifestations in the Census data of the practice of regularly moving between several residences.

The *Guidelines* describe the problem of numerator–denominator bias. Data linkage to count Aboriginal and Torres Strait Islander people for Closing the Gap measure numerators will increase the accuracy of the numerator, but exacerbate numerator–denominator bias when Census–derived estimates are used as the denominator, even when these estimates are adjusted for under-count and under-identification. There is limited access to Census address variables, even within the ABS, to address this problem using data linkage. It may therefore be necessary to consider alternative ways to derive Closing the Gap measure denominators other than from ABS data, at least in regard to the Aboriginal and Torres Strait Islander population size.

Name changes and aliases

Aboriginal and Torres Strait Islander people may use different names at various times and may have specific cultural practices relating to changing their names, such as changing the name of children to that of a recently deceased member of the community after the latter has died, or ceasing to use a name also used by someone who has died. These cultural practices are not uniform across all Aboriginal and Torres Strait Islander people, but may be common in the places they occur. Sayers and others (2003) checked personal details for children enlisted in the Aboriginal Birth Cohort (ABC) Study, which covers 686 children born to Indigenous mothers between January 1987 and March 1990 at Royal Darwin Hospital, and found that by age 4, over one-third (34.4%) of children tracked had changed their birth name at least once. The ABC researchers have handled this issue by collecting alternate names as a separate variable as they hear about them in the course of contacting and following up participants, and maintaining a single record with all aliases recorded.

For health data linkage the ABC researchers rely on the unique Hospital Registration Number. A HRN is allocated to all people born in the Northern Territory, or who have received services at any Northern Territory health service and is used for all subsequent contacts. This HRN allows for deterministic matching. However, for non-health related data linkage, probabilistic matching is needed using the multiple identifiers of current name, aliases, date of birth, place of birth, usual place of residence, mother's name, and carer's name.

There is considerable cultural diversity in the use of alternate names in both the Indigenous and non-Indigenous community, which has presented a problem for data collectors and

custodians. The solutions used to address this problem are just as diverse. It is beyond the scope of the *Guidelines* to document the full range of solutions. However, it is worthwhile for data linkers and analysts to be aware of this issue and to explore the efforts data custodians may have made to record alternate names, since these efforts may have produced useful resources for data linkage. For example, the Northern Territory hospital system maintains lists of aliases with each patient record, each of which has a unique HRN. Queensland's Vaccination Information/Vaccination Administration System (VIVAS) deals with alternate names by creating a new record with the 'new' name for a person while recording this 'new' name in an alias field in the old record. In this way, the new record could be linked to the old record for the same individual.

Children with non-Indigenous mothers and Indigenous fathers

Midwives' notification data (i.e. perinatal data) often does not record details about the father of a newborn, including Indigenous status. If using perinatal data alone, it is therefore not possible to distinguish between babies with an Indigenous father or a non-Indigenous father if the baby's mother is non-Indigenous. This is a substantial problem for identifying Indigenous babies in perinatal data. Kennedy and others (2009a) found that more than one-quarter (29.1%) of Indigenous babies in Queensland births data born between 2003 and 2006 had a non-Indigenous mother and an Indigenous father. In 2007, the ABS reported that 29.3% of Indigenous babies registered as being born in 2006, had an Indigenous father and a mother whose Indigenous status was either not stated or non-Indigenous; the comparable figure for Victoria alone was 41.3% (ABS 2007).

Indigenous babies with non-Indigenous mothers in perinatal data could be detected by linking perinatal Datasets to datasets that record either Indigenous status of the baby or Indigenous status of the father. The most logical Dataset may appear to be birth registration data, yet this is not straightforward. Since 2006, New South Wales births registration data for babies with a non-Indigenous mother have only recorded the Indigenous status of the father as either non-Indigenous or not stated (ABS 2010a). Hence there are no currently available complete national figures for Indigenous births to non-Indigenous mothers later than 2006. Moreover, there is some evidence that registration of births of Indigenous babies can be delayed, sometimes by years, in numbers disproportionately greater than for non-Indigenous babies (see below).

In some instances, birth registration data have not been available for linkage with perinatal data, so other datasets containing Indigenous status of fathers have been used instead. Kennedy and others (2009a), for example, linked hospital morbidity data to assess Indigenous babies with non-Indigenous mothers in Queensland perinatal data (see *Thematic list*, pages 29–30). However, there may be few other options available. Heffernan and others (2009) found that of 20 Victorian datasets, only birth registrations, perinatal death registrations (separate from regular death registrations), and the Aboriginal Hospital Liaison Officer Collection contained Indigenous status of fathers; only the latter Dataset had Indigenous status of babies, mothers and fathers.

Delays in being entered into datasets – notably, registering births

Delays in recording data about an individual may affect the use of a Dataset as a source to link to data about that individual. This will be an issue for linked data about Aboriginal and Torres Strait Islander people if delays in recording disproportionately affect Aboriginal and Torres Strait Islander people compared to other Australians. This has been shown to be the

case in birth registrations (ABS 2007, 2010a). In examining delays in registration of births in 2006, the ABS found that while 12.1% of all births registered in 2006 occurred before 2005, almost one-quarter (23.9%) of Indigenous births registered in 2006 occurred before 2005 (ABS 2007). More recently, while 88.1% of all birth registrations in 2009 were for births in 2009, and 1.0% were for births that occurred before 2004, for Aboriginal and Torres Strait Islander Australians, only 73.4% of birth registrations in 2009 were for births in 2009, and 6.0% were for Indigenous births prior to 2004 (ABS 2010a). The proportion of delays in registering Indigenous births also differs a great deal between states and territories. In 2009, 91.6% of birth registrations in the Northern Territory for Indigenous births were for births in 2009, while only 56.0% of birth registrations in Queensland for Indigenous births were for births in 2009 (ABS 2010a). Under-registration and the delay in registering Indigenous births has been shown to distort calculations of fertility rates (ABS 2010a), and is thus likely to affect other statistics that rely on birth registration data including, potentially, COAG PIs.

Incomplete information in birth records may disproportionately affect Aboriginal and Torres Strait Islander people, especially incomplete Indigenous status. The Queensland RBDM embarked on a comprehensive exercise to complete data in incomplete birth registrations (which were previously not counted in Queensland births data), called the Retrospective Births Project (ABS 2010a). This project completed 1,780 registrations, of which approximately 40% were Indigenous (ABS 2010a).

These documented delays in registering births mean that births registration data may not include as many births which may be recorded in other datasets as might be expected, even for births that are not recent. Moreover, births of Indigenous babies are more likely to be missing from births data than non-Indigenous births, which may contribute to under counting of Aboriginal and Torres Strait Islander people if birth registrations are being linked in order to improve Indigenous identification in other datasets.

Delayed recording of data may differentially affect Aboriginal and Torres Strait Islander people in other datasets as well. Cunningham and Paradies (2000) previously found that 15% of Indigenous deaths were registered late (i.e. after 1998) during 1995–97 compared to 4% for non-Indigenous deaths in combined data from Western Australia, South Australia and the Northern Territory. It may be useful for data linkers and analysts to determine if delayed recording in the datasets they are using has been previously assessed by Indigenous status.

Incomplete date of birth or 'dummy' dates of birth

In some datasets, a disproportionate number of records within which the date of birth is partly or completely missing, or else is a 'dummy date' (an approximation of date of birth) substituted for an unknown actual birthdate, may be for Aboriginal and Torres Strait Islander people. This is well–illustrated in a 1997 audit of the Northern Territory Hospital Morbidity Dataset (Condon et al. 1998, see *Thematic list*, page 14). Indigenous status recorded in this Dataset was highly consistent with responses gained from patients when they (or for children, their parents) were personally interviewed (94%). However, date of birth was not completely recorded at interview for 74 out of 400 patients; 71 of these were Aboriginal patients. Fifteen of these did provide a year of birth. Despite this, all 74 had a date of birth recorded in the hospital morbidity Dataset. In seeking to explain these results, Condon and others (1998) point out that many older Aboriginal people, notably in remote communities, may not know their full date of birth, since 'dates were not used in their society at the time of their birth'. However, they found that the date of birth recorded in the hospital morbidity Dataset for these 74 patients indicated the patients were of all ages, including 14 children

aged less than 3-years of age. The age of these patients suggests that this may remain an issue in at least some datasets today.

This issue is not confined to the Northern Territory. Large numbers of records of Aboriginal and Torres Strait Islander people in the Residential Aged Care and National Hospital Morbidity Datasets have also been found to have dates of birth typically used as dummy dates, see *Guidelines*, section 3.2 (AIHW 2012; AIHW & ABS 2012).

For datasets containing records with incomplete or dummy dates of birth, linkage using components of date of birth as linking variables may fail to link a disproportionately high number of Aboriginal and Torres Strait Islander people. It may be useful for the candidate datasets to be assessed for incomplete dates of birth, and to see if there are disproportionately large numbers of births on typically chosen dummy dates such as 1 January or 1 July. These aspects could be quantified and interaction with recorded Indigenous status could be analysed. Using just the year of birth may reduce the number of missed links due to this problem. However, it is possible that uncertainty about a person's date of birth may result in them having different dates of birth in their records between or even within datasets.

5 Conclusion

The promise of data linkage as a tool for informing how the gap can be closed has been noticed by many, and is driving some exciting developments. Some useful lessons for data linkage that could assist the measurement of COAG Closing the Gap Performance Indicators can be learnt from the applications of data linkage described in this review:

1. The use of survey data as a 'gold standard' for Indigenous status information when linking with other datasets has great potential.

2. To date, data linkage has rarely been used for exploring topics related to COAG Closing the Gap Performance Indicators. This is especially so for child mortality, smoking, some other lifestyle factors, antenatal care, and access to health care compared to need.

3. There has been little consistency about how Indigenous data linkage has been done or evaluated.

4. There is a small evidence base for the relative merits of different approaches to identifying Aboriginal and Torres Strait Islander people through data linkage.

5. The existing evidence base may not be universally applicable across all Australian jurisdictions and datasets.

However, an emphasis on the use of data linkage to improve accuracy of Indigenous status in datasets is beginning to emerge. This trend is particularly evident for death, perinatal, birth and hospital data. These datasets are often used for government program or policy performance indicators and, to a large extent, already have a standardised Indigenous status question. Additionally, these datasets will have already been assessed for the quality of their Indigenous status data. datasets with Indigenous status data enhanced through data linkage are being established in New South Wales and at the AIHW.

There are still great differences between states and territories in the amount of work using linked data about Aboriginal and Torres Strait Islander people. This partly reflects how long data linkage nodes have existed in each jurisdiction. The establishment of CHeReL years earlier than most PHRN nodes was a catalyst for innovative research and analysis in New South Wales. PHRN nodes in other states are also starting to produce new insights through data linkage.

It should be noted that the differences in legislative regimes between states and territories appears to be affecting how easy it is to establish data linkage projects regarding Aboriginal and Torres Strait Islander people, even when funding and expertise are available. Only the AIHW and ABS are currently linking data related to Aboriginal and Torres Strait Islander people on a national scale.

In moving forward toward a consistent approach to Indigenous data linkage, an evidence base needs to be developed that is broadly comparable. The *Guidelines* suggest approaches to Indigenous data linkage, which, if followed, will allow a robust evidence base for best practice to be built as quickly as possible.

References

ABS 2007. Births, 2006, Australia. ABS Cat. no. 3301.0. Canberra: Australian Bureau of Statistics.

ABS 2008a. Information Paper: Census Data Enhancement - Indigenous Mortality Quality Study. ABS Cat. no. 4723.0. Canberra: ABS.

ABS 2008b. Technical Note: Estimated Aboriginal and Torres Strait Islander Australian Resident Population - method of calculation. 19/08/2008. Canberra, Australia. Viewed 24/10/2011,
<http://www.abs.gov.au/ausstats/abs@.nsf/Products/3238.0.55.001~Jun+2006~Technical+Note~Estimated+Aboriginal+and+Torres+Strait+Islander+Australian+Resident+Population+%E2%80%93+Method+of+Calculation+(Technical+Note)?OpenDocument>.

ABS 2009. Experimental Life Tables for Aboriginal and Torres Strait Islander Australians 2005-07. ABS Cat. no. 3302.0.55.003. Canberra: ABS.

ABS 2010a. Births, 2009, Australia. ABS Cat. no. 3301.0. Canberra: ABS.

ABS 2010b. Census Data Enhancement Project: An Update. 15/10/2010. Canberra, Australia: Australian Bureau of Statistics. Viewed 28 April 2011 2011,
<http://www.abs.gov.au/ausstats/abs@.nsf/mf/2062.0>.

ABS 2010c. Linking Census Records to Death Registrations. ABS cat. no. 1351.0.55.030. Canberra: ABS.

ABS 2011a. How Australia Takes a Census, 2011: Collection. 5 May 2011. Viewed 28 February 2012,
<http://www.abs.gov.au/ausstats/abs@.nsf/lookup/2903.0Main%20Features82011>.

ABS 2011b. How Australia Takes a Census, 2011: Planning the 2011 Census. 5 May 2011. Viewed 28 February 2012,
<http://www.abs.gov.au/ausstats/abs@.nsf/lookup/2903.0Main%20Features42011>.

AIHW 2009. Australia's mothers and babies 2007. Canberra: Australian Institute of Health and Welfare.

AIHW 2010. Indigenous identification in hospital separations data: quality report. Health Service Series No 35, Cat. no. HSE 85. Canberra: AIHW.

AIHW 2012. An enhanced mortality database for estimating Indigenous life expectancy: A feasibility study. Cat. no. IHW 75. Canberra: AIHW.

AIHW & ABS 2012. National best practice guidelines for data linkage activities relating to Aboriginal and Torres Strait Islander people: 2012. Canberra: AIHW.

AIHW: Laws P, Abeywardana S, Walker J & Sullivan E 2007. Australia's mothers and babies 2005. Canberra: AIHW.

Alessandri LM, Chambers HM, Garfield C, Vukovich S & Read AW 1999. Cumulative mortality in children aged 1 to 6 years born in Western Australia from 1980-89. Archives of Disease in Childhood 80:15.

Allen & Clarke 2011. Evaluation of the Child Health Check Initiative and the Expanding Health Service Delivery Initiative: Final report. Canberra: Department of Health and Ageing.

Anderson P & Wild R 2007. Ampe Akelyernemane Meke Mekarle - Little Children are Sacred. Northern Territory Board of Inquiry into the Protection of Aboriginal Children from Sexual Abuse.

Blakely T, Woodward A & Salmond C 2000. Anonymous linkage of New Zealand mortality and Census data. Australian and New Zealand Journal of Public Health 24:92-5.

Bohensky MA, Jolley D, Sundararajan V, Evans S, Ibrahim J & Brand C 2011. Development and validation of reporting guidelines for studies involving data linkage. Australian and New Zealand Journal of Public Health 35:486-9.

Bradshaw PJ 2010. Personal communication. August 2010.

Bradshaw PJ, Alfonso HS, Finn J, Owen J & Thompson PL 2009. Measuring the gap: accuracy of the Western Australian hospital morbidity data in the identification of adult urban Aboriginal and Torres Strait Islander people. Australian and New Zealand Journal of Public Health 33:276-9.

Bradshaw PJ, Alfonso HS, Finn J, Owen J & Thompson PL 2010a. A comparison of coronary heart disease event rates among urban Australian Aboriginal people and a matched non-Aboriginal population. Journal of Epidemiology and Community Health.

Bradshaw PJ, Alfonso HS, Finn J, Owen J & Thompson PL 2010b. The use of coronary revascularisation procedures in urban Australian Aboriginals and a matched general population coronary procedures in Aboriginals. Heart Lung and Circulation 19:247-50.

Brameld K, Holman D, Thomas M & Bass J 1999. Use of a state data bank to measure incidence and prevalence of a chronic disease: end-stage renal failure. American Journal of Kidney Diseases 34:1033-9.

Briffa TG, Sanfilippo FM, Hobbs MS, Ridout SC, Katzenellenbogen JM, Thompson PL et al. 2010. Under-ascertainment of Aboriginality in records of cardiovascular disease in hospital morbidity and mortality data in Western Australia: a record linkage study. BMC Medical Research Methodology 10:111.

Burke V, Zhao Y, Lee AH, Hunter E, Spargo RM, Gracey M et al. 2007. Health-related behaviours as predictors of mortality and morbidity in Australian Aborigines. Preventive Medicine 44:135-42.

Burns L & Mattick RP 2007. Using population data to examine the prevalence and correlates of neonatal abstinence syndrome. Drug and Alcohol Review 26:487-92.

Burns L, Mattick RP & Cooke M 2006a. Use of record linkage to examine alcohol use in pregnancy. Alcoholism-Clinical and Experimental Research 30:642-8.

Burns L, Mattick RP & Cooke M 2006b. The use of record linkage to examine illicit drug use in pregnancy. Addiction 101:873-82.

Calver J, Wiltshire A, Holman CD, Hunter E, Garfield C & Rosman DL 2005. Does health assessment improve health outcomes in indigenous people? An RCT with 13 years of follow-up. Australian and New Zealand Journal of Public Health 29:107-11.

Campbell SJ, Nissen MD & Lambert SB 2009. Rotavirus epidemiology in Queensland during the pre-vaccine era. Communicable Diseases Intelligence 33:204-8.

Carson BE, Guthridge SL, Li SQ & Measey ML 2006. Growing up in the Territory: parent survey. Darwin: Department of Health and Community Services.

Carson D 1994. The nature of movement in to and out of the Northern Territory as shown by the 1991 census. Cat. no. 3402.7. Darwin: Australian Bureau of Statistics.

Carville KS, Lehmann D, Hall G, Moore H, Richmond P, de Klerk N et al. 2007. Infection is the major component of the disease burden in Aboriginal and non-Aboriginal Australian children - A population-based study. Pediatric Infectious Disease Journal 26:210-6.

CHeReL 2006. Published studies and other projects using linked data in NSW. Sydney, NSW: Centre for Health Record Linkage.

CHeReL 2008. Centre for Health Record Linkage Record Linkage Outputs. Sydney, NSW: Centre for Health Record Linkage.

COAG 2008. National Indigenous Reform Agreement (Closing the Gap), Schedule F. COAG.

Condon J 2011. Personal communication.

Condon J, Barnes A, Cunningham J & Smith L 2004. Demographic Characteristics and Trends of the Northern Territory Indigenous Population 1966 to 2001. Darwin: Cooperative Research Centre for Aboriginal Health.

Condon J, Williams DJ, Pearce MC & Moss E 1998. Northern Territory hospital morbidity dataset: validation of demographic data 1997. Darwin: Northern Territory Health Services.

Coory M & Johnston T 2006. Trends in the annual prevalence of hospitalisation for remote indigenous communities in Queensland, 1997/98 to 2004/05. Australian and New Zealand Journal of Public Health 30:440-3.

Coory M, Scott IA & Baade P 2002. Differential effect of socioeconomic status on rates of invasive coronary procedures across the public and private sectors in Queensland, Australia. Journal of Epidemiology and Community Health 56:233-4.

Coory MD & Walsh WF 2005. Rates of percutaneous coronary interventions and bypass surgery after acute myocardial infarction in Indigenous patients. Medical Journal of Australia 182:507.

Cunningham J, O'Dea K, Dunbar T, Weeramanthri T, Zimmet P & Shaw J 2006. Study protocol--diabetes and related conditions in urban indigenous people in the Darwin, Australia region: aims, methods and participation in the DRUID Study. BMC Public Health 6:8.

Cunningham J & Paradies Y 2000. Mortality of Aboriginal and Torres Strait Islander Australians. ABS cat. no. 3315.0. Australian Bureau of Statistics.

DoHA & AIHW 2008. Progress of the Northern Territory Emergency Response Child Health Check Initiative: health conditions and referrals. Canberra: AIHW.

Draper G 2010. Personal communication.

Draper G, Unwin E, Serafino S, Somerford P & Price S 2005. Health Measures 2005: A report on the health of the people of Western Australia. Perth: Department of Health WA.

Draper GK, Somerford PJ, Pilkington AS & Thompson SC 2009. What is the impact of missing Indigenous status on mortality estimates? An assessment using record linkage in Western Australia. Australian and New Zealand Journal of Public Health 33:325-31.

Freemantle CJ, Read AW, Klerk NHd, McAullay D, Anderson IP & Stanley FJ 2006. Patterns, trends, and increasing disparities in mortality for Aboriginal and non-Aboriginal infants born in Western Australia, 1980-2001: population database study. The Lancet 367:1758.

Freemantle J, Stanley F, Read A & de Klerk N 2004. The First Research Report: patterns and trends in mortality of Western Australian infants, children and young people 1980-2002. The Department for Community Development, Government of Western Australia.

Glasson EJ, de Klerk NH, Bass AJ, Rosman DL, Palmer LJ & Holman CD 2008. Cohort profile: The Western Australian Family Connections Genealogical Project. International Journal of Epidemiology 37:30-5.

Glasson EJ, Sullivan SG, Hussain R & Bittles AH 2005. An assessment of intellectual disability among Aboriginal Australians. Journal of Intellectual Disability Research 49:626-34.

Gordon A & Jeffery HE 2008. Classification and description of stillbirths in New South Wales, 2002-2004. Medical Journal of Australia 188:645-8.

Hall SE & Holman CD 2003. Inequalities in breast cancer reconstructive surgery according to social and locational status in Western Australia. European Journal of Surgical Oncology 29:519-25.

Haynes K & Stone C 2004. Predictors of incomplete immunisation in Victorian children. Australian and New Zealand Journal of Public Health 28:72-9.

Heffernan B, Sheridan S & Freemantle CJ 2009. An Overview of Statutory and Administrative Datasets: Describing the Health of Victoria's Aboriginal Infants, Children and Young People. Melbourne: Onemda VicHealth Koori Health Unit, The University of Melbourne.

Holman CD, Wisniewski ZS, Semmens JB, Rouse IL & Bass AJ 2000. Population-based outcomes after 28,246 in-hospital vasectomies and 1,902 vasovasostomies in Western Australia. BJU International 86:1043-9.

Holman RC, Curns AT, Cheek JE, Singleton RJ, Anderson LJ & Pinner RW 2003. Infectious disease hospitalizations among American Indian and Alaska native infants. Pediatrics 111:E176-82.

Kariminia A, Butler T, Corben S, Kaldor J, Levy M & Law M 2005. Mortality among prisoners: how accurate is the Australian National Death Index? Australian and New Zealand Journal of Public Health 29:572-5.

Kariminia A, Butler TG, Corben SP, Levy MH, Grant L, Kaldor JM et al. 2007a. Extreme cause-specific mortality in a cohort of adult prisoners--1988 to 2002: a data-linkage study. International Journal of Epidemiology 36:310-6.

Kariminia A, Law MG, Butler TG, Corben SP, Levy MH, Kaldor JM et al. 2007b. Factors associated with mortality in a cohort of Australian prisoners. European Journal of Epidemiology 22:417.

Katzenellenbogen JM, Sanfilippo FM, Hobbs MS, Briffa TG, Ridout SC, Knuiman MW et al. 2010. Incidence of and case fatality following acute myocardial infarction in Aboriginal and non-Aboriginal Western Australians (2000-2004): a linked data study. Heart Lung and Circulation 19:717-25.

Katzenellenbogen JM, Vos T, Somerford P, Begg S, Semmens JB & Codde JP 2011. Burden of Stroke in Indigenous Western Australians: A Study Using Data Linkage. Stroke 42:1515-21.

Kennedy B, Cornes S, Martyn S, Wills R & Breckell C 2009a. Measuring Indigenous perinatal outcomes – should we use the Indigenous status of the mother, father or baby? Health Statistics Centre, Queensland Health.

Kennedy B, Howell S & Breckell C 2009b. Indigenous identification in administrative data collections and the implications for reporting Indigenous health status. Brisbane: Health Statistics Centre, Queensland Health.

Kliewer EV & Stanley FJ 1993. Stillbirths, neonatal and post-neonatal mortality by race, birthweight and gestational age. Journal of Paediatrics and Child Health 29:43-50.

Langridge AT 2010. Personal communication.

Langridge AT, Nassar N, Li J & Stanley FJ 2010. Social and racial inequalities in preterm births in Western Australia, 1984 to 2006. Paediatric and Perinatal Epidemiology 24:352-62.

Lawrence D, Jablensky AV, Holman CDJ & Pinder TJ 2000. Mortality in Western Australian psychiatric patients. Social Psychiatry and Psychiatric Epidemiology 35:341.

Lee AH, Flexman J, Wang K & Yau KK 2004a. Recurrent gastroenteritis among infants in Western Australia: a seven-year hospital-based cohort study. Annals of Epidemiology 14:137-42.

Lee AH, Somerford PJ & Yau KK 2003. Factors influencing survival after stroke in Western Australia. Medical Journal of Australia 179:289-93.

Lee AH, Somerford PJ & Yau KK 2004b. Risk factors for ischaemic stroke recurrence after hospitalisation. Medical Journal of Australia 181:244-6.

Leonard H, Nassar N, Bourke J, Blair E, Mulroy S, de Klerk N et al. 2008. Relation between intrauterine growth and subsequent intellectual disability in a ten-year population cohort of children in Western Australia. American Journal of Epidemiology 167:103-11.

Leonard H, Petterson B, Bower C & Sanders R 2003. Prevalence of intellectual disability in Western Australia. Paediatric and Perinatal Epidemiology 17:58-67.

Leonard H, Petterson B, De Klerk N, Zubrick SR, Glasson E, Sanders R et al. 2005. Association of sociodemographic characteristics of children with intellectual disability in Western Australia. Social Science and Medicine 60:1499-513.

Li SQ, Cass A & Cunningham J 2003. Cause of death in patients with end-stage renal disease: assessing concordance of death certificates with registry reports. Australian and New Zealand Journal of Public Health 27:419-24.

Li SQ, Jacklyn SP, Carson BE, Guthridge S & Measey ML 2006. Growing up in the Territory: Social-emotional Wellbeing and Learning Outcomes. Darwin: Department of Health and Community Services.

Mak DB & Watkins RE 2008. Improving the accuracy of Aboriginal and non-Aboriginal disease notification rates using data linkage. BMC Health Services Research 8:118.

Martin JH, Coory MD, Valery PC & Green AC 2009. Association of diabetes with survival among cohorts of Indigenous and non-Indigenous Australians with cancer. Cancer Causes and Control 20:355-60.

Memmott P 2011. Personal communication. 5 October.

Memmott P, Long S, Bell M, Taylor J & Brown D 2004. Between places: Indigenous mobility in remote and rural Australia. AHURI Positioning Paper No. 81. Melbourne: Australian Housing and Urban Research Institute, Queensland Research Centre, RMIT University.

Memmott P, Long S & Thomson L 2006. Indigenous mobility in rural and remote Australia. AHURI Final Report No. 90. Melbourne: Australian Housing and Urban Research Institute, Queensland Research Centre, RMIT University.

Meuleners L, Hendrie D & Lee A 2008. Hospitalisations due to interpersonal violence: a population-based study in Western Australia. Medical Journal of Australia 188:572.

Meuleners L, Lee A & Hendrie D 2009. Effects of demographic variables on mental illness admission for victims of interpersonal violence. Journal of Public Health 31:162.

Moore H, Burgner D, Carville K, Jacoby P, Richmond P & Lehmann D 2007. Diverging trends for lower respiratory infections in non-Aboriginal and Aboriginal children. Journal of Paediatrics and Child Health 43:451-7.

Moore HC, de Klerk N, Richmond P & Lehmann D 2010. A retrospective population-based cohort study identifying target areas for prevention of acute lower respiratory infections in children. BMC Public Health 10:757.

Moore HC, Lehmann D, de Klerk N, Jacoby P & Richmond PC 2011. Reduction in disparity for pneumonia hospitalisations between Australian indigenous and non-Indigenous children. Journal of Epidemiology and Community Health.

Neville SE, Taylor LK, Moore H, Madden R, Ring I, Pulver LJ et al. 2011. Using linkage between hospital and ABS mortality data to enhance reporting of deaths among Aboriginal and Torres Strait Islander peoples. Australian and New Zealand Journal of Public Health.

O'Brien J 2010. Personal communication.

O'Donnell M 2011. Personal communication.

O'Donnell M, Nassar N, Leonard H, Jacoby P, Mathews R, Patterson Y et al. 2010a. Characteristics of non-Aboriginal and Aboriginal children and families with substantiated child maltreatment: a population-based study. International Journal of Epidemiology 39:921-8.

O'Donnell M, Nassar N, Leonard H, Mathews R, Patterson Y & Stanley FJ 2010b. Monitoring child abuse and neglect at a population level: Patterns of hospital admissions for maltreatment and assault. Child Abuse and Neglect 34(11):823-32.

O'Grady KA, Lee KJ, Carlin JB, Torzillo PJ, Chang AB, Mulholland EK et al. 2010a. Increased risk of hospitalization for acute lower respiratory tract infection among Australian indigenous infants 5-23 months of age following pneumococcal vaccination: a cohort study. Clinical Infectious Diseases 50:970-8.

O'Grady KF, Carlin JB, Chang AB, Torzillo PJ, Nolan TM, Ruben A et al. 2010b. Effectiveness of 7-valent pneumococcal conjugate vaccine against radiologically diagnosed pneumonia in indigenous infants in Australia. Bulletin of the World Health Organization 88:139-46.

Oddy WH, Kickett-Tucker C, De Maio J, Lawrence D, Cox A, Silburn SR et al. 2008. The association of infant feeding with parent-reported infections and hospitalisations in the West Australian Aboriginal Child Health Survey. Australian and New Zealand Journal of Public Health 32:207-15.

Petterson B, Leonard H, Bourke J, Sanders R, Chalmers R, Jacoby P et al. 2005. IDEA (Intellectual Disability Exploring Answers): A population-based database for intellectual disability in Western Australia. Annals of Human Biology 32:237-43.

Read AW, Gibbins J & Stanley FJ 1996. Hospital admissions for lower respiratory tract illness before the age of two years in western Australia. Paediatric and Perinatal Epidemiology 10:175-85.

Read AW, Gibbins J, Stanley FJ & Morich P 1994. Hospital admissions before the age of 2 years in Western Australia. Archives of Disease in Childhood 70:205-10.

Robbins J 1999. The Northern Territory Client Master Index. Health Information Management Journal 29:35-7.

Rosenwax L & McNamara B 2006. Who receives specialist palliative care in Western Australia - and who misses out. Palliative Medicine 20:439.

Rosman DL, Ferrante AM & Marom Y 2001. A linkage study of Western Australian drink driving arrests and road crash records. Accident Analysis and Prevention 33:211-20.

Sayers SM, Mackerras D, Singh G, Bucens I, Flynn K & Reid A 2003. An Australian Aboriginal birth cohort: a unique resource for a life course study of an Indigenous population. A study protocol. BMC International Health and Human Rights 3:1.

Schultz R, Read AW, Straton JA, Stanley FJ & Morich P 1991. Genitourinary tract infections in pregnancy and low birth weight: case-control study in Australian aboriginal women. British Medical Journal 303:1369-73.

Selvey CE 2001. Conjugate pneumococcal vaccine coverage. Northern Territory Disease Control Bulletin 8:29-31.

Semmens JB, Lawrence-Brown MM, Fletcher DR, Rouse IL & Holman CD 1998. The Quality of Surgical Care Project: a model to evaluate surgical outcomes in Western Australia using population-based record linkage. Australian and New Zealand Journal of Surgery 68:397-403.

Shepherd C 2011. Personal communication.

Silburn SR, Blair E, Griffin JA, Zubrick SR, Lawrence DM, Mitrou FG et al. 2007. Developmental and environmental factors supporting the health and well-being of Aboriginal adolescents. International Journal of Adolescent Medicine and Health 19:345-54.

Silburn SR, Zubrick SR, De Maio JA, Shepherd C, Griffin JA, Mitrou FG et al. <Telethon Institute for Child Health Research> 2006. The Western Australian Aboriginal Child Health Survey: Strengthening the capacity of Aboriginal children, families and communities. Perth. Viewed October 2010, <http://www.ichr.uwa.edu.au/waachs/publications/volume_four>.

Silburn SR, Zubrick SR, Garton A, Gurrin LC, Burton PR, Dalby RB et al. 1996. Western Australian Child Health Survey: Family and Community Health. ISBN 0 642 20759 3). Perth, Western Australia: Australian Bureau of Statistics and the TVW Telethon Institute for Child Health Research.

Stanley FJ, Croft ML, Gibbins J & Read AW 1994. A population database for maternal and child health research in Western Australia using record linkage. Paediatric and Perinatal Epidemiology 8:433-47.

Stanley FJ, Read AW, Kurinczuk JJ, Croft ML & Bower C 1997. A population maternal and child health research database for research and policy evaluation in Western Australia. Seminars in Neonatology 2:195-201.

Sullivan M 2010. Personal communication.

Taylor L & Bejuk B 2009. New South Wales Mothers and Babies 2006. NSW Department of Health.

Valery PC 2010. Personal communication.

Valery PC, Coory M, Stirling J & Green AC 2006. Cancer diagnosis, treatment, and survival in Indigenous and non-Indigenous Australians: a matched cohort study. The Lancet 367:1842.

von Katterfeld B 2010. Personal communication.

WADLS 2010. Data Linkage Branch – Project Applications 1995 – 2009. Project list. Perth. Viewed 29 March 2010.

Williamson A, Banks E, Redman S, Craig J, Cass A, Fernando D et al. 2010. The Study of Environment on Aboriginal Resilience and Child Health (SEARCH): study protocol. BMC Public Health 10:287.

Xu F, Sullivan EA, Madden RC, Black D & Pulver LR 2012. Improvement of maternal Aboriginality in NSW births data. BMC Medical Research Methodology 12.

You J, Condon JR, Zhao Y & Guthridge S 2009. Incidence and survival after acute myocardial infarction in Indigenous and non-Indigenous people in the Northern Territory, 1992-2004. Medical Journal of Australia 190:298-302.

Young MJ 2001. Assessing the quality of identification of Aboriginal and Torres Strait Islander people in Western Australian hospital data, 2000. Perth: Health Information Centre, Health Department of Western Australia.

Zubrick SR, Lawrence DM, Silburn SR, Blair E, Milroy H, Wilkes E et al. <Telethon Institute for Child Health Research> 2004. The Western Australian Aboriginal Child Health Survey: The Health of Aboriginal Children and Young People. Perth. Viewed October 2010, <http://www.ichr.uwa.edu.au/waachs/publications/volume_one>.

Zubrick SR, Mitrou F, Lawrence D & Silburn SR 2011. Maternal death and the onward psychosocial circumstances of Australian Aboriginal children and young people. Psychological Medicine 41(9):1971-80.

Zubrick SR, Silburn SR, De Maio JA, Shepherd C, Griffin JA, Dalby RB et al. <Telethon Institute for Child Health Research> 2006. The Western Australian Aboriginal Child Health Survey: Improving the educational experiences of Aboriginal children and young people. Perth. Viewed 26 October 2011 2010, <http://www.ichr.uwa.edu.au/waachs/publications/volume_three>.