APPENDIX A

AUSTRALIAN JURISDICTIONAL DIFFERENCES: A QUALITATIVE DESCRIPTION

INTRODUCTION

In Australia, health-care is delivered via a complex interplay of three systems. These three systems comprise the national provision of primary specialist and pharmaceutical services, a loosely federated secondary and public hospital system, along with profit-based private insurance/private hospitals. Given these arrangements it is reasonable to expect that culture, clinical behaviours and history of practice contribute to variations in utilisation at the end-of-life. Additionally, Australia comprises six states and two territories. The various operational dynamics in each State and Territory may have led to different utilisation patterns in respect to end-of-life care.

There is a dearth of published or grey literature examining jurisdictional differences in the delivery of health care in Australia. Consequently, this candidate undertook a small qualitative study to better understand jurisdictional nuances in the delivery of end-of-life care.

METHOD

Research Question: What particular characteristics do you believe defines your jurisdiction in relation to the provision of public health care?

Jurisdictions: The four jurisdictions (Queensland, New South Wales, South Australia and Western Australia) in which the quantitative research for this thesis was to be undertaken.

Participants: Five[1] key experts from the four participating jurisdictions. Each expert was/had been a senior policy maker from the respective jurisdiction with extensive policy experience concurrent with the thesis period.

Process: Qualitative interviews were held with each participant. In addition to the research question stated above, participants were requested to describe:

- the financial incentives that may affect utilisation patterns (particularly as it might relate to end-of-life care),
- their palliative care systems,
- how their organisation/governance systems may contribute to system outcomes,
- how the centralisation versus devolved administration might contribute to hospital utilisation.

Analysis: The analysis involved a descriptive synthesis, describing the dynamics apparent in each jurisdiction.

[1] **Professor Michael Reid**; Ex Director-General Health, New South Wales & Ex Director-General Health Queensland; **Mr Jim Birch;** Ex XEO South Australian Health; **Professor Jane Hall**, Professor of Health Economics, Centre for Health Economic Research and Evaluation, University of Technology Sydney; **Associate Professor Keith Evans;** Ex Director, Drug and Alcohol, Queensland Health & Ex Director, Primary Care, South Australian Health; **Mr Neale Fong**, Ex Director General WA Department of Health.

FINDINGS

The findings are arranged in terms of jurisdiction: Queensland, New South Wales, South Australia and Western Australia.

Queensland

Queensland Health was administered by Labor Governments for the majority of the period 1990-2010. These administrations were largely interventionist in nature with 'command and control' from the centre being the prominent paradigm. Whilst this period was associated with rhetorical commitment to out-of-hospital and primary focused care, the majority of funding was directed towards rebuilding infrastructure and an emphasis on technology led innovations.

During this period of time, Queensland Health suffered significant reputational harm related to the registration of overseas qualified medical practitioners. This emerged when questions were raised about the medical practices of the Director of Surgery, Bundaberg Base Hospital. Consequently Queensland Health has exercised substantial caution in the credentialing and employment of overseas trained doctors and has adopted a 'defence' culture regarding system/hospital clinical failure. This may contribute to conservative "quick to admit" behaviours.

Queensland Health was restructured towards the end of 2005. Part of the restructure was a reduction in the number of Health Districts from 38 to 20. In September 2008 the number of Health Districts was further reduced to 15. It is only since 2005/06 that the Queensland government increased the autonomy of the Health Districts supporting the move to a service funder and provider model.

Following a review of Queensland Health systems, an Independent Health Quality and Complaints Commission was established in July 2006. The Health Quality and Complaints Commission also has a role in the development and implementation of quality, safety and clinical practice standards throughout Queensland's public and private services and monitor best practice clinical governance and patient safety.

New South Wales (NSW)

During the research period of this study, the economic context in NSW was not particularly constrained, although the NSW government was endeavouring to address economic pressures in the health sector. The Government funded specific programs to address inequity in access to health-care. In particularly these programs were targeted at Indigenous people and also communities with socio-economic disadvantage.

The prevailing funding policy context was the implementation of a “Population Based Resource Allocation” model that effectively allocates to the various health regional authorities a funding quantum based on the population serviced, although it was adjusted for some specified known factors (such as age, socio economic status and Indigenous status). Activity based measures, such as casemix, were only referred to for benchmarking purposes and not used to determine funding acquittal parameters.

The ‘Garling Report’ featured during the period which examined a range of significant public quality problems in the Campbelltown Hospital and found not only avoidable harm but a culture of secrecy, and lack of transparency in management practice that allowed poor practice to prevail [65]. Again, these are not the subject of this research. However, possibly as a consequence of the Garling report recommendations, the Department of Health maintained a centrally controlled reign over the regional health services leading to a possible orientation towards risk mitigation in terms of quality issues rather than utilisation, or budget concerns.

South Australia

South Australia is centrally administered system. Regional health networks were formed quite late in the research period. While in the first formation of the area health services, the boards were given a reasonably high level of authority; the underlying system remained quite centrally controlled. Subsequently, the government progressively took back direct control of the system and boards were eliminated. In the later part of the research period, governing councils were reinstated although with advisory mandate only.

Importantly, South Australia implemented “casemix” funding in the early part of the research period (approximately 1993), the objective of which was to effectively set a determined price for each hospital service event (weighted for complexity). This system of accountability may affect utilisation patterns as it ties hospital funding to activity thereby rewarding additional activity. One might assume it would drive technical efficiency at the expense of allocative efficiency as it applies to hospital demand.

The South Australian economy was under financial pressure throughout the period and remains so and this may be expected to have a dampening effect on service availability. A relatively low investment in hospital demand management strategies has been apparent with the exception of a large scale coordinated care program in the period 1996 to 2000. This program aimed to reduce the hospitalisation of patients with chronic illnesses though was not able to demonstrate significant effect and, for the purposes of this research, is unlikely to show a material impact on end-of-life utilisation [57-64].

Western Australia

Western Australia's health system remained relatively stable with a centralised administration. The formation of Area Health Services was relatively recent and, even when implemented the governing councils of the regions were advisory rather than fully accountable statutory bodies. The medical peak body remains relatively influential and confers significant weight to the interest of the medical professions. The financing system in WA has been block funded to hospitals and admission activity was subject to relatively little constraint. The WA government throughout the period benefited from the mining resources "boom" and was, as such, relatively wealthy. This may have had the effect of alleviating financial pressure compared to other states. WA also embarked on a well-resourced community palliative care system with a non-government organisation (Silver Chain), which provides a single metropolitan wide non-admitted palliative care service. Uptake of the community palliative care services by palliative patients and palliative care physicians/GPs effects a significant percentage of predictable deaths [13, 55]. Patients who use the Silver Chain palliative care service typically use 1.2 admissions in the last 80 days of life and this may cause some observable utilisation patterns that distinguish it from other states. Overall, Western Australia's palliative care activity is three times the nation average (0.3 admissions per 1000 population vs national mean of 0.1 admissions) [56], most of which were absent from the in-patient data in this research.

Conclusion

The interviews with key experts indicated that there were nuances in healthcare delivery in different Australian jurisdictions. These differences are the likely outcomes of political, economic and social variances between Australian states. These differences were considered when interpreting the findings from the thesis study that addressed jurisdictional differences in end-of-life care.

APPENDIX B

EOL –VARIABLES X JURISDICTION

1.1 MORTALITY DATA

1.1.1 Cause of Death

The following cause of death information was collected for each jurisdiction. Coded cause of death was not available for 2010 decedents or for SA decedents.

	Format	NSW	WA	SA	QLD
2010	Free Text	I. Direct Cause of Death (COD1-COD5) I. Antecedent causes of death II. Contributing Factors	I. Direct Cause of Death (CODText) I. Antecedent causes of death II. Contributing Factors	codeath1-codeath7 Duration 1-7	I. Direct Cause of Death (D_CAUSE) I. Antecedent causes of death II. Contributing Factors
2005	ICD10	Icd10dth “cause of death” icd10r1..r20 “contributing causes”, sorted in ICD10 order (max 13 causes) Icd10dth is duplicated in the contributing causes of death in 99% of cases.	Primary COD only (ICD10) “COD_Code”	Not supplied	d_ICD_COD “cause of death” d_ICD_COD1- COD13 ICD10 underlying causes of death, sorted in ICD10 order d_ICD_COD is duplicated in the underlying causes of death in 49% of cases
	Free Text	I. Direct Cause of Death (COD1-COD5) I. Antecedent causes of death	N/A	codeath1-codeath7 Duration 1-7	Direct Cause of Death (D_CAUSE) I. Antecedent causes of death

	Format	NSW	WA	SA	QLD
		II. Contributing Factors -			II. Contributing Factors
2000	ICD10	-	Primary COD only (ICD10) "COD_Code"	-	-
1995	ICD9	-	Primary COD only (ICD9) "COD_Code"	-	-

1.1.2 Geography

A number of geographical variables were not available for the 2010 NSW deaths data, as these would have been sourced from the ABS coded deaths had it been available:

- Statistical Local Area *
- Statistical Division *
- Current Health Area of Residence *
- State of usual residence *
- State *

Local Health District and State of Residence were instead derived from the morbidity data (for those decedents with hospital admissions) – based upon the most frequent value.

Geography data was not made available by the Qld data custodians.

	Format	NSW	WA	SA	QLD
2010		State of Residence (derived – 93.2%) Local Health Districts (derived 93.2%) Area Health Service (derived	Area Health Service Health District Metro/ Country (derived)	Postcode SLA SEIFA (IRSD PCod06)	None

Format	NSW	WA	SA	QLD
	93.2%)	SEIFA (various)		
2005	State of Residence Statistical Division Statistical Local Area Local Health Districts (2011) Area Health Service	Area Health Service Health District Metro/ Country (derived) SEIFA (various)	Postcode SLA SEIFA (IRSD PCod06)	None
2000		Area Health Service Health District Metro/ Country (derived) SEIFA (various)		
1995		Area Health Service Health District Metro/ Country (derived) SEIFA (various)	-	-

1.1.3 Demographics

Sex was not available for the 2010 NSW deaths data, as this would have been sourced from the ABS coded deaths had it been available. Sex derived from the morbidity data was used instead (for those decedents with hospital admissions) – based upon the most frequent value.

Date of death was not available for SA or Qld decedents.

	NSW	WA	SA	QLD
2010	Sex (derived – 93.2%) Age Group (any) Age at Death Date of Death ddmmyyyy	Sex Age Group (any) Age at Death Date of Death ddmmyyyy Aboriginality	Sex Age Group (any) Age at Death Date of Death yyyymm Aboriginality	Sex 5 yrs age groups to 0-4…109 yrs N/A Date of Death yyyymm
2005	Sex Age Group (any) Age at Death Date of Death ddmmyyyy	Sex Age Group (any) Age at Death Date of Death ddmmyyyy Aboriginality	Sex Age Group (any) Age at Death Date of Death yyyymm Aboriginality	Sex 5 yrs age groups to 0-4…109 yrs N/A Date of Death yyyymm
2000		Sex		

NSW	WA	SA	QLD
	Age Group (any) Age at Death Date of Death ddmmyyyy Aboriginality		
1995	Sex Age Group (any) Age at Death Date of Death ddmmyyyy Aboriginality	-	-

1.2 Morbidity Data

1.2.1 Demographics

The Qld data custodians did not permit the release of sex information.

WA 1995-2010	Qld 2005 & 2010	SA 2005 & 2010	NSW 2005 & 2010
-	-	Age at separation	Age at separation
Gender	(use sex from deaths data)	Sex	Sex
DOB (yyyymm)	-	-	Country of Birth
IndigStat	-	Indigenous_Status	-

1.2.2 Geography

The Qld data custodians did not permit the release of geographical information.

WA 1995-2010	Qld 2005 & 2010	SA 2005 & 2010	NSW 2005 & 2010
Postcode		-	State of Residence
		Usual Residence (Country Health SA, Central Adelaide, Northern Adelaide, Southern Adelaide)	Local Health Districts (LHD11)
SEIFA (various)			Area Health Service of Residence
-		-	

1.2.3 Episode Related

Day of separation was not made available by the Qld data custodians.

WA 1995-2010	Qld 2005 & 2010	SA 2005 & 2010	NSW 2005 & 2010
Admission Date (yyyymm)	Admission Date (yyyymm)	Admission Date (yyyymmdd)	Admission Date (yyyymm)
-	Admission Time (hh:mm)	-	-
SepDateD (yyyymmdd)	Separation Date (yyyymm)	Separation Date (yyyymmdd)	Separation Date (yyyymmdd)
-	Separation time (hh:mm)	-	-
Admission Type/ Separation Type	-	Admission Type e.g. Ordinary, Long Stay Acute, Long Stay NHT	
SourceReferral	q_refsource	Source of Referral e.g. ED, self, Nursing Home,	Source of Referral e.g. ED, self, Nursing Home, Community

WA 1995-2010	Qld 2005 & 2010	SA 2005 & 2010	NSW 2005 & 2010
		Community Health	Health
SourceReferralLocation	-	-	-
-	-	Same Day Flag	Stay Flag e.g. Inpatient, SameDay
-	q_stnd_unit_start/ end	-	Unit/ Bed type e.g. rehab, palliative, coronary care, HDU, ICU etc
SeparationMode	q_sepn_mode	Separation Mode e.g. discharged to home, discharged to nursing home, died, transferred to psychiatric hospital	Separation Mode e.g. discharged to home, discharged to nursing home, died, xferred to psychiatric hospital
EpisodeCare	q_epis_type	Episode of Care e.g. Acute, Rehab, Newborn, Palliative, Maintenance	Episode of Care e.g. Acute, Rehab, Newborn, Palliative, Maintenance
-	-	-	Acute Flag - whether service at an acute facility
HospitalCategory e.g. Tertiary	q_hosptype (100% = public)	-	Hospital Role e.g. Major Metro, District, Community, Psychiatric
DRG, DRG version	ARDRG 41/42 (2005) 50/51 (2010)	ARDRG	ARDRG 4.1/5.0
-	MDC 41/42 (2005) 50/51 (2010)	-	-

WA 1995-2010	Qld 2005 & 2010	SA 2005 & 2010	NSW 2005 & 2010
-	-	-	Service Related Group V3
PrinDiag, CoDiagnosis, AddDiag01-AddDiag20	qPD1, qOD1..qOD30	Principal Diagnosis, additional diagnoses 1-29	Diagnoses ICD10d1..55
-	-	External Cause (1 only)	External Causes ICD10ex1..8
PrinProc (primary procedure)	qPR1..qPR10	Procedure 1-Procedure 25	MBS Item Codes 1..50
-	-	Separation Election (hospital/ private)	-
PatientFundingSource	-	-	Payment Status e.g. public, private, compensable, DVA, etc.
-	-		Insurance Status e.g. Basic, Ancillary, Comprehensive
LOS (days)	q_stay (capped at 60 days)	Length of Stay (days)	Length of Stay (days)
-	-	-	Day Only length of stay (hours)
-	-	-	Leave Days
ICUDays	-	ICU hours	-
	q_icu flag (0/1)		
PsychDays			

APPENDIX C

DEATH CERTIFICATE CODING

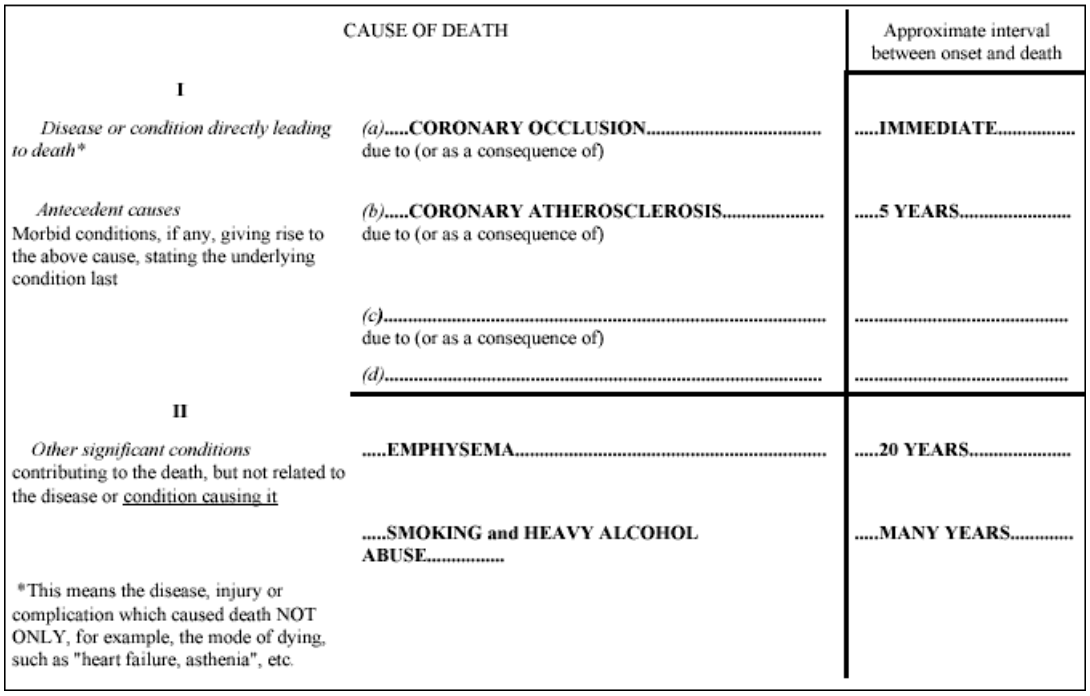
2.1 Death Certificates

Death Certificates in Australia comprise two separate forms; a medical certificate which indicates the cause(s) of death, and a questionnaire providing personal information about the deceased. The medical certificate is completed either by a doctor who was in attendance at the time of death or who can certify as to the cause of death, or by the coroner when the death was unexpected or unexplained. The personal information questionnaire is completed by the next of kin. This normally takes place at the funeral parlour with the help of the funeral director. Both forms are collected by the Registrar of Births and Deaths in each state or territory.

There are two parts to the standard medical certificate of cause of death.

Part I is the area "above the line". This is where the disease or condition directly leading to death is stated followed by any conditions which have given rise to this disease or condition (these are called antecedent causes). Any conditions listed above the line should form what is termed the "morbid train of events" that have led to death. That is, they form a sequence starting at the disease or condition which directly led to death. This condition may then have been "due to (or as a consequence of)" an antecedent cause which was in turn, "due to (or as a consequence of)" another antecedent cause et cetera. The Underlying Cause is the cause which is listed last. That is, it is the cause that is deemed to have started the morbid train of events.

Part II is the area "below the line". This area is to be used to list other significant conditions which have contributed to the death but which are not deemed to be part of the morbid train of events leading to the death.



CAUSE OF DEATH		Approximate interval between onset and death
I		
*Disease or condition directly leading to death**	*(a)*.....**CORONARY OCCLUSION**..... due to (or as a consequence of)	.....**IMMEDIATE**.....
Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last	*(b)*.....**CORONARY ATHEROSCLEROSIS**..... due to (or as a consequence of)	.....**5 YEARS**.....
	(c)..... due to (or as a consequence of)	.....
	(d).....	.....
II		
Other significant conditions contributing to the death, but not related to the disease or condition causing it	.....**EMPHYSEMA**.....	.....**20 YEARS**.....
	.....**SMOKING and HEAVY ALCOHOL ABUSE**.....	.....**MANY YEARS**.....
*This means the disease, injury or complication which caused death NOT ONLY, for example, the mode of dying, such as "heart failure, asthenia", etc.		

Figure1 - Example of a completed medical certificate of cause of death

Source:

http://www.abs.gov.au/AUSSTATS/abs@.nsf/956c382b0b05ba7d4a2568010004e173/875c0bee2d2512ebca256d6b0004830f!OpenDocument

2.2 ABS Coding of Causes of Death

The ABS coding process classifies the text describing causes of death into ICD codes for statistical purposes. The data are processed using the *Mortality Medical Data System* (MMDS). The information is then made available in two forms - **entity-axis-data** and **record-axis-data**.

- **Entity-axis-data** contain codes for all causes of death in the order and position that they appeared on the death certificate.
- **Record-axis-data** contain codes for all causes of death, in alphanumeric order, following processing in accordance with ICD coding rules and rules associated with the automated processing system.

Record-axis-data have multiple and underlying causes assigned and have had duplicate and superfluous codes removed, and are thus generally more appropriate for output purposes. In addition, some specific combinations of causes are 'linked', that is they are replaced by a single code. For example, if 'pulmonary oedema' (J81) appears on the death certificate along with 'heart failure unspecified' (I50.9), these two causes are replaced by 'left ventricular failure' (I50.1). Record-axis-data are generally used for ABS publications and have been used for all analyses in this paper.

The rules governing the selection of an underlying cause of death under ICD-10 are highly prescriptive. A couple of the scenarios illustrate this:

In the first scenario, the deceased fell and fractured a femur. After surgery to rectify the fracture the person had a myocardial infarction. Myocardial ischaemia due to atherosclerotic heart disease were specified on the first two lines of Part 1 of the certificate.
Surgical repair of the fracture appeared on Part 2. The sequence myocardial ischaemia due to atherosclerotic heart disease automatically dictated that the underlying cause of death would be coded as I21.9 Acute myocardial infarction, unspecified.

A second scenario provides a similar example. The person lost balance, fell, and sustained a fractured femur. The person was admitted to hospital, and later died there of a heart attack. The first two lines of the death certificate specified that the deceased had died of a myocardial infarction due to ischaemic heart disease. This sequence dictated an underlying cause of I21.9 Acute myocardial infarction, unspecified. The fact of the fall did not appear on the death certificate.

A third example provides an example of the importance placed by coders on the wording of the certificate. The deceased had fallen and fractured a femur. The person was admitted to hospital and subsequently suffered a cardiac arrest. Ischaemic heart disease (IHD) appeared on the first line of Part 1 of the certificate. IHD is considered to be a long-term condition which, under the ICD-10 coding rules, automatically lead to the assignment of I25.9 Chronic ischaemic heart disease, unspecified as the underlying cause of death. Had the first line of the certificate instead stated acute ischaemia, coders would have looked at information contained on Part 2 of the certificate, which mentioned the fracture.

The rules governing certification of deaths of people aged 75 years and over who fractured their femur are illustrated by Scenario 11. In this case, a fall that resulted in a fractured neck of femur lead to admission to hospital. During surgery to repair the fracture, the patient suffered suspected heart failure. I50.0 Congestive heart failure was coded as the Underlying cause of death. The coder has commented that she had some suspicions about the accuracy of this UCoD but, because the person was aged over 74, the case would not have been queried.

Source: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442458803

APPENDIX D

JURISDICTIONAL DATA LINKAGE OVERVIEWS

3.1 WA Data Linkage Overview

Source: http://www.datalinkage-wa.org/data-linkage/linkage-process

Where truly unique personal identification numbers are not available across all information sources, probabilistic linkage allows connections (or linkages) to be created by comparing the personal information available and calculating the likelihood that records belong to the same person, place or event.

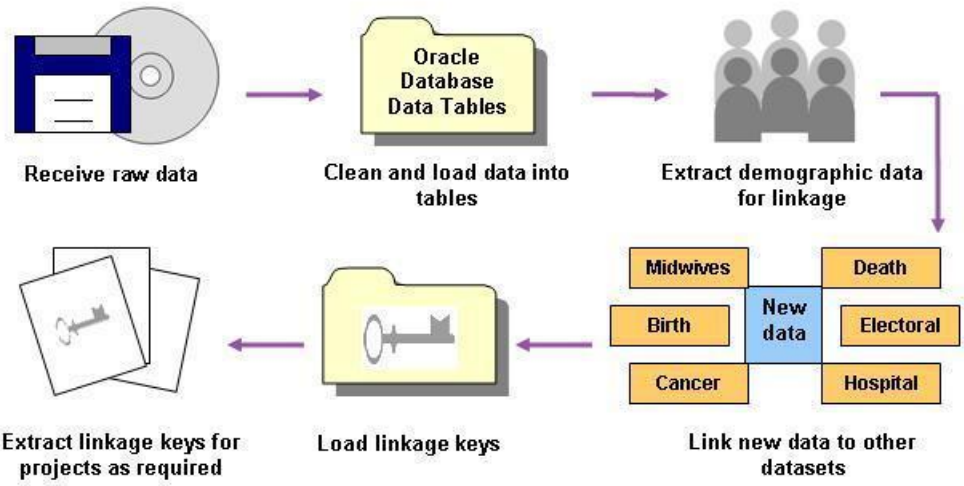


Figure 2 - WA Data Linkage Overview

Linkage is a complex process that uses many 'passes' through datasets using different arrangements of the data items at each pass. Weights are assigned based on the likelihood of a "true match" and thresholds are set to separate "probable" and "improbable links". Linkage strategies are designed so that these thresholds are as close together as possible to minimise the number of matches that need manual review and a decision by a Linkage Officer.

3.1.1 Linkage Process

The linkage process can be split into the following steps:

3.1.1.1 Obtain demographic data

Raw data is provided for linkage. All or some of the following demographic fields are included:

- Name (first name, second name, family name, aliases)
- Date of Birth
- Address (house number, street name, suburb, postcode)
- Sex
- Record date
- Other unique identifiers (e.g. Hospital Unique Medical Record Number)

3.1.1.2 Clean and standardise data

The data fields are cleaned and put into a standard format that can be used for linkage. Customised identifiers are assigned. For example:

MC DONALD > MCDONALD

O'CONNOR > OCONNOR

12th August 1982 > 19820812

3.1.1.3 Load demographic tables

The demographic details are loaded into tables in an Oracle database. There are different tables for different datasets since not all datasets have the same variables.

3.1.1.4 Extract linkage variables

Customised scripts are used to extract only those records and fields required for a given linkage into "flat data files".

3.1.1.5 Run linkage engine

The linkage program is used to run comparisons between two flat data files. Linkage officers can customise their linkage strategies according to the individual characteristics of each dataset. Some links pass as automatic matches, some are automatic rejections, and some fall into a "grey area" in between where links are manually checked for validity.

3.1.1.6 Load links

The IDs of linked records are assigned an identical "chain number", which is stored in a separate database.

3.1.1.7 Update links as required

Linkages are regularly revisited to ensure that the system of links is continually refined and improved.

3.1.1.8 Extract linkage keys

Customised project specific linkage keys are extracted by encrypting the "master ID" for each chain of records. These are the keys that have service data attached by the various data collections.

3.1.1.9 Extraction process

The following diagram shows how linked data is extracted:

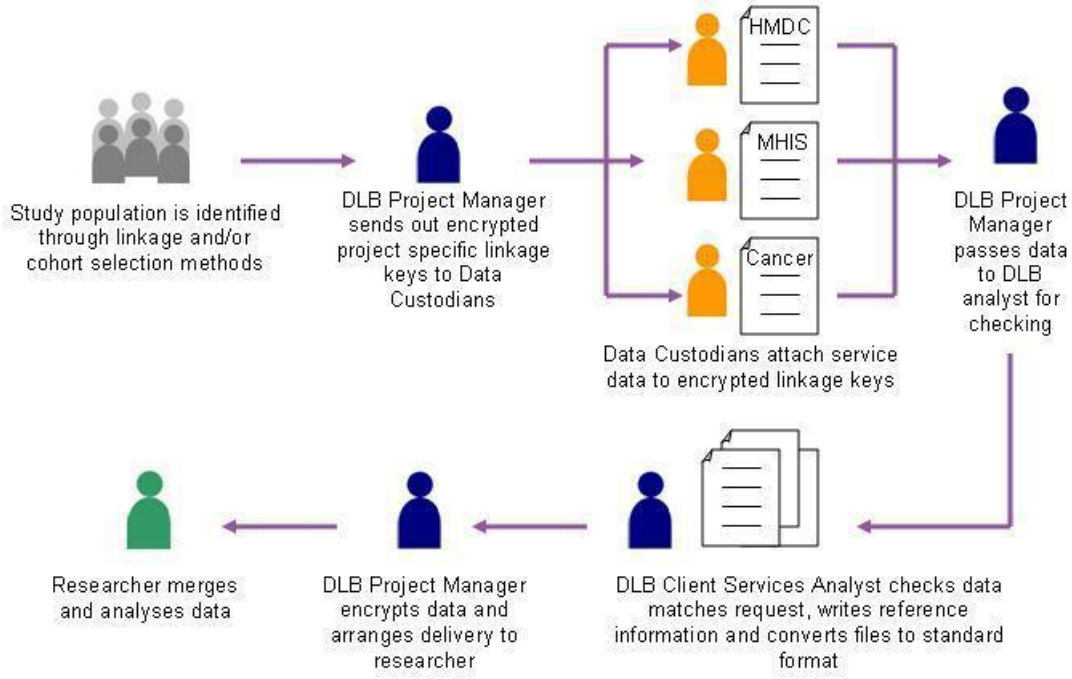


Figure 3 - WA Data Extraction Process

3.1.2 Extraction Process

3.1.2.1 Identify study population

The study population first needs to be selected. This can be done via linkage, where the researcher already has the study population chosen, or via selection from one or more of the health data collections. For example:

- All people who went to hospital for a colonoscopy (from HMDC)
- All people with colorectal cancer (from WA Cancer Registry)
- People in both these groups (from both HMDC and WA Cancer Registry).

Control populations are also identified (e.g. a random sample of people from the electoral roll who are the same age and gender as the cases).

3.1.2.2 Extract linkage keys

Once the study population is defined, the linkage team extracts the linkage keys for each requested dataset. The project manager then distributes these lists of keys to the relevant data collections for the service data to be attached.

3.1.2.3 Attach service data

The data custodians arrange for the requested service data from their collection to be attached to the linkage keys. For core data collections, the files are sent back to the DLB project manager. For some external datasets, the service data is released directly to the researcher.

3.1.2.4 Checking

The service data files come back in various formats. The DLB project manager arranges for a DLB analyst to check that the data matches the request and convert all the data to fixed width text files. Supporting documentation is also written to describe the data requested.

3.1.2.5 Data release

The DLB project manager prepares the data for release by encrypting it and burning it to a disc, then arranges for secure delivery to the researcher.

3.2 SA Data Linkage Overview

A data linkage was conducted whereby 24,872 Births, Deaths and Marriages (BDM) death records for 2005 and 2010 year of death were linked to 3,704,846 ISAAC public hospital inpatient records for the separation years 2009 to 2011. A total of 20,024 death records were found to link to the hospital inpatients file giving a linking success rate of 80.5%.

Automatch Probabilistic Record Linkage Version 4.3 computer software (Matchware Technologies Inc, Kennebunk, ME) was used for the data linkage.

Prior to data linking, a unique study identifier was assigned to the deaths file. Both files were also cleaned to increase the probability of linking records from both files whereby given names were separated and placed into separate fields, and any erroneous characters, hyphens, comments, or aliases in name fields, were removed prior to linking. NYSIIS and soundex phonetic codes, HASAC identifier (i.e. first four characters of surname, first given and middle initials) altered given name and altered surnames were computed using algorithms written in Microsoft Visual Foxpro 7.0.

For the Automatch linkage, a total of eight linkage passes were conducted with a manual clerical review of the linked record pairs (and any associated linked duplicate records) performed after each completed linkage pass. Cutoff weights for all passes were initially set to a weight of 1.0. However, these were adjusted upwards where required in accordance with the quality of the linked pairs of records for each pass. Blocking variables for the different passes included the use of surname, first given name, sex, exact date of birth, hospital, medical record number, NYSIIS and Soundex of surname, modified surname where vowels were removed, modified year of birth whereby the last digit was excluded, short given name and first initial, and combinations of date of birth components. For pass 1, blocking variables were only used whilst for the remaining passes, both blocking and matching variables were used.

A copy of the Automatch program used for this linkage is included as follows:

```
;Automatch.mat
;2005 & 2010 BDM deaths vs Isaac Public 2009-2011
PROGRAM MATCH
DICTA bdmdths
DICTB isacpub
;
;Pass: 1
; *** Blocking variables
BLOCK1 CHAR clsname clsname
BLOCK1 CHAR given1  given1
BLOCK1 CHAR nsex    nsex
BLOCK1 CHAR dobyyyy dobyyyy
BLOCK1 CHAR dobmm   dobmm
BLOCK1 CHAR dobdd   dobdd
;
;Pass: 2
BLOCK2 CHAR hospital hospital
BLOCK2 CHAR urno urno
;
MATCH2 UNCERT clsname clsname 0.70 0.00001 700.0
MATCH2 ARRAY UNCERT name name 0.70 0.00001 700.0
MATCH2 UNCERT nysiis nysiis 0.90 0.0001 700.0
MATCH2 UNCERT hasac hasac 0.90 0.00001 700.0
MATCH2 UNCERT given1 given1 0.90 0.0001 700.0
;MATCH2 UNCERT given2 given2 0.90 0.0001 700.0
MATCH2 ARRAY UNCERT given given 0.90 0.02 900.0
MATCH2 ARRAY UNCERT initials initials 0.70 0.02 700.0
MATCH2 UNCERT sh_g1 sh_g1 0.90 0.02 700.0
MATCH2 UNCERT sh_gn1 sh_gn1 0.90 0.02 700.0
```

```
MATCH2 UNCERT sh_g2 sh_g2 0.90 0.02 700.0
MATCH2 ARRAY UNCERT sh_given sh_given 0.70 0.02 700.0
MATCH2 PRORATED dobyyy dobyyy 0.90 0.02 0
MATCH2 PRORATED dobyyyy dobyyyy 0.90 0.02 0
MATCH2 PRORATED dobmm dobmm 0.90 0.001 0
MATCH2 PRORATED dobdd dobdd 0.90 0.03 5
MATCH2 ARRAY UNCERT dobc8 dobc8 0.70 0.0002 700.0
;
;Pass: 3
BLOCK3 CHAR nysiis  nysiis
BLOCK3 CHAR given1  given1
BLOCK3 CHAR nsex    nsex
;BLOCK3 CHAR dobyyyy dobyyyy
;BLOCK3 CHAR dobmm   dobmm
;BLOCK3 CHAR dobdd   dobdd
;
MATCH3 UNCERT clsname clsname 0.70 0.00001 700.0
MATCH3 ARRAY UNCERT name name 0.70 0.00001 700.0
;MATCH3 UNCERT nysiis nysiis 0.90 0.0001 700.0
MATCH3 UNCERT hasac hasac 0.90 0.00001 700.0
;MATCH3 UNCERT given1 given1 0.90 0.0001 700.0
MATCH3 UNCERT given2 given2 0.90 0.0001 700.0
MATCH3 ARRAY UNCERT given given 0.90 0.02 900.0
MATCH3 ARRAY UNCERT initials initials 0.70 0.02 700.0
MATCH3 UNCERT sh_g1 sh_g1 0.90 0.02 700.0
MATCH3 UNCERT sh_gn1 sh_gn1 0.90 0.02 700.0
MATCH3 UNCERT sh_g2 sh_g2 0.90 0.02 700.0
MATCH3 ARRAY UNCERT sh_given sh_given 0.70 0.02 700.0
MATCH3 PRORATED dobyyyy dobyyyy 0.90 0.02 0
MATCH3 PRORATED dobyyy dobyyy 0.99 0.001 0
```

```
MATCH3 PRORATED dobmm dobmm 0.90 0.001 0
MATCH3 PRORATED dobdd dobdd 0.90 0.03 5
MATCH3 ARRAY UNCERT dobc8 dobc8 0.70 0.0002 700.0
;
;Pass: 4
BLOCK4 CHAR altsname altsname
BLOCK4 CHAR given1   given1
;BLOCK4 CHAR nsex     nsex
BLOCK4 CHAR dobyyy   dobyyy
BLOCK4 CHAR dobmm    dobmm
BLOCK4 CHAR dobdd    dobdd
;
; *** Matching variables
MATCH4 UNCERT clsname clsname 0.70 0.00001 700.0
MATCH4 ARRAY UNCERT name name 0.70 0.00002 700.0
MATCH4 UNCERT nysiis nysiis 0.90 0.0002 700.0
MATCH4 UNCERT hasac hasac 0.70 0.0002 700.0
MATCH4 UNCERT given2 given2 0.90 0.0001 700.0
MATCH4 ARRAY UNCERT given given 0.90 0.02 900.0
MATCH4 ARRAY UNCERT initials initials 0.70 0.02 700.0
MATCH4 UNCERT sh_gn1 sh_gn1 0.90 0.0002 700.0
MATCH4 UNCERT sh_g1  sh_g1  0.90 0.0001 700.0
MATCH4 UNCERT sh_g2 sh_g2 0.90 0.02 700.0
MATCH4 PRORATED dobyyyy dobyyyy 0.90 0.02 0
;MATCH4 PRORATED dobyyy dobyyy 0.99 0.001 0
;MATCH4 PRORATED dobmm dobmm 0.90 0.001 0
;MATCH4 PRORATED dobdd dobdd 0.90 0.03 5
MATCH4 ARRAY UNCERT dobc8 dobc8 0.70 0.0002 700.0
;
;Pass: 5
```

```
BLOCK5 CHAR altsname altsname
BLOCK5 CHAR sh_gn1   sh_gn1
BLOCK5 CHAR nsex     nsex
BLOCK5 CHAR dobyyyy  dobyyyy
BLOCK5 CHAR dobdd    dobdd
;
; *** Matching variables
MATCH5 UNCERT clsname clsname 0.90 0.00002 700.0
MATCH5 ARRAY UNCERT name name 0.70 0.00002 700.0
MATCH5 UNCERT hasac hasac 0.70 0.02 700.0
MATCH5 ARRAY UNCERT initials initials 0.70 0.03 700.0
MATCH5 UNCERT given1 given1 0.90 0.00001 700.0
MATCH5 UNCERT given2 given2 0.99 0.001 900.0
MATCH5 UNCERT sh_g1 sh_g1 0.90 0.0001 700.0
MATCH5 UNCERT sh_g2 sh_g2  0.99 0.03 700.0
MATCH5 ARRAY UNCERT given given 0.90 0.02 700.0
MATCH5 ARRAY UNCERT sh_given sh_given 0.70 0.0001 700.0
MATCH5 PRORATED dobyyyy dobyyyy 0.90 0.02 0
;MATCH5 PRORATED dobmm dobmm 0.90 0.001 0
;MATCH5 PRORATED dobdd dobdd 0.90 0.03 5
MATCH5 ARRAY UNCERT dobc8 dobc8 0.70 0.0002 700.0
;
;Pass: 6
BLOCK6 CHAR altsname altsname
BLOCK6 CHAR init1 init1
BLOCK6 CHAR nsex nsex
BLOCK6 CHAR dobyyy dobyyy
BLOCK6 CHAR dobmm dobmm
BLOCK6 CHAR dobdd dobdd
;
```

```
; *** Matching variables
MATCH6 UNCERT clsname clsname 0.90 0.00002 700.0
MATCH6 ARRAY UNCERT name name 0.70 0.00002 700.0
MATCH6 UNCERT hasac hasac 0.70 0.02 700.0
MATCH6 ARRAY UNCERT initials initials 0.70 0.03 700.0
MATCH6 UNCERT given1 given1 0.90 0.00001 700.0
MATCH6 UNCERT given2 given2 0.99 0.001 900.0
MATCH6 UNCERT sh_g1 sh_g1 0.90 0.0001 700.0
MATCH6 UNCERT sh_g2 sh_g2  0.99 0.03 700.0
MATCH6 ARRAY UNCERT given given 0.90 0.02 700.0
MATCH6 ARRAY UNCERT sh_given sh_given 0.70 0.0001 700.0
MATCH6 PRORATED dobyyyy dobyyyy 0.90 0.02 0
MATCH6 PRORATED dobyyy dobyyy 0.90 0.02 0
;MATCH6 PRORATED dobmm dobmm 0.90 0.001 0
;MATCH6 PRORATED dobdd dobdd 0.90 0.03 5
MATCH6 ARRAY UNCERT dobc8 dobc8 0.70 0.0002 700.0
;
;Pass: 7
BLOCK7 CHAR soundex soundex
BLOCK7 CHAR sh_g1   sh_g1
BLOCK7 CHAR nsex    nsex
BLOCK7 CHAR dobyyy  dobyyy
BLOCK7 CHAR dobmm   dobmm
BLOCK7 CHAR dobdd   dobdd
;
; *** Matching variables
MATCH7 UNCERT clsname clsname 0.90 0.00002 700.0
MATCH7 ARRAY UNCERT name name 0.70 0.00002 700.0
;MATCH7 UNCERT nysiis nysiis 0.70 0.00002 700.0
MATCH7 UNCERT hasac hasac 0.70 0.02 700.0
```

```
MATCH7 ARRAY UNCERT initials initials 0.70 0.03 700.0
MATCH7 UNCERT given1 given1 0.90 0.00001 700.0
MATCH7 UNCERT given2 given2 0.99 0.001 900.0
MATCH7 UNCERT sh_gn1 sh_gn1 0.90 0.0002 700.0
;MATCH7 UNCERT sh_g1 sh_g1 0.90 0.0001 700.0
MATCH7 UNCERT sh_g2 sh_g2 0.99 0.03 900.0
MATCH7 ARRAY UNCERT given given 0.90 0.02 700.0
MATCH7 ARRAY UNCERT sh_given sh_given 0.70 0.0001 700.0
MATCH7 PRORATED dobyyyy dobyyyy 0.99 0.02 0
;MATCH7 PRORATED dobdd dobdd 0.90 0.03 5
MATCH7 ARRAY UNCERT dobc8 dobc8 0.70 0.0002 700.0
;
;Pass: 8
BLOCK8 CHAR altsname altsname
BLOCK8 CHAR init1 init1
;BLOCK8 CHAR nsex nsex
BLOCK8 CHAR dobmm dobmm
BLOCK8 CHAR dobyyy dobyyy
;
; *** Matching variables
MATCH8 UNCERT clsname clsname 0.90 0.00002 700.0
MATCH8 ARRAY UNCERT name name 0.70 0.00002 700.0
MATCH8 UNCERT hasac hasac 0.70 0.02 700.0
MATCH8 ARRAY UNCERT initials initials 0.70 0.03 700.0
MATCH8 UNCERT given1 given1 0.90 0.00001 700.0
MATCH8 UNCERT given2 given2 0.99 0.001 900.0
MATCH8 UNCERT sh_g1 sh_g1 0.90 0.0001 700.0
MATCH8 UNCERT sh_g2 sh_g2  0.99 0.03 700.0
MATCH8 ARRAY UNCERT given given 0.90 0.02 700.0
MATCH8 ARRAY UNCERT sh_given sh_given 0.70 0.0001 700.0
```

```
MATCH8 PRORATED dobyyyy dobyyyy 0.90 0.02 0
;MATCH8 PRORATED dobmm dobmm 0.90 0.001 0
MATCH8 PRORATED dobdd dobdd 0.90 0.03 5
MATCH8 ARRAY UNCERT dobc8 dobc8 0.70 0.0002 700.0
;
; Match, clerical & duplicate cutoff weights
CUTOFF1   1.0   1.0    1.0
CUTOFF2   1.0   1.0    1.0
CUTOFF3  65.0  65.0   65.0
CUTOFF4   1.0   1.0    1.0
CUTOFF5   1.0   1.0    1.0
CUTOFF6   1.0   1.0    1.0
CUTOFF7   1.0   1.0    1.0
CUTOFF8  61.0  61.0   61.0
```

3.3 Qld Data Linkage Overview

Reference: http://www.health.qld.gov.au/hsu/pdf/other/qlddatalinkframework.pdf

3.3.1 Datasets for linkage

The most common datasets used for linkage include:

- Queensland Hospital Admitted Patient Data Collection (QHAPDC)
- Queensland Perinatal Data Collection (QPDC)
- Queensland Cancer Registry (QCR)
- Registrar General (RG) deaths*
- Emergency Department Information System (EDIS)
- Community Integrated Mental Health Application (CIMHA)

*For data linkage purposes, Queensland Health acts as proxy data custodian under a Memorandum of Understanding (MOU) with the Queensland Registrar General.

There is also a Master Linkage File (MLF) containing permanently linked references to QHAPDC, QPDC, RG births and RG deaths. The use of this file saves a significant amount of RLG resources and results in a faster processing time, benefiting researchers. The intention is to increase the currency of the linked data and to develop methodology for near-real time linkage and to expand the MLF with data from further sources such as the Emergency Department Information System and Queensland Ambulance Service (QAS) to improve its coverage and therefore usefulness.

3.3.2 Current Master Linkage File coverage:

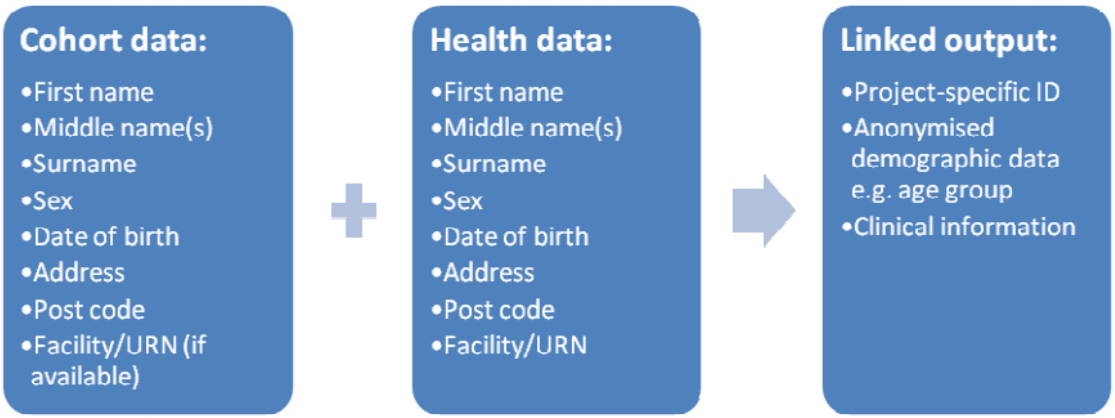
- QHAPDC Hospital Admissions* 01 July 2003 to 30 June 2012**
- Queensland Perinatal Data Collection* 01 July 2003 to 30 June 2012**
- Registrar General births 01 July 2009 to 30 June 2012
- Registrar General deaths 01 July 2003 to 30 June 2012**

*names and addresses for private facilities and in perinatal data were not supplied to the central data collection prior to 01 July 2007 so the quality of the linkage for these patients is questionable for the period from 01 July 2003 to 30 June 2007. This may result in bias in the analyses undertaken using data from this period. For example, readmission rates for private hospitals may appear lower than for public hospitals due to failure to link data for private patients.

**Data were linked in two stages using different tools. As a result, pre-linked data are available for 01 July 2003 to 30 June 2010 or 01 July 2009 to 30 June 2012, but the data from these two periods have not been internally linked. That is, there is no person ID that covers the entire period at this stage (updated 13/12/13).

3.3.3 Data Linkage Process Overview

The figure below is a simplified overview of the linkage process involving linking a client's cohort file to Queensland Health datasets such as QHAPDC, resulting in a file containing only anonymous summarised clinical information relevant to the research proposal:



1 Client identifies the population of interest. This could be a supplied cohort group from the researcher or Queensland Health clinical database, or a specific group from a health-related data collection(s) (e.g. admitted or emergency department heart failure patients transported by the Qld Ambulance Service).

2 Cohort file containing only identifiers for linkage plus a person and (if applicable) record ID sent securely to RLG. Linkage staff import, clean and standardise identifying demographic information for linkage. Issues such as missing data and duplicates are noted for reporting back to the client.

3 Linkage staff use most efficient/appropriate method of linking cohort to required dataset(s), for example if MLF can be used this greatly reduces clerical checking of uncertain 'grey area' links. Similarly, for linkages involving QHAPDC if compatible identifying reference numbers are available in the cohort data, a deterministic matching exercise can be done, with any remaining unmatched cohort linked probabilistically. A project-specific linkage key is created during this process.

4 The project-specific linkage key is attached to the relevant clinical service records and required variables are extracted. Where necessary, data items are manipulated to

meet agreed privacy protocols e.g. date of admission formatted as year/month, age transformed to 5-year bands.

5 The whole linkage process is quality-checked by a separate member of linkage staff. Quality assurance is discussed in further detail below.

6 Methodology and any data issues are summarised in a report, along with output variable descriptions and details of what is contained in output worksheet(s)/file(s). Where possible output is provided in Excel, but large files may be supplied in SAS or as text files.

7 Final output and report are sent to the client via a secure method of transmission. Normally WinZip is used to password protect the file(s) with 256 bit encryption. If the client does not have use of WinZip an alternative encryption method is agreed. Passwords are always given separately by telephone.

3.3.4 Quality Assurance (QA) of the linkage

A number of checks are carried out at each stage of the linkage process, from receipt of cohort data (if applicable) to ensuring the final output meets the client's requirements and complies with privacy protocols.

These checks are applied as an additional layer of checking following matching by linkage software and manual review of all grey area identified. These checks involve automated identification of any records or groups of records meeting the criteria and manual review to determine whether they should or should not be considered a match. The checks listed are used on the Master Linkage File and, where applicable, for linkage conducted for research requests. These checks are based on the list of quality assurance checks used by the Centre for Health Research Linkage in NSW (CHeReL), personal communication with other linkage units within Australia, and quality issues identified with Queensland data or with the linkage process and tools being used in Queensland.

Although quality checks are carried out, there are limitations to data linkage within Queensland data sets that may make linkage output appear to be incorrect. For example, when hospitalisations are linked to death there may be instances when a discharge type of 'death' is not linked to an RG death record. This could be due to the death being registered in a different state – for example, residents of northern NSW may be admitted to facilities in southern QLD and vice versa.

3.4 NSW Data Linkage Overview

3.4.1 Linkage process

Identifying information such as name, address, date of birth, date of death and gender obtained from RBDM deaths and APDC are included in the Master Linkage Key (MLK), which is being constructed by the Centre for Health Record Linkage (CHeReL). [1] No health data are used in this process.

RBDM Deaths and APDC records were linked using probabilistic record linkage methods and *ChoiceMaker* software.[2] *ChoiceMaker* uses 'blocking' and 'scoring' to identify definite and possible matches. During blocking, *ChoiceMaker* searches the target datasets for records which are possible matches to each other. There are two types of blocking. The exact blocking algorithm requires records to have the same set of valid fields and the same values for these fields. The automated blocking algorithm builds a set of conditions that are used to find as many as possible records that potentially match each other.

Scoring employs a combination of a probabilistic decision, which is computed using a machine learning technique, and absolute rules, which include upper and lower probability cut-offs, to determine whether each potential match denotes or possibly denotes the same person. Upper and lower probability cut-offs initially start at 0.75 and 0.25 for a linkage and are adjusted for each individual linkage to ensure false links are kept to a minimum.

At the completion of the process, each record in the MLK is assigned a record identification number and a MLK person ID to allow linked records for the same individual to be identified and extracted.

3.4.2 *Extraction of records from Master Linkage Key (MLK)*

All RBDM records for person who died in the years 2005 and 2010 were extracted from CHeReL Master Linkage Key (MLK). The APDC records which were linked to the persons in the RBDM records for the five years prior to death were also extracted.

3.4.3 Deterministic linkage of RBDM deaths with ABS mortality data

The RBDM death records in the MLK extract were deterministically linked to the ABS mortality records. This linkage was carried out using a five-step procedure:

First pass	year of registration, encrypted registration number and exact date of death.
Second pass	year of registration, encrypted registration number and either: 1 day difference in date of death or same year of death or date of birth.
Third pass	year of registration, date of death, sex, postcode and date of birth
Fourth pass	year of registration, date of death, sex, date of birth
Fifth pass	date of death, sex and date of birth.

Once the linkage was finalised, the CHeReL created a Project Person Number (PPN) for each Person ID and assigned the PPN to all records in the linked datasets. The PPN and the encrypted record number from each source database were returned to the respective data custodians. Each data custodian sent to the study investigators a dataset comprising the approved information from the source database plus the PPNs. The investigators can merge the datasets using the PPN.

3.4.4 Results

Table 1 shows the number and type of records from each data source.

Table 2 summarises the outcome of extracting the APDC records which matched the persons in the RBDM records from the MLK version [2013_10].

The MLK is regularly checked for false positive linkages.

False positive rate = 3/1,000 records (0.3%)

Table 1 - Data sources and record types

Data Source	Description	Number
NSW Register of Births, Deaths and Marriages (RBDM)	RBDM death registrations in NSW. Death date: Jan 2005 – Dec 2005 Jan 2010 – Dec 2010	 60,543 records 47,478 records
NSW Admitted Patient Data Collection (APDC)	All episodes of care selected for the following parameters: Separation date: Jul 2000 – Dec 2005 Jan 2006 – Dec 2010	 11,494,314 records 12,336,480 records
Australian Bureau of Statistics (ABS) mortality data	All ABS Mortality Data in NSW Death date: Jan 2005 – Dec 2005 Jan 2010 – Dec 2010	 44,548 records 0 records*

**ABS mortality data is only available up to Dec 2007.*

Table 2 - Summary of records returned to Study Investigators

Data Source	Description	Number
RBDM **(Cohort)**	RBDM death registrations in NSW. Death date: Jan 2005 – Dec 2005	60,543 records (44,629 persons)
	Death date: Jan 2010 – Dec 2010	47,478 records (47,282 persons)
	Total RBDM records	**108,021 records** **(91,911 persons)**
APDC	APDC records that linked to the above RBDM (2005 and 2010) cohorts Separation date: Jul 2000 – Dec 2005	401,556 records (41,395 persons) 92.75% of cohort
	Separation date: Jan 2006 – Dec 2010	514,230 records (44,274 persons) 93.25% of cohort
	Total linked APDC records	**915,786 records** **(85,669 persons)**
ABS Mortality	ABS death records that linked to the above RBDM 2005 cohort (ABS data not available for 2010 cohort) Death date: Jan 2005 – Dec 2005	44,522 records (44,509 persons) 99.73% of cohort
Total Records to be returned to Study Investigators: Total Project Person Numbers (PPN):		**1,068,329 records** **(91,911 persons)**

3.4.5 References

The Centre for Health Record Linkage at: www.cherel.org.au.

Borthwick, A, Buechi, M, and Goldberg, A. (2003). Key concepts of the ChoiceMaker 2 record Matching System. In: Proceedings of the KDD-2003 Workshop on Data Cleaning, Record Linkage, and Object Consolidation, 7-12. Washington DC

APPENDIX E

PATIENT COHORT DEFINITIONS

NEOPLASM COHORT

Definition

Decedents whose primary cause of death was a cancer-related condition.

ICD10 Code Range

C00-C17, C18-C21, C22, C23-C32, C33-C34, C35-C49, C50, C51-C97, D60-D63, D37-D48, B21

Detail:

Malignant neoplasms C00-C97

HIV disease resulting in malignant neoplasms B21

Neoplasms of uncertain or unknown behaviour D37-D48

Preventable Subsets[2]

Lung Cancer 0-74 years: C33-C34

Liver Cancer 0-74 years: C22

Colorectal cancer 0-74 years: C18–C21

Treatable Subsets

Breast cancer 0-74 years: C50

Exclusions

Benign neoplasms D10-D36

Chronic Myeloproliferative disease D47.1

[2] Preventable/ treatable deaths based upon ICD10 codes and age group were analysed. Reference*: 'Healthy Communities. Avoidable deaths and life expectancies in 2009–2011' www.nhpa.gov.au National Health Performance Authority.*

ICD9 Code Range

140-165, 170-174, 176, 179-208, 238, 273, 284-285, 289Neurodegenerative Cohort

Definition

Decedents whose primary cause of death was a neurodegenerative-related condition

ICD10 Code Range

G10-G14, G20-G26, G31-G32, G35-G37

Detail

Systemic atrophies primarily affecting the central nervous system G10-G14

Extrapyramidal and movement disorders G20-G26

Other degenerative diseases of the nervous system G31-G32

Demyelinating diseases of the central nervous system G35-G37

Exclusions:

Alzheimer's disease G30

ICD9 Code Range

330-335, 341, 348, 359

ORGAN FAILURE COHORT

Definition

Decedents whose primary cause of death was organ failure.

ICD10 Code Range

E09,E10-E14, E15-E16, E70-E89, F04-F09, I05-I09, I10-I15, I20-I25, I26-I28, I30-I52, I60-I69, J09-J18, J30-J39, J40-J44,J45-J47, J60-J70, J80-J84, J85-J86, J90-J94, J95-J99, N00-N08, N10-N16, N17-N19, O10-O16, U04, K70-K77, K55, K56

Detail

Conditions more likely to features as the primary or first secondary cause of death:

- Impaired glucose regulation and diabetes mellitus E09-E14
- Other disorders of glucose regulation and pancreatic internal secretion E15-E16
- Metabolic disorders E70-E89 includes Cystic fibrosis, Amyloidosis, Dehydration
- Organic, including symptomatic, mental disorders F00-F09 includes Dementia
- Chronic rheumatic heart diseases I05-I09
- Hypertensive diseases I10-I15
- Ischaemic heart diseases I20-I25
- Pulmonary heart disease and diseases of pulmonary circulation I26-I28

Conditions more likely to feature as a secondary cause of death:

- Other forms of heart disease I30-I52
- Cerebrovascular diseases I60-I69
- Influenza and pneumonia J09-J18 includes avian influenza virus, Swine Flu
- Other diseases of upper respiratory tract J30-J39
- Chronic lower respiratory diseases J40-J47 includes Bronchitis, COAD, Asthma
- Lung diseases due to external agents J60-J70
- Other respiratory diseases principally affecting the interstitium J80-J84 includes ARDS

- Suppuratives and necrotic conditions of lower respiratory tract J85-J86
- Other diseases of pleura J90-J94 includes Pneumothorax
- Other diseases of the respiratory system J95-J99 includes Respiratory failure
- Glomerular diseases N00-N08
- Renal tubulo-interstitial diseases N10-N16
- Kidney failure N17-N19
- Oedema, proteinuria and hypertensive disorders in pregnancy, childbirth and the puerperium O10-O16
- Severe acute respiratory syndrome [SARS] U04
- Diseases of Liver K70-K77
- Vascular disorders of intestine K55
- Paralytic ileus and intestinal obstruction without hernia K56

50% Preventable, 50% Treatable Subsets

Ischaemic Heart Disease (IHD) 0-74 years: I20-I25

Cerebrovascular Diseases (CVD) 0-74 years: I60–I69

Diabetes 0-74 years: E10-E14

Preventable Subsets

Chronic Obstructive Pulmonary Disease (COPD) 45-74 years: J40-J44

Conditions associated with 1000 or more deaths (based on 2005 NSW decedents):

- Other forms of heart disease I30-I52 (16181) deaths
- Ischaemic heart diseases I20-I25 (15341) deaths
- Cerebrovascular diseases I60-I69 (8116) deaths
- Influenza and pneumonia J09-J18 includes avian influenza virus, Swine Flu (5623) deaths
- Kidney failure N17-N19 (5190) deaths
- Chronic lower respiratory diseases J40-J47 includes Bronchitis, COAD, Asthma (4512) deaths

- Organic, including symptomatic, mental disorders F00-F09 includes Dementia (4469) deaths
- Other diseases of the respiratory system J95-J99 includes Respiratory failure (3849) deaths
- Impaired glucose regulation and diabetes mellitus E09-E14 (3838) deaths
- Lung diseases due to external agents J60-J70 (1806) deaths
- Diseases of Liver K70-K77 (1652) deaths
- Metabolic disorders E70-E89 includes Cystic fibrosis, Amyloidosis, Dehydration (1486) deaths
- Other respiratory diseases principally affecting the interstitium J80-J84 includes ARDS (1219) deaths
- Pulmonary heart disease and diseases of pulmonary circulation I26-I28 (1101) deaths

Conditions associated with fewer than 1000 deaths (based on 2005 NSW decedents):

- Hypertensive diseases I10-I15 (546) deaths
- Paralytic ileus and intestinal obstruction without hernia K56 (489) deaths
- Vascular disorders of intestine K55 (282) deaths
- Chronic rheumatic heart diseases I05-I09 (193) deaths
- Renal tubulo-interstitial diseases N10-N16 (163) deaths
- Glomerular diseases N00-N08 (64) deaths
- Other diseases of pleura J90-J94 includes Pneumothorax (58) deaths
- Suppurative and necrotic conditions of lower respiratory tract J85-J86 (51) deaths
- Other diseases of upper respiratory tract J30-J39 (38) deaths
- Other disorders of glucose regulation and pancreatic internal secretion E15-E16 (15) deaths

ICD9 Code Range

046, 078, 250, 270-272, 275-277, 290, 293-294, 310, 330, 394-397, 402-404, 410-411, 413, 415-416, 420-434, 436-437, 440-444, 447-448, 470-478, 480-484, 487, 491-496, 501-507, 511-513, 518-519, 580-586, 590-591, 593, 599, 642, 646, 997

SUDDEN DEATH COHORT

Definition

Decedents whose primary cause of death was due to sudden (unexpected) death.

ICD10 Code Range

S00-S09, S10-S19, S20-S29, S30-S39, S40-S49, S50-S59, S60-S69, S70-S79, S80-S89, S90-S99, T00-T07, T08-T14, T15-T19, T20-T31, T33-T35, T36-T50, T51-T65, T66-T78, T79, T80-T88, T89, V00-V99, W00-X59, X60-X84, X85-Y09, Y10-Y34, Y35-Y36, Y40-Y84, V00-V99, W00-X59, X60-X84, X85-Y09, Y10-Y34, Y35-Y36, Y40-Y84

- Injuries to the head S00-S09
- Injuries to the neck S10-S19
- Injuries to the thorax S20-S29
- Injuries to the abdomen, lower back, lumbar spine and pelvis S30-S39
- Injuries to the shoulder and upper arm S40-S49
- Injuries to the elbow and forearm S50-S59
- Injuries to the wrist and hand S60-S69
- Injuries to the hip and thigh S70-S79
- Injuries to the knee and lower leg S80-S89
- Injuries to the ankle and foot S90-S99
- Injuries involving multiple body regions T00-T07
- Injuries to unspecified part of trunk, limb or body region T08-T14
- Effects of foreign body entering through natural orifice T15-T19
- Burns T20 - T31
- Frostbite T33 - T35
- Poisoning by drugs, medicaments and biological substances T36-T50
- Toxic effects of substances chiefly non-medicinal as to source T51-T65
- Other and unspecified effects of external causes T66-T78
- Certain early complications of trauma T79
- Complications of surgical and medical care, not elsewhere classified T80-T88

- Other complications of trauma not elsewhere classified T89
- Transport accidentsV00 - V99
- Other external causes of accidental injury W00-X59
- Intentional self-harm X60 - X84
- Assault X85 - Y09
- Event of undetermined intent Y10-Y34
- Legal intervention and operations of war Y35-Y36
- Complications of medical and surgical care Y40-Y84

ICD9 Code Range

798, 800-848, 850-855, 858, 860-876, 878-888, 890-914, 916-976, 978, 980

FRAILTY COHORT

Definition

Decedents whose primary cause of death was frailty or debility.

Pattern Matching

The ICD coding system does not reliability identify frail individuals. Accordingly it was necessary to define frailty as the free text cause of death matching any of the following patterns:

‘* AGE’

‘*ADVANCED AGE*’

‘*DEBILITY*’

‘*FRAIL*’

‘*FRAILTY*’

‘*GENERAL DECLINE*’

‘*INANITION*’

‘*OLD AGE*’

‘*SENILITY*’

‘*SENILTY*’

‘AGE’

‘AGED’

PROBABILISTIC PATTERN MATCHING ALGORITHM

Overview

For 2010 decedents and for other decedents where ICD-coded cause of death was not available, it was necessary to develop a probabilistic pattern matching algorithm to assign patient cohorts using free-text cause of death information sourced from Part I of the death certificates.

The patient cohorts assigned were:

1 Neoplasms

2 Neurodegenerative

3 Organ Failure

4 Sudden Death

5 Frailty

2005 mortality data for NSW and Qld contained both free-text cause of death and ICD-coded cause of death and was used to calibrate the pattern matching algorithm.

STEP 1 – Word Sequence Extraction

Words sequences from the free-text cause of death information (from one to three words in sequence) were extracted as per the following example:

"(I) METASTATIC LARGE CELL CARCINOMA OF LUNG, 6 MONTHS" was output as:

```
METASTATIC
      LARGE
            CELL
                  CARCINOMA
                              OF
                                    LUNG
METASTATIC LARGE
      LARGE CELL
            CARCINOMA OF
                  OF LUNG
METASTATIC LARGE CELL
      LARGE CELL CARCINOMA
            CELL CARCINOMA OF
                  CARCINOMA OF LUNG
```

Common words (e.g. “FROM”, “WHERE” etc.) and other non-clinical terms such as dates or geographical names were excluded from the word sequences where possible.

STEP 2- Calculation of Word Sequence Frequencies

The frequencies of agreement for each word sequence against the patient cohort of interest were derived along with the frequencies of agreement in unrelated patient cohorts. The ratio of agreement frequencies was subsequently calculated:

$$Ratio(word\ sequence_i, patient\ cohort_n) = \frac{Frequency(word\ sequence_i, patient\ cohort_n)}{Frequency\ (word\ sequence_i, other\ patient\ cohorts)}$$

A denominator of 100 was substituted for cases with zero agreements in other cohorts to allow these cases to be separately identified. 3 contains agreement frequencies and ratios for selected cause of death word sequences for 2005 NSW decedents for the Neoplasm Cohort:

Table 3 - Select Word Sequence

Word Sequence	Ratio	Agree - Neoplasm Cohort	% Neoplasm Cohort	Agree – Other Cohorts
RECTAL CANCER	"18,300"	183	1.38%	0
RECTAL CARCINOMA	"15,000"	150	1.13%	0
CARCINOMA OF COLON	"14,600"	146	1.10%	0
CARCINOMA BOWEL	"3,700"	37	0.28%	0
DISSEMINATED CANCER	"2,900"	29	0.22%	0
ADENOCARCINOMA OF PANCREAS	"2,400"	24	0.18%	0

Word Sequence	Ratio	Agree - Neoplasm Cohort	% Neoplasm Cohort	Agree – Other Cohorts
CANCER COLON	“1,300”	13	0.10%	0
BRAIN CANCER	“1,100”	11	0.08%	0
CANCER OF STOMACH	“1,100”	11	0.08%	0
PROGRESSIVE METASTATIC	“1,100”	11	0.08%	0
CANCER OF RECTUM	“900”	9	0.07%	0
ADENOCARCINOMA OF BOWEL	“800”	8	0.06%	0
PANCREAS CANCER	“800”	8	0.06%	0
ADENOCARCINOMA RIGHT LUNG	“600”	6	0.05%	0
SIGMOID COLON CANCER	“500”	5	0.04%	0
PANCREATIC CANCER	258.0	258	1.95%	1
MALIGNANT ADENOCARCINOMA	“100”	1	0.01%	0
CARCINOMA COLON	99.0	99	0.75%	1
CELLULAR CARCINOMA	95.0	95	0.72%	1
LARGE CELL CARCINOMA	68.0	68	0.51%	1
CANCER	53.3	4580	34.62%	86

Word Sequence	Ratio	Agree - Neoplasm Cohort	% Neoplasm Cohort	Agree – Other Cohorts
METASTATIC TRANSITIONAL CELL	51.0	51	0.39%	1
METASTATIC LUNG CARCINOMA	42.0	42	0.32%	1
CELL LYMPHOMA	35.0	70	0.53%	2
NON HODGKINS LYMPHOMA	31.0	31	0.23%	1
PROSTATIC CARCINOMA	30.0	30	0.23%	1
CHOLANGIO CARCINOMA	28.0	28	0.21%	1
CARCINOMA OF BLADDER	26.0	78	0.59%	3
METASTATIC BOWEL CARCINOMA	23.0	23	0.17%	1
LEUKEMIA	22.0	44	0.33%	2
NON HODGKIN'S LYMPHOMA	21.7	65	0.49%	3
MONOCYTIC LEUKAEMIA	18.0	18	0.14%	1
CARCINOMA LEFT	17.0	51	0.39%	3
CANCER OF BOWEL	15.0	15	0.11%	1
ADVANCED CANCER	14.0	14	0.11%	1
LARGE CELL LYMPHOMA	13.0	13	0.10%	1

Word Sequence	Ratio	Agree - Neoplasm Cohort	% Neoplasm Cohort	Agree – Other Cohorts
BASAL CELL CARCINOMA	11.0	11	0.08%	1
TERMINAL METASTATIC	7.0	7	0.05%	1
ACUTE MYEOBLASTIC LEUKAEMIA	6.0	6	0.05%	1

The Neoplasm Cohort was defined as per section **Error! Reference source not found.**.

The number of decedents in the Neoplasm Cohort for 2005 was 13,299. The number of decedents in other cohorts for 2005 was 31,400.

STEP 3 – Choosing Word Sequences for the Pattern Matching Algorithm

A balance was required between the use of highly specific terms that might only be present in the calibration data sets vs more generalised terms that might be more widely applicable to other data sets.

Table 4 - Term Applicability – Cancer Example

Term	Ratio	Cases	Specificity	Applicability
CANCER	53.3	4580	Medium	High
PANCREATRIC CANCER	258.0	258	Medium	Medium
RUPTURE OF CANCEROUS	“100”	1	High	Lower

For example, in 4 although the term “RUPTURE OF CANCEROUS” is highly specific, it is only present for one decedent and might not be applicable to other datasets. While the term “PANCREATIC CANCER” has medium specificity, similar terms for every other type of cancer would also need to be included in the pattern matching algorithm. In comparison, the term “CANCER” has medium specificity and high applicability. In this example it was preferable to add the word “CANCER” to the pattern matching algorithm, rather than specifying every specific type of cancer.

Table 5 - Term Applicability - Neoplasm Example

Term	Ratio	Cases	Specificity	Applicability
NEOPLASM	5.75	23	Low	High
LUNG NEOPLASM	“400”	4	High	Medium
MALIGNANT NEOPLASM	“800”	800	High	Medium

In the second example in the term Neoplasm has a high applicability, but has a low specificity, most likely due to the number of NEOPLASMS which are either being benign or are secondary to the cause of death. In this example, it was advisable to add more specific terms to the pattern matching algorithm such as “LUNG NEOPLASM” or “MALIGNANT NEOPLASM”.

STEP 4 – Determination of Cut-offs

An model developed using MS Access 2010 Visual Basic for Applications was used to model the impact of various ratio cut-offs upon overall cohort matching accuracy, as well as the number of false positives and false negatives. Optimal results were achieved using word sequences with a specificity ratio of at least 6.0 and more than two cases.

Table 6 - NSW Decedents - Neoplasms Cohort - Matching Accuracy

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	12,799	96.7%		
Yes	No	368		2.8%	
No	Yes	314			2.4%

PATTERN MATCHING METRICS

2005 Decedents

Neoplasms

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	12,799	96.7%		
Yes	No	368		2.8%	
No	Yes	314			2.4%

Neurodegenerative

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	511	87.5%		
Yes	No	70		12.0%	
No	Yes	92			15.8%

Organ Failure

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	18,508	81.0%		
Yes	No	3,876		17.0%	
No	Yes	3,247			14.2%

Metrics exclude 819 missing free-text CODs.

Sudden Death

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	1,323	56.1%		
Yes	No	281		11.9%	
No	Yes	149			6.3%

Metrics exclude 752 missing free-text CODs.

2005 Qld Decedents

Neoplasms

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	7,185	95.6%		
Yes	No	313		4.2%	
No	Yes	162			2.2%

Metrics excludes 20 cases where COD was "*NOT YET DETERMINED*" and 20 cases with missing ICD codes.

Neurodegenerative

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	511	87.5%		
Yes	No	70		12.0%	
No	Yes	92			15.8%

Metrics excludes 3 cases of "NOT YET DETERMINED" and one case of a missing ICD10 code.

Organ Failure

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	18,508	81.0%		
Yes	No	3,876		17.0%	
No	Yes	3,247			14.2%

Metrics exclude 309 cases of "NOT YET DETERMINED" and 164 cases of null ICD10 codes.

Sudden Death

ICD10 Code Agrees	Pattern Match Agrees (search Part I)	No. Decedents	Match (%)	`False (-) %	False (+) %
Yes	Yes	1,323	56.1%		
Yes	No	281		11.9%	
No	Yes	149			6.3%

Metrics exclude 278 cases of “NOT YET DETERMINED" and 44 cases of null ICD10 codes.