Comparing name-based and event-based strategies for data linkage

A study linking hospital and residential aged care data for Western Australia

The Australian Institute of Health and Welfare is Australia's national health and welfare statistics and information agency. The Institute's mission is *better information and statistics for better health and wellbeing*.

Please note that as with all statistical reports there is the potential for minor revisions of data in *Comparing name-based and event-based strategies for data linkage: a study linking hospital and residential aged care data for Western Australia* over its life. Please refer to the online version at <www.aihw.gov.au>.

DATA LINKAGE SERIES
Number 3

Comparing name-based and event-based strategies for data linkage

A study linking hospital and residential aged care data for Western Australia

Rosemary Karmel
Australian Institute of Health and Welfare

and

Diana Rosman
Department of Health, Western Australia

2007

Australian Institute of Health and Welfare
Canberra

AIHW cat. no. CSI 3

This publication is part of the Australian Institute of Health and Welfare's Data linkage series. A complete list of the Institute's publications is available from the Institute's website <www.aihw.gov.au>.

ISSN 1833 1238

ISBN 978 1 74024 693 4

Suggested citation

AIHW: Karmel R & Rosman DL 2007. Comparing name-based and event-based strategies for data linkage: a study linking hospital and residential aged care data for Western Australia. Data linkage series no. 3. Cat. no. CSI 3. Canberra: Australian Institute of Health and Welfare.

Any enquiries about or comments on this publication should be directed to:

Community Services Integration and Linkage Unit
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601

Phone: (02) 6244 1000

Published by the Australian Institute of Health and Welfare
Printed by

Contents

Acknowledgments

This study could not have been carried out without the vital input of a number of people. The contributions of several people within the Australian Institute of Health and Welfare are especially acknowledged. In particular, the event-based linkage used in the study builds on an idea originally developed by Diane Gibson for linking event data when name data are not available. In addition, Peter Braun put considerable time and effort into preparing the residential aged care datasets used in the data linkage and analysis. Valuable comments on drafts of the report were provided by Phil Anderson and Jonas Lloyd.

We also gratefully acknowledge the staff of the Health Information Linkage Branch (formerly the Data Linkage Unit) in the Western Australian Department of Health on whom the accuracy and completeness of the Western Australian Data Linkage System relies. Within the Health Information Linkage Branch we thank Emma Brook who undertook the linkage to the Residential Aged Care records and Carol Garfield who provided technical support and managed the linkage and extraction processes. John Bass made significant contributions to the project as facilitator, coordinator and custodian of the linkage keys for the cross-jurisdictional arrangements.

Discussions with Dr Terry Neeman, consultant statistician with Covance, and Greg Griffiths concerning the analysis and presentation of comparison results were also very fruitful.

This project was funded primarily by the Australian Institute of Health and Welfare and the Western Australian Department of Health. The Australian Health Ministers' Advisory Council through its Statistical Information Management Committee also contributed funds towards this project.

Symbols in tables

—	Nil or rounded to zero
. .	Not applicable
n.p.	Not publishable

Abbreviations

ABS	Australian Bureau of Statistics
ACAT	Aged Care Assessment Team
ACCMIS	Aged and Community Care Management Information System
ACT	Australian Capital Territory
BSESLA	basic SLA group E linkage strategy
CPC	constrained postcode E linkage strategy
CPC_f	constrained postcode E linkage strategy, includes 2-digit postcode matching
CPC_s	constrained postcode E linkage strategy, not including 2-digit postcode matching
CSLA	constrained SLA group E linkage strategy
$CSLA_f$	constrained SLA group E linkage strategy, includes 2-digit postcode matching
$CSLA_s$	constrained SLA group E linkage strategy, not including 2-digit postcode matching
DOB	date of birth
DoHA	Australian Government Department of Health and Ageing
E linkage	event-based data linkage (name data not used)
HiL	Health Information Linkage Branch, Western Australian Department of Health
ICD-10-AM	International Classification of Diseases, 10th revision, Australian modification
MDC	major diagnostic category
NHMD	National Hospital Morbidity Database
N linkage	name-based data linkage
NSW	New South Wales
NT	Northern Territory
PPV	positive predictive value
Qld	Queensland
RAC	residential aged care
RCS	resident classification scale (RCS1–4 = high care, RCS5–8 = low care)
SA	South Australia
SLA	statistical local area
stat. adm.	statistical admission into hospital episode
stat. sep.	statistical separation from hospital episode
Tas	Tasmania
US	United States of America
Vic	Victoria
WA	Western Australia
WADLS	Western Australian Data Linkage System
YOB	year of birth

Key points

Background to the study:

1. Statistical analysis of movement from hospital to residential aged care would provide a strong evidence base to inform policy development for the hospital–aged care interface. Linking currently available data for such analysis has the advantages of reduced cost, timeliness and no additional data collection imposition.
2. Data linkage of records for individuals is commonly carried out using detailed demographic data, including name and/or a person identification number. While neither name data nor a unique person identifier are available on the national datasets for both the hospital and residential aged care systems for such data linkage, a range of demographic and event data is available.
3. Two earlier projects by the Australian Institute of Health and Welfare investigated the theoretical and practical feasibility of a linkage strategy (termed event-based linkage) which uses demographic data (excluding name) and geographic and event data to link episode records from the hospital and residential aged care national databases.

Current study:

4. The current study directly compares results from linking hospital and residential aged care data using a name-based linkage strategy without specific event information, with those from a group of event-based linkage strategies which did not use any name information. For this comparison, unidirectional hospital-to-residential-care event links were selected, from the full set of annual events in Western Australia, independently using the name-based and event-based linkage strategies.
5. Prior to detailed assessment, the Institute's event-based linkage was refined through initial comparisons with the name-based linkage established in the Western Australian Department of Health, with the results being used to distinguish between effective and poorly performing match tactics, thereby preventing the collection of large quantities of incorrect, or false, links.
6. Within transition destinations, links identified by these refined event-based strategies were highly reliable: overall 98% of links identified by the event-based strategies were true matches when compared with links from the name-based strategy.
7. As expected from earlier studies, the relatively high incidence of missed links (8–14% among preferred strategies) resulted in the event-based linkage strategies underestimating the volume of flow from hospital to residential aged care. There is also the potential for bias if these links relate to a particular subset of the population.
8. However, this study confirmed that the event-based linkage strategies resulted in linked data that were representative of the name-based linked data in terms of the distributions across key variables and within types of transitions. In particular, illustrative examples looking at patterns of use and characteristics of people making particular types of transitions between the two sectors show that analyses using linkage results derived from the name-based and event-based strategies lead to very similar conclusions.
9. Overall, the reliability of the links identified by event-based strategies—especially when movement is known to have occurred, dates are expected to coincide and region information is reliable—suggests that such an approach could be used in other areas. In

particular, although this has not yet been tested, a similar strategy could be used to derive whole-of-stay hospital records for cases when patients move within the hospital sector. In such cases, date of transfer and region information (postcode) should be highly consistent on the before- and after-transfer records.

10. The detailed technical processes and close collaboration required for this project have confirmed that detailed knowledge of both the service systems and the data collection practices within those systems are essential when identifying transition events using any data linkage method.

11. Key findings for event-based linkage are that:

    - Using detailed comparisons of name-based and event-based links, and expanding the event data used to identify links, it has been possible to improve the general performance of the event-based linkage strategy as put forward in the initial feasibility study; in particular, event-based linkage strategies that adjust the linkage procedures according to available information (constrained event-based linkage) perform better than those that apply the same approach to all records (basic 1event-based linkage).
    - Conservative event-based linkage more often tends to miss matches than make false matches, with inconsistent event date and/or region data on the two databases being the main causes of missed links. This sensitivity of event-based linkage varies with both the strategy and transition destination.
    - Within event-based linkage, broader geographic matching can be used for areas with small populations, but not among those with denser populations. Within these limits, methods using broader regions when matching perform better than those using smaller regions.

A broad summary of the methodology and results of the study follows.

Summary

Background

The interface between acute hospital care and residential aged care has long been recognised as an important issue in aged care services research. Despite this, existing national data provide very poor information on the movement of clients between the residential and acute care sectors. Current national datasets on the two sectors are derived from administrative collections, and have been designed primarily to provide information on the specific program which they describe, rather than to provide information on program interfaces or system-level information.

Data linkage is a statistical tool that can be used to link data from different sources, thereby expanding the types of statistical investigations that can be carried out (including analysis of movement over time) without increasing the reporting load of service providers or requiring special surveys. Given the existence of national datasets for the two sectors, this suggests that data linkage could be used to develop datasets suitable for investigating movement between the hospital and aged care sectors.

While neither name nor a common person identification number are available for linking data from the two sectors, some demographic data are available. In addition, information on transition dates, that is entry and exit dates, is available for all periods of hospitalisation and use of residential aged care. To see if such event data are sufficient to allow data linkage, investigations into the feasibility of linking hospital morbidity and residential aged care datasets using a combination of demographic and event data were conducted by the Australian Institute of Health and Welfare (AIHW). Findings from the feasibility study suggested that the set of linked client records resulting from an event-based anonymous linkage strategy could provide a valuable source of information on the client characteristics and service use patterns associated with movements between the two sectors (AIHW 2003).

In the feasibility study, matching was based on date of birth, sex, region of usual residence, and hospital separation and residential aged care (RAC) entry dates. In the absence of validation against a gold standard linkage, doubts concerning the efficacy of the strategy were raised because of the lack of either name or a common person identifier on the two datasets. This issue has been addressed using two distinct methods: first, by employing statistical theory to investigate the effectiveness of the AIHW event-based strategy in a range of linkage situations (AIHW: Karmel 2004); second, by measuring any differences between the two approaches through direct comparisons between a name-based linkage and the event-based linkage. It is these comparative analyses that are the subject of this report. Opportunities to compare such linkages against a 'gold standard' are rare but have been undertaken previously by comparing deliberatively reduced identifiers within fully identified data extracts (for example, Rosman 1996).

In the current analysis, we compare two distinct strategies for linking hospital and RAC events. The first is a name-based process carried out by the Health Information Linkage Branch (HiL) of the Western Australian Department of Health, here termed N linkage. The second is the AIHW event-based linkage strategy—E linkage—which matches event records using date of birth, sex, geographic region of usual residence and event dates. The scope of the comparisons was limited to movements from hospital to RAC in Western Australia during the financial year 2000–01. Western Australia was chosen for this study because of its unique position in having a well-established data linkage system containing links within and between a number of health data collections which was recently expanded to include links to RAC data.

In summary, the report:

- examines ways to refine the original E linkage strategy
- examines missed and false links made by E linkage (when compared with N linkage)
- compares the E and N links as a whole, looking for similarities and differences in distribution across variables of interest
- compares results from analyses using E and N linked datasets to identify the types of analyses for which the results are comparable and those for which there are differences that need to be taken into account.

Ethical approval for the project was obtained from the AIHW Ethics Committee, and permission to use the relevant data was obtained from both the Steering Committee for Cross-Jurisdictional Linked Administrative Health Data (a joint Australian and Western Australian government committee) and the Confidentiality of Health Information Committee of the Western Australian Department of Health.

Context

The purpose of the data linkage described here was to identify movements of people from hospital into RAC. Before discussing the linkage it is important to understand both the circumstances under which people move between the two sectors, and the data available for linkage and analysis.

Movement between services

When people finish an episode of care in hospital there are a number of possible outcomes: they may transfer within the hospital system, transfer to a residential health care service, go back to their home in the community, go home to RAC, enter RAC (either temporarily for respite care or permanently), or their hospital episode may have ended with their death. Similarly, people entering RAC may be coming from a range of settings: they may be moving from their home into RAC to live either permanently or for respite care, they may be returning after an episode in hospital (commonly while on RAC hospital leave, but possibly also while on social leave from RAC), or they may be entering residential care—either as a permanent admission or for short-term care (a respite care admission)—at the end of a period of hospitalisation (Section 2).

When looking at movement from hospital into RAC, the aim of the data linkage is to link people's exits from hospital with their relevant entries into residential aged care, thereby allowing analysis of patterns of movements between the two sectors.

Data

Datasets were limited to events within Western Australia occurring in 2000–01 (and a few days either side). In addition, as 95% of admissions into RAC are for people aged 65 years and over (AIHW 2002), the data were limited to events for people who were aged at least 65 years by 30 June 2001. Two extracts were obtained for both the hospital and RAC data: one to be used to establish links (a linkage dataset), and a second to be used for analysing link bias (an analysis dataset) (Section 3).

Hospital data

Hospital event data were extracted from the Hospital Morbidity Data System, which is managed and maintained within the Information Collection and Management Directorate of the Western Australian Department of Health. This system supplies similar data for the National Minimum Dataset held by the AIHW. The extract for this project was organised and extracted through the processes established for the Western Australian Cross Jurisdictional Data Linkage Protocol. The extract contained one record per hospital discharge, but with statistical discharges and same-day events removed. These inpatient hospital episodes included both public and private hospital separations for all of the Western Australian population for the period of interest. The hospital data items for linkage and those for analysis were supplied separately.

For the E linkage strategy (carried out by the AIHW) a special extract of hospital separations was derived. This contained only data that are available nationally on the National Hospital Morbidity Database (NHMD) so that the E linkage would reflect results that would be obtained if the strategy were to be applied to link NHMD and RAC data nationally or in other jurisdictions. The hospital linkage dataset contained only the information required for establishing and checking E links, including date of birth, sex, postcode of usual residence, admission and separation dates, and modes of hospital admission and separation.

Two sets of records were excluded from the hospital dataset used for E linkage: hospital episodes that ended with a transfer within the hospital system, and same-day hospital episodes (that is, without staying overnight). The former were excluded as these people do not leave the hospital system at the end of the episode of care, while the latter were excluded to avoid unnecessary erroneous matches as people are unlikely to go from hospital into residential aged care on a single day. For N linkage, it was not necessary to exclude within sector transfers as the person-based linkage allows such transfers to be combined into a single period of hospitalisation.

In 2000–01, in Western Australia there were just over 86,200 hospital separations with a non-transfer discharge that lasted one or more nights. Note that to allow for small gaps between hospital separation and RAC admission (and for small differences in recording dates), separations for a few days either end of the financial year were also included in the dataset for linkage.

To allow detailed comparison of the linked datasets resulting from the N and E strategies, a dataset specifically for analysis was also derived. This analysis dataset contained the information to be used when investigating possible analytical bias, and included age, sex, region of usual residence, country of birth, marital status, modes of hospital admission and separation, length of stay, and hospital diagnosis and procedure variables.

Residential aged care data

The RAC data included all permanent and respite admissions and hospital and social leave events for 2000–01, totalling slightly more than 19,600 events for Western Australia. The data were derived from the Australian Department of Health and Ageing's Aged and Community Care Management Information System (ACCMIS).

As for the hospital data, two extracts were obtained: one for linkage and one for analysis. While the N linkage process is based on established person links for people using hospital and RAC services, the appropriate event links still had to be derived. Therefore, both the N and E linkage used the same RAC event linkage file to establish final event links.

The RAC linkage file, containing data for establishing and checking event links, included date of birth, sex, postcode of usual residence (for leave events this is the RAC facility's postcode), country of birth, marital status, event type, admission and discharge dates, leave event start and end dates, place and date of Aged Care Assessment Team (ACAT) assessment, and mode of discharge. The RAC analysis file, with information to be used when investigating possible analytical bias, contained age, sex, region of residence before current event, region of RAC facility, country of birth, marital status, event dates, care needs, place and date of ACAT assessment, data on the previous RAC admission event (if applicable), and mode of discharge.

The linkage strategies

Name-based (N) linkage

The name-based linkage was undertaken by the Health Information Linkage Branch (HiL) within the Western Australian Department of Health using the two-phase cross-jurisdictional linkage protocol described by Kelman et al. (2002). The linkage process uses as much personal information as is available to create and load links as they are discovered. The links are stored in the Western Australian Data Linkage System (WADLS) which is a dynamic on-line system able to be accessed and updated continuously by linkage staff within HiL (Section 4). Links between a large variety of health-related datasets are established, stored and maintained in this system. Linkages to RAC records for 1990–2003 were performed during 2004. Each link created between a RAC client record and any of the other sources contributing to the WADLS records was loaded into quarantined tables within the WADLS so that access could be managed separately from the main system, but links could be updated at a later stage. For the RAC linkage, demographic information included surname, given names, sex, date of birth and address. Information on the dates of events within the RAC was not available for linkage. Separate linkages between RAC and Western Australian electoral, ambulance, hospital, emergency and death records were performed.

Using HiL's hospital–RAC person links, all relevant hospital and RAC event records were retrieved from their respective data custodians and then matched to form combined strings of hospital events and RAC events for the same person. The most appropriate hospital–RAC event link was then chosen by measuring the closeness of hospital and RAC event dates. Where there was a choice between matches to different RAC event types, matches to RAC hospital leave had priority over matches to admissions.

Overall, the N strategy resulted in 8,106 links between hospital separations and entry (or re-entry) into RAC.

Event-based (E) linkage

The event-based E linkage strategies link records by using limited demographic information in conjunction with event dates (Section 5). Additional data on event characteristics can also be used. In this study, the information used to establish links included: date of birth, sex, postcode of usual residence, postcode of hospital, episode start and end dates, hospital episode admission and separation modes, and time and place of assessment for entry into RAC. The resulting linked dataset varies depending on whether all this information is used when identifying links, and how particular data items are used to specify the linkage process.

Overall, for this project three types of E linkage were investigated:

- constrained matching within statistical local area (SLA) group (abbreviated to CSLA)
- constrained matching within postcode (CPC)
- basic matching within SLA group (BSESLA).

To ensure that the privacy of individuals was maintained, the E linkage was carried out within the AIHW using the Institute's protocol for protecting privacy when carrying out data linkage, and data were protected according to standard AIHW procedures (AIHW 2006).

Constrained E matching

The purpose of constrained E matching is to find the best match using all available event date information and event descriptors. To achieve this, matching procedures are tailored specifically for comparisons between different subsets of RAC and hospital events defined in terms of their type and/or admission and separation characteristics. Matches are then established using date of birth, sex, region, and event dates. Knowledge of both the service systems and data collections is used to determine the specific comparisons, with the specific matching processes varying depending on: whether a hospital episode began with a within-sector transfer, the patient's reported destination after discharge from hospital, the type of RAC event (admission or leave event), and the place and time of ACAT assessment for a RAC admission. Because two dates are available for RAC hospital leave, and the related hospital episode may end in a number of ways, match procedures for these events are the most complicated (see Appendix 2).

Constrained E matching is carried out in two stages. Initial matches are selected using 1-to-1 probability matching. Relatively broad match criteria are used to identify possible matches between particular subsets of RAC and hospital data, with the selected RAC record match for a particular hospital record being that with the highest probability of matching the hospital record given the data used for matching; for this project, 12 hospital–RAC dataset pairs were used and each pair was compared using up to seven distinct match passes. Some variation in date of birth, region and event dates is allowed when establishing matches. Finer match rules are then applied to select the final matches, using, first, deterministic rules to exclude poor matches and then specified match priorities to choose between duplicate links.

Two variants of constrained E linkage were considered, differing only in the size of the region used when establishing links. CSLA linkage is based on matching within SLA groups, where an SLA group is that set of SLAs that overlaps a postcode and two postcodes are said to match if they have a common SLA in their SLA groups. CPC linkage restricts matching to regions based on the four digits of the reported postcode, and so generally uses smaller match regions than CSLA.

Before refining the constrained strategies using information from initial comparisons with N linkage, the CSLA and CPC strategies resulted in 7,802 and 7,781 linked events, respectively.

Basic E matching

In addition to constrained linkage, basic E linkage was used. Under this strategy, the same linkage process is used for all hospital and RAC events, irrespective of their admission and separation characteristics. As a consequence, match data are limited to information available for all events, namely date of birth, sex, postcode of usual residence and a single movement date (that is, hospital separation matching a RAC entry date). Matches are selected using 1-to-1 probability matching, although only minor variation from exact matching is allowed (in event dates).

Basic matching using SLA group as the match region (BSESLA) was used in the initial feasibility study of hospital–RAC event-based linkage (AIHW 2003). It is therefore important to compare the results from this approach with those obtained using the person-based linkage. It is also useful to compare the effectiveness of the BSESLA strategy with that of the constrained approaches to measure the gains in moving from a simple strategy to a more complex one.

Using BSESLA linkage resulted in 6,693 event matches.

Comparing linkage strategies

When linking records, four outcomes are possible: a true link, no link, a false link (false positive) and a missed link (false negative). The correspondence between two strategies can be gauged by seeing how many of the links are the same and how many are different. While some constraints were imposed on the HiL by the RAC data provider when linking Western Australia RAC and health data, the use of name and address in N linkage, and the availability of name and address reporting history across a range of health service events, results in this linkage being highly reliable. Consequently, in this study it was used as the reference standard against which the E linkage results were compared, that is, to determine whether an E link was 'true' or 'false'. Linkage checks indicate a very low error rate in the established WADLS.

For this project, two key measures were used when comparing matches:

- Positive predictive value (PPV): the percentage of E links that are true links
  = E true links/E links
- Sensitivity: the percentage of all links that are identified by the E linkage strategy
  = E true links/N links.

Using N linkage as the reference standard, N-only links represent those links missed by E linkage, while E-only links represent false matches made under the strategy.

Refining constrained E linkage

As stated above, constrained E linkage was carried out by matching between 12 hospital–RAC dataset pairs, with up to seven different (and pair-specific) match criteria being applied to each pair. The reliability of a particular match criterion can be gauged by comparing the resulting E links with N links. In particular, the PPV of E matches made via specific match criteria can be used to detect particular match passes that result in unacceptably high levels of false matches. Such passes can then be dropped from the constrained E linkage strategy without compromising the reproducibility of the linkage process for other jurisdictions (Section 6).

To this end, poorly performing matching algorithms used in the constrained E linkage were identified using the condition that the underlying PPV should be above 60% with 95% probability. As a result, a number of match criteria were tightened or dropped altogether. In addition, analysis of the distribution of the population across geographic areas indicated that allowing substantial variation in the region match by relaxing the match regions to those defined by the first two digits of a postcode could lead to relatively low PPVs for some of the resulting regions in the more populous states.

After refining the constrained strategies as above, the CSLA strategy resulted in identifying 7,595 transitions from hospital into RAC when broad region matching was allowed (that is, matching within regions identified by the first two digits of the person's postcode) and 7,253 when it was excluded. For CPC linkage, the corresponding numbers were 7,587 and 7,078 respectively. For both the SLA- and postcode-based constrained linkage approaches, removing match passes that lead to unacceptably high false match rates improved their overall PPV by two percentage points (up from slightly under 96% to almost 98% for both strategies).

Missed links

After refining the constrained strategies, compared with N linkage the E strategies still missed some links and falsely identified others, with missed links considerably outnumbering false links (688 versus 177 for CSLA linkage with broad region matching—see Table S.1). Poor region matching was the main reason for missing links to RAC admissions under the CSLA strategy. This was a less important, but still significant, reason for missing matches among RAC leave events. Because of the very limited geographic data available on the hospital and RAC datasets for use in E linkage, matches missed due to poor region matching on the two datasets cannot be retrieved by adjusting the E linkage strategy.

Missed matches to RAC leave events were primarily the result of poorly matching event dates for related hospital and RAC events. The large gaps between the recorded dates for the linked events in these missed matches (commonly 3 or more days) indicate that it would not be possible to adjust the CSLA matching strategy to allow capture of these matches without risking the introduction of large numbers of false matches.

False matches

The majority of false E matches is for admissions: for CSLA almost two-thirds of the small number of false matches were for admissions, with another third being for RAC hospital leave events and a small number relating to RAC social leave.

To ensure that CSLA false links were not the result of person links being missed by N linkage, the reliability of the person links on which N linkage was based was re-examined by reviewing HiL's person links related to the CSLA-only links. Investigations by HiL of the client links implied by the CSLA-only links led to HiL identifying just two additional person links between hospital and RAC clients that had previously been missed.

Many CSLA-only links have exact matches on both date of birth and event dates. This suggests these false matches are caused by similar hospital and/or RAC activity by similar people (in terms of date of birth and sex) living in a particular region. Comparisons indicate that one of the most effective ways of reducing the number of false matches made under the CSLA strategy would be to reduce the size of the geographic region used in matching. Analysis of CSLA-only links indicates that this would be more effective for links to RAC admissions than for links to RAC leave. However, overly narrowing the geographic matching criteria could result in dropping many more true links than false links (see below).

Quality of E linked datasets

There are five E linkage strategies that could be used in different situations—depending on the regional distribution of the population under study and the human resources available to undertake the data linkage:

- $CSLA_s$—constrained SLA group linkage strategy, not including matching within expanded regions defined by the first two digits of the postcode (termed 2-digit postcode matching)
- $CSLA_f$—constrained SLA group linkage strategy, including 2-digit postcode matching
- CPC_s—constrained postcode linkage strategy, not including 2-digit postcode matching
- CPC_f—constrained postcode linkage strategy, including 2-digit postcode matching
- BSESLA—basic SLA group linkage strategy.

These five strategies resulted in very similar total positive predictive values but a range of sensitivities (Table S.1) (Section 7): overall, both the refined constrained and basic E linkage strategies had PPVs of around 98%. Excluding matching on 2-digit postcode had very little effect on the PPV of the constrained strategies. However, disallowing such matches noticeably reduced the total number of matches (by 342 and 509 matches for CSLA and CPC, respectively) and consequently reduced the sensitivity of the strategies—from 91.5% down to 87.6% for CSLA, with a similar shift for CPC. Nonetheless, the sensitivity of the constrained strategies, both with and without 2-digit postcode matching, was considerably higher than that for BSESLA: the constrained E linkages resulted in over 5% more links than the basic strategy.

If all the E linkage strategies result in linked datasets that are equally representative of movements of people from hospital to residential aged care, the above results indicate that if 2-digit postcode matching can be included there is little to choose between the $CSLA_f$ and CPC_f strategies. In this case, since straight postcode matching is much easier both to understand and carry out than SLA group matching, the CPC_f strategy would be preferred over the $CSLA_f$ strategy. However, if 2-digit postcode matching cannot be included, the $CSLA_s$ strategy would be preferred because of its greater sensitivity while having almost the same PPV as CPC_s matching. Because of its relatively poor sensitivity, BSESLA matching would only be used if there were insufficient resources to undertake the more complex constrained matching. Analyses suggest that in this case basic matching using 3-digit postcode could provide similar results without introducing the complication of SLA group matching (see Section 5).

Table S.1: Positive predictive value and sensitivity of event-based E linkage strategies, using name-based N linkage as the reference standard

Match strategy		True links (A)	False links (B)	Missed links (C)	Total links (D = A+B)	PPV (A/D)	Sensitivity (A/F)	Relative size (D/F)
			Number				Per cent	
Name-based N linkage	**(F)**	**8,106**	. .	. .	. .	. .	. .	**100.0**
Event-based E linkage								
$CSLA_s$		7,100	153	1,006	7,253	97.9	87.6	89.5
$CSLA_f$		7,418	177	688	7,595	97.7	91.5	93.7
CPC_s		6,936	142	1,170	7,078	98.0	85.6	87.3
CPC_f		7,418	169	688	7,587	97.8	91.5	93.6
BSESLA		6,539	154	1,567	6,693	97.7	80.7	82.6

Source: Table 6.17.

The similar total PPVs but varying sensitivities of the E linkage strategies raises the question of whether differences in their linked datasets are in size only, as the consistent PPVs suggest, or whether there are some underlying distributional differences in the matches made under the various strategies.

Detailed analyses of the various sets of links show that the PPV and, in particular, the sensitivity of E links vary with RAC event type (Table S.2). Differences were also apparent for categories within a range of other variables. However, much of this variation can be explained by the mix of RAC event types within particular groups. When modelling the propensity of E linkage to miss N links within RAC event type, only a small number of variables were found to have statistically significant effects, with more variables being identified for the less sensitive basic linkage than for constrained linkage.

In practice, the dominant effect of RAC event type on the efficiency of E linkage is largely mitigated by the logical requirement of separate examination of different transition types, that is, movement into permanent admissions, respite admissions, RAC leave returns and the community. Furthermore, given that the E linkage strategies are much more likely to miss N links than make false matches, analysis indicates that E links provide a good basis for examining the demographic profile of people undertaking various types of transitions. There was some evidence, however, that there could be some small regional and/or hospital episode differences in the profile of N and E links for RAC admissions and RAC leave events. Such differences are likely to be greater for basic than for constrained E links.

Noting the above, the question then arises as to whether differences in the profiles of N and E links affect the utility of the E linked datasets for looking at movements between sectors.

Table S.2: Positive predictive value, sensitivity and relative size, by RAC event type and E linkage strategy, using name-based N linkage as the reference standard

	RAC event type				
Match strategy	**Permanent admission**	**Respite admission**	**Hospital leave**	**Social leave**	**Total**
Event-based E linkage strategy			**PPV (per cent)**		
$CSLA_s$	95.1	98.2	98.7	95.5	97.9
$CSLA_f$	94.8	97.4	98.7	95.5	97.7
CPC_s	95.0	98.0	98.7	99.3	98.0
CPC_f	95.0	97.2	98.7	99.3	97.8
BSESLA	95.5	97.8	98.5	89.2	97.7
			Sensitivity (per cent)		
$CSLA_s$	75.0	81.1	92.5	91.3	87.6
$CSLA_f$	88.5	86.4	93.3	91.3	91.5
CPC_s	69.5	78.5	91.7	90.7	85.6
CPC_f	88.9	87.0	93.1	90.7	91.5
BSESLA	65.2	73.8	86.4	91.9	80.7
			Relative size (per cent)		
$CSLA_s$	78.9	82.6	93.8	95.7	89.5
$CSLA_f$	93.3	88.7	94.6	95.7	93.7
CPC_s	73.2	80.2	92.9	91.3	87.3
CPC_f	93.6	89.4	94.3	91.3	93.6
BSESLA	68.3	75.5	87.7	103.1	82.6
Name-based N linkage (number)[(a)]	**1,723**	**852**	**5,370**	**161**	**8,106**

(a) Analysis indicated that for a small number of links the event match chosen by the N linkage strategy was not the preferred link. In particular, for 18 matches (0.2% of N links) the preferred link was to a RAC hospital leave event rather than the chosen (earlier) admission event for the same person.

Source: Table 7.3.

Utility of E linked data for analysis

When undertaking analysis of transitions, it is the combined effect of missed N links and false E links that determines the overall utility of an E-linked dataset. Examination of all records linked under the particular strategies shows that, overall, the E linkage strategies resulted in linked data that largely reflected the N linkage match set in terms of the distributions across key variables (Section 8). That is, while not exactly the same, the E linkage match sets' distributions looked highly similar to those for the N linkage match set. In this respect, constrained methods using SLA group when matching performed better than those using straight postcode, and constrained matching (even without using 2-digit postcode matching) performed at least as well as or better than basic E matching.

Many analyses of movement between hospital and RAC will want to compare people who have moved between the two sectors with those who have not. In this case, both linked and unlinked records are examined. In this study, examples of such analyses considered three broad groups of analysis: movement from hospital, movement into RAC, and an example looking at a specific issue—dementia. In all cases, analysis was carried out taking into account the type of transition into RAC, thereby removing one of the greatest sources of possible bias identified in the earlier analyses. Comparisons were limited to the three E linkage strategies $CSLA_s$, CPC_s and BSESLA. The 2-digit postcode E linkage strategies were not considered as they could not be used for national analysis due to the high population density in some regions in the larger states.

In terms of practical utility, analysis by post-hospital destination or source of RAC admission indicated that, as expected from the sensitivity estimates, the E linkage strategies underestimate the volume of movement between hospital and RAC, with permanent RAC admissions being particularly affected (see relative size in Table S.2). Nevertheless, illustrative examples looking at patterns of use and characteristics of people moving between the two sectors—such as those in Tables S.3 and S.4—show that analyses using links derived from the N and E strategies lead to very similar conclusions. Examination of results also indicate that, irrespective of the linkage strategy, care needs to be taken when drawing conclusions as some differences may not be statistically different due to small numbers in some cross-classifications.

Table S.3: Analysis example 1: Median length of hospital episode, by transition type and sex, for name-based N linkage and event-based $CSLA_s$ linkage, separations for people aged 65 years and over, 2000–01

	Name-based N linkage			Event-based $CSLA_s$ E linkage		
Movement type	Males	Females	All	Males	Females	All
	Median length of hospital episode (days)					
Returning to permanent RAC[(a)]	6	6	6	6	6	6
Hospital to permanent RAC[(a)]	33	31	32	32	29	30
Hospital to respite RAC[(a)]	16	14.5	15	16	15	15
Hospital to community/other[(b)]	4	4	4	4	4	4
Died in hospital[(b)]	8	8	8	8	8	8

(a) Based on linked hospital and RAC records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

Note: Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals. Length of stay excludes days on leave from hospital.

Source: Table 8.6.

Table S.4: Analysis example 2: Care level and dementia status for RAC entries, by transition type, for name-based N linkage and event-based CSLA$_s$ linkage, 2000–01

	Name-based N linkage					Event-based CSLA$_s$ E linkage				
	Care level					Care level				
Transition type	High	Low	All	Total		High	Low	All	Total	
	Row %			Col. %	N	Row %			Col. %	N
Return from hospital[(a)]										
With dementia	64.5	35.5	100.0	31.0	1,711	64.9	35.1	100.0	30.7	1,583
Without dementia	33.6	66.4	100.0	69.0	3,802	33.3	66.7	100.0	69.3	3,575
All	*43.2*	*56.8*	*100.0*	*100.0*	*5,513*	*43.0*	*57.0*	*100.0*	*100.0*	*5,158*
Into permanent RAC from hospital[(a)]										
With dementia	83.1	16.9	100.0	45.9	769	83.9	16.1	100.0	45.9	598
Without dementia	72.6	27.4	100.0	54.1	905	74.5	25.5	100.0	54.1	705
All	*77.4*	*22.6*	*100.0*	*100.0*	*1,674*	*78.8*	*21.2*	*100.0*	*100.0*	*1,303*
Into respite RAC from hospital[(a)]										
With dementia	32.5	67.5	100.0	24.6	209	32.4	67.6	100.0	24.7	173
Without dementia	19.1	80.9	100.0	75.4	640	19.6	80.4	100.0	75.3	526
All	*22.4*	*77.6*	*100.0*	*100.0*	*849*	*22.7*	*77.3*	*100.0*	*100.0*	*699*

(a) Based on linked hospital and RAC records. See also notes (a) and (b) to Table 8.4 for additional information.

Note: Diagnosis of dementia includes diagnoses of dementia and Alzheimer's disease (ICD-10-AM Ed. 1 categories F00–F03, and G30—see NCCH 1998). Table excludes 115 cases with missing RCS, and all unlinked RAC hospital leave events.

Source: Table 8.17.

1 Background

The interface between acute hospital care and residential aged care (RAC) has long been recognised as an important issue in aged care services research. Despite this, existing national data provide very poor information on the movement of clients between the residential and acute care sectors. Current national datasets on the two sectors are derived from administrative by-product collections, and have been designed primarily to provide data on the specific program which they describe, rather than to examine program interfaces or provide system-level information.

Data linkage is a statistical tool that can be used to link data from different sources, thereby expanding the types of statistical investigations that can be carried out (including analysis of movement over time) without increasing the reporting load of service providers or requiring special surveys. This suggests that data linkage could be used to develop datasets suitable for investigating movement between the hospital and aged care sectors. Data linkage of records for individuals is commonly carried out using detailed demographic data, including name and/or a person identification number. However, while the national RAC data and some of the state and territory hospital morbidity datasets may contain name, the national hospital morbidity dataset does not contain such information. Furthermore, the national hospital and RAC datasets do not contain a common person identification number which could facilitate data linkage.

While neither name nor a common person identification number are available for linking data from the two sectors, some demographic data are available. In addition, information on transition dates, that is, entry and exit dates, is available for all periods of hospitalisation and residential aged care. To see if such event data are sufficient to allow data linkage, an investigation into the feasibility of linking hospital morbidity and RAC datasets using a combination of demographic and event data was conducted. The study was carried out in 2001 and 2002 by the Australian Institute of Health and Welfare (AIHW) under the auspices of the Australian Health Ministers' Advisory Council's Care of Older Australians Working Group. Findings from the feasibility study suggested that the set of linked client records resulting from an event-based anonymous linkage strategy could provide a valuable source of information on the client characteristics and service use patterns associated with movements between the two sectors (AIHW 2003).

In the feasibility study, matching was based on date of birth, sex, region of usual residence, and hospital separation and RAC entry dates. In the absence of validation against a gold standard linkage, doubts concerning the efficacy of the strategy were raised because of the lack of either name or a common person identifier on the two datasets. This issue has been addressed using two distinct methods.

First, in 2004 a study employing statistical theory investigated the effectiveness of the AIHW event-based strategy in a range of linkage situations, and, using the results, refined the strategy for use in future work (AIHW: Karmel 2004). Overall, estimates based on the theoretical analysis indicated that the Institute's linkage strategy results in acceptably low rates of false matches, and so can be used to derive a dataset useful for investigating the hospital–aged care interface.

Second, direct comparisons between a name-based linkage strategy and the event-based linkage strategy have been carried out to measure any differences between the two approaches, and to determine whether event-based matching leads to any biases in the match dataset compared with name-based matching. It is these comparative analyses that are the subject of this report. The scope of the comparisons was limited to movements from hospital to RAC in Western Australia during the financial year 2000–01. Western Australia was chosen for this study because of its unique position in having a well-established data linkage system which was recently expanded to include RAC data (see Holman et al. 1999; Brook et al. 2005).

In summary, this report compares results obtained using name-based and event-based matching strategies, for ease of reference termed N and E linkage respectively. In particular, the report:

- examines ways to refine the original E linkage strategy
- examines missed and false links made by E linkage (when compared with N linkage)
- compares the E and N links as a whole, looking for similarities and differences in distribution across variables of interest
- compares results from analyses using E and N linked data to identify the types of analyses for which the results are comparable and those for which there are differences that need to be taken into account.

Before data linkage was undertaken for this project, ethical approval for the project was obtained from the AIHW Ethics Committee, and permission to use the relevant data was obtained from the Steering Committee for Cross-Jurisdictional Linked Administrative Health Data (a joint Australian and Western Australian government committee) and the Confidentiality of Health Information Committee of the Western Australian Department of Health.

1.1 Report structure

In covering the various phases of the comparison, the report is structured as follows. First, the context of movements from hospital into RAC is outlined (Section 2), and the data used to link the datasets and to allow analysis of movements are described (Section 3). The N and E linkage strategies are then described in detail in Sections 4 and 5, respectively. As there may be data and resource limitations when undertaking data linkage, two types of E linkage are discussed.

Section 6 contains comparisons of results from the N linkage strategy with initial results from the E linkage strategies. Refinements to the E linkage strategies to improve their efficiency are identified. Reasons for the event-based strategy missing some of the links identified using the name-based approach are examined, as are causes of additional (or false) links. In addition, the effect of population size on the accuracy of the E linkage is considered. Five variants of the E linkage strategy are put forward for detailed investigation.

The completeness and reliability of the E-linked datasets, when compared with that resulting from the N linkage strategy, are examined in Section 7, while Section 8 illustrates the analytical effect of any differences in the linked datasets using a number of illustrative analyses. Finally, a number of appendices contain further details of the methods used and comparisons undertaken during this project.

2 Types of transitions

When identifying transition events, detailed knowledge of both the service systems and the data collection practices within those systems is essential, irrespective of the data linkage method used. Such awareness is required to ensure that the most appropriate event is selected as the transition event. In linking hospital and RAC data to investigate movement from hospital into RAC, the aim of any linkage strategy is to match hospital separations by people who then go into RAC.

In general, an episode in hospital may end with the patient either:

a returning home (in the community) or going to live temporarily with family and friends

b going to live temporarily with family and friends before returning to a RAC service

c going into a RAC service

d returning to the RAC service where they usually live

e transferring to another hospital

f transferring to residential health care services

g changing episode type within the hospital

h dying.

Similarly, a person may enter residential aged care in a number of ways. An entry may be for a person:

A being admitted into RAC for permanent care

B being admitted into RAC for respite care

C transferring between RAC services (for either permanent or respite care)

D returning to RAC following an episode in hospital (termed 'hospital leave') or

E returning to RAC following a stay with family or friends (termed 'social leave').

F In addition, a RAC permanent resident may go to hospital and die while there; that is, die while on hospital leave or die in hospital while on social leave.

In terms of the events described above, linkage strategies try to match hospital separations of type *c* to the relevant RAC entries of types *A* and *B*; separations of type *d* to entries of type *D* (or occasionally *E*); separations of type *b* to the relevant entries of type *E*; and separations of type *h* to deaths of aged care residents while in hospital (type *F*) (AIHW: Karmel 2004:1).

3 Data

In the current analysis, we are comparing two distinct linkage strategies. The first is a name-based process carried out by the Health Information Linkage Branch (HiL) of the Western Australian Department of Health, that is, N linkage. The second is the AIHW event-based linkage strategy—E linkage—which matches event records using date of birth, sex, geographic region of usual residence and event dates.

Data for hospital episodes were linked to RAC events, for 2000–01, using the N and E linkage strategies. Both datasets were limited to events within Western Australia. This excluded the possibility of identifying interstate movements; however, the extent of such moves is likely to be small for Western Australia. In addition, the data were limited to events for people who were aged at least 65 years by 30 June 2001.

Two extracts were obtained for both hospital and RAC data: one to be used to establish links (a linkage dataset), and a second to be used for analysing link bias (an analysis dataset).

3.1 Hospital data

Hospital event data consistent with the National Hospital Morbidity Database (NHMD) was prepared by HiL. This restriction was applied so that the results would indicate outcomes if the strategy were to be applied to link NHMD and RAC data in other jurisdictions. The data included both public and private hospital separations.

For the E linkage strategy (carried out at the AIHW) a special extract of hospital separations was derived. The hospital linkage dataset contained only the information required for establishing and checking E links, and included date of birth, sex, postcode of usual residence, country of birth, marital status, admission and separation dates, and modes of hospital admission and separation.

Two sets of records were excluded from the hospital dataset used for E linkage:

- Statistical discharges: a 'statistical discharge' means that the person in question changes from one hospital episode care type to another (for example, from acute care to rehabilitation) or transfers from one hospital to another. As these people do not leave the hospital system, trying to link to a RAC admission is not appropriate. In addition, if the person were on RAC hospital leave, statistical discharges would not relate to a return to RAC. Consequently, separation records relating to statistical discharges were excluded.
- Same-day hospital separations: people admitted and discharged on the same day are unlikely to be discharged to a RAC facility unless they are going from a RAC facility to hospital for a day procedure. In this case the person is unlikely to be recorded as being on hospital leave by the RAC facility. Therefore, to avoid spurious matches between aged care admissions and hospital separations into the community, records with the same admission and separation dates were excluded.

Statistical discharges were included in the N linkage for this project as the Western Australian data linkage system allows the derivation of complete hospital stay, and this would aid linkage to RAC hospital leave. On the other hand, same-day separations were excluded from the N linkage, both for consistency with the E linkage approach and because people entering and leaving hospital on the same day are unlikely to be newly admitted to a RAC facility on discharge, and reported RAC leave events involve at least one night.

In 2000–01, in Western Australia there were just over 86,200 hospital separations with a non-statistical discharge that lasted one or more nights (Table 3.1). Note that to allow for small gaps between hospital separation and RAC admission (and for small differences in recording dates), separations for a few days either end of the financial year were also included in the dataset for linkage.

To allow detailed comparison of the linked datasets resulting from the N and E strategies, a dataset specifically for analysis was also derived. This analysis dataset contained the information to be used when investigating possible analytical bias, and included age, sex, region of usual residence, country of birth, marital status, modes of hospital admission and separation, length of stay, and hospital diagnosis and procedure variables.

Table 3.1: Hospital separations, by mode of admission and separation, 2000–01

	Mode of admission		
Mode of separation	**Statistical**	**Non-statistical**	**Total**
To RAC	862	1,370	2,232
To other health care establishment	422	774	1,196
Left against medical advice/ statistical discharge from leave/unknown	161	338	499
Death	909	3,057	3,966
To usual residence/other	6,896	71,432	78,328
Total	**9,250**	**76,971**	**86,221**

Note: Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals.

Source: AIHW analysis of hospital morbidity linkage dataset.

3.2 Residential aged care data

The RAC data included all RAC permanent and respite admissions and hospital and social leave events for the year of interest, totalling slightly more than 19,600 events (Table 3.2). The data were derived from the Department of Health and Ageing's Aged and Community Care Management Information System (ACCMIS).

Table 3.2: RAC events, by event type, 2000–01

RAC event type	Number
Permanent admission	4,752
Respite admission	3,297
Hospital leave	6,878
Social leave	4,709
Total	**19,636**

Note: Table includes admissions within the year, and RAC leave events (including deaths in hospital and discharges to hospital) with at least some days in the period 1 July 2000–30 June 2001.

Source: AIHW analysis of ACCMIS.

As for the hospital data, two extracts were obtained: one for linkage and one for analysis. While the N linkage process had already established person links between hospital and RAC services, the appropriate event links still had to be derived. Therefore, both the N and E linkage used the same RAC event linkage file to establish final event links.

The RAC linkage file, containing data for establishing and checking event links, included a project-specific RAC client number, date of birth, sex, postcode of usual residence (for leave events this is the RAC facility's postcode), country of birth, marital status, event type, admission and discharge dates, leave event start and end dates, place and date of Aged Care Assessment Team (ACAT) assessment, and mode of discharge. The RAC analysis file, with information to be used when investigating possible analytical bias, contained the project-specific RAC client number, age, sex, region of residence before current event, region of RAC facility, country of birth, marital status, event dates, care needs assessment, place, level and date of ACAT assessment, and mode of discharge.

4 N linkage strategy

The purpose of the data matching was to link a hospital episode (where relevant) to a single RAC entry event for the same person, that is, a 1-to-1 match between a hospital (non-statistical) separation and RAC entry event. While some constraints were imposed on HiL by the RAC data provider when linking RAC and health data, the use of name and address in N linkage, and the availability of name and address reporting history across a range of health service events, results in this linkage being highly reliable. Consequently, in this report it is used as the reference standard against which the E linkage results are compared. Linkage checks indicate a very low error rate in the established Western Australian health system – RAC system person links.

4.1 Linkage protocol

A memorandum of understanding signed by the Western Australian Department of Health and the Australian Government Department of Health and Ageing in December 2002 specified an agreed protocol to be used for linkages between Western Australian and Australian government health datasets, including those for RAC records. This protocol is referred to as the 'Bass Protocol', the 'WA Best Practice Protocol' or the 'Two Stage Protocol' (Kelman et al. 2002).

The first essential element of this protocol is that the linkage stage must be performed separately from the analysis stage. The second essential element is that full identifying information is available for the linkage stage, but is not available for the analysis stage. Similarly the information available for linkage does not include clinical or service details, including the dates of service. The third element is that the people involved in the linkage stage (linkage staff) do not have access to the clinical or service data and therefore cannot take part in the analysis stage. This protocol requires a dedicated linkage team with the ability to perform linkages that span a variety of data sources and time periods. The result is improved link quality and efficiency across all data sources.

4.2 Linkage process

The name-based linkage was undertaken using the Western Australian Data Linkage System (WADLS) which is a dynamic on-line system able to be accessed continuously by the team of linkage officers within HiL. The linkage process uses as much personal information as is available to create and load links as they are discovered. Any new links are loaded into the system as they are identified. The system contains links between population-based Western Australian health datasets dating from as early as 1966. Data sources contributing to this system include electoral records, hospital admissions, emergency presentations, cancer registrations, mental health contacts, midwives' notifications as well as birth, death and marriage registrations.

For the RAC linkage, the demographic information used for linkage included surname, given names, sex, date of birth and address. Separate linkages between RAC and Western Australian electoral, ambulance, hospital, emergency and death records were performed. For each linkage, there were up to eight passes through the records using different blocking and match strategies. A de-duplication linkage run was also performed on RAC records alone in order to detect groups of RAC records thought to belong to the same person, resulting in identifying 1,320 people with more than one record among the 57,145 RAC person records from ACCMIS for 1997–2005. Each link created between a RAC client record and any of the other Western Australian health records was loaded into the special tables within the WADLS. These separate 'extra links' tables were established for linkages such as the cross-jurisdictional arrangements so that access could be managed separately but links could still be updated if found at a later stage.

The result of the linkage process was a file of 'chains' of linked hospital and RAC records for each person aged 65 years or over who had either been in hospital or a RAC facility during the 12-month period of interest. However, because of the demographic history available through the WADLS, many of these links may have been created by using information other than that contained within particular RAC or hospital admission records.

From the hospital–RAC person links, hospital events relating to a particular RAC event had to be selected for each person. The most appropriate hospital–RAC event link was chosen by measuring the proximity of hospital and RAC event dates. For RAC leave events, both the start and end dates of the hospital stay and RAC leave were considered, with the selected match being that with the smallest difference between the hospital separation date and RAC re-entry date. Considerable differences in the end dates were allowed because people may have been discharged from RAC while they were in hospital (discharge to hospital). In addition, deaths in hospital were identified to establish which RAC hospital leave episodes could not involve a return to RAC. For RAC admissions (permanent and respite), only the hospital separation and RAC admission dates were considered. To allow for small delays in entering RAC, as well as reporting discrepancies, these dates could differ by up to 2 days and still be considered a match.

There are situations where a RAC resident can have two RAC entry events on the same day. In particular, a person may finish a period of leave and move to a different RAC facility on the same day so that they would have both a leave end date and RAC admission date on the same day. Since, logically, a person must leave one facility before entering another, the match to the hospital leave is preferred. Also, choosing the hospital leave event indicates that the person was already in the RAC system, and not a new entrant. To put this preference into practice, where there was a choice between matches to different RAC event types, the match to the event with the earlier start and end dates was chosen. For cases where a person ended a period of RAC hospital leave and changed RAC facility on the same day, this choice by event date resulted in the RAC hospital leave event being the selected match, as the exit date for the RAC leave was earlier than the exit date for the RAC admission.

A detailed description of the N linkage match specifications is given in Appendix 1. Overall, the N strategy resulted in 8,106 links between hospital separations and entry, or re-entry, into RAC (Table 4.1). In addition, just over 440 links were established with RAC hospital leave that ended with the death of the patient. A further 575 links matched to RAC events that involved a discharge to hospital while on RAC hospital leave.

Table 4.1: Summary of N linkage, matched events for people moving from hospital to RAC, 2000–01 (number)

Match type		[(a)]Matches to deaths in hospital	Other matches	Total
RAC hospital leave				
A.	Exact date match of RAC leave and hospital stay, without hospital transfers	299	3,826	4,125
B.	Exact date match of RAC leave and hospital stay, with hospital transfers	46	425	471
C.	Exact end date match of RAC leave and hospital stay	43	625	668
D.	RAC leave encompasses hospital stay	47	189	236
E.	Hospital stay encompasses RAC leave	3	250	253
F.	RAC leave starts before hospital stay, and ends before hospital stay	1	30	31
G.	RAC leave starts after hospital stay, and ends after hospital stay	3	25	28
H.	Discharge to hospital (excluded from linkage)	. .	. .	[(b)]575
Total, excluding H		*442*	*5,370*	*5,812*
RAC social leave				
I.	RAC leave encompasses hospital stay	20	161	181
Admissions				
J.	Permanent RAC admissions	2	1,723	1,725
K.	Respite RAC admissions	—	852	852
Total		*2*	*2,575*	*2,577*
Total (excluding H)		**464**	**8,106**	**8,570**

(a) Death in hospital identified from date of death and hospital date of separation.

(b) May include links to hospital episodes not available to the AIHW for E linkage (due to date of extraction). 118 matches were dropped for this reason from match types A–G and I–K.

Notes

1. Coverage includes links involving either a hospital separation ending in the financial year 2000–01, or RAC admission within the year; return from hospital leave within the year, or on RAC social leave at some time in the year (may either start and/or end in the financial year).
2. Duplicate matches have been deleted using priority ranking (see Appendix 1).
3. In 20 cases, a permanent admission (out of 1,725) was selected rather than the relevant RAC hospital leave event. These cases (including two deaths in hospital) were for people who started a period of hospital leave on the same day as, or very soon after, admission into permanent RAC.

5 E linkage strategies

5.1 Linkage protocol

To protect the privacy of individuals, the E linkage was carried out by the AIHW using the Institute's protocol *Data linkage and protecting privacy: a protocol for linking between two or more data sets held within the Australian Institute of Health and Welfare* (AIHW 2006). The principles underlying this protocol are that:

- Data linkage is not carried out directly between original complete data sets;
- Data linkage is undertaken using purpose-specific linkage data sets that contain only the data required for establishing and validating links;
- Links between data sets are recorded using project specific unique record identifiers so that links identified for a particular project (including longitudinal analyses) cannot be used to establish links between data sets outside the scope of the project using a chain of links ('consequential' linking);
- Analysis files do not contain identifying data (such as name, date of birth and address, or the record number from the original data set); and
- Intermediate data sets and the project specific record identifiers are deleted following completion of the final linked analysis data sets.

5.2 Linkage process: constrained E matching

The purpose of constrained E matching is to find the best match using all available event date information and event descriptors available in the NHMD and ACCMIS data. To achieve this, matching procedures are specified separately for comparisons between different subsets of RAC and hospital events defined in terms of their type and/or admission and separation characteristics. Because two dates are available for RAC hospital leave (start and end dates), and the related hospital episode may end in a number of ways, match procedures for these events are the most complicated.

Event information may suggest that some matches are more likely to be correct than others (for example, a link of *hospital discharge reported as to RAC*:*RAC admission* has greater face validity than a link of *hospital discharge reported as to usual residence*:*RAC admission*). In addition, matching within subsets defined in terms of event characteristics both minimises coincident records within datasets (with respect to the match variables) by reducing the number of records being compared, and allows link priorities to be set later in the matching process if non-unique matches are identified. That is, matching within partitioned datasets reduces the incidence of duplicate links and facilitates selection of the most likely match if duplicate links occur when the links from the partitioned datasets are combined. For this project, 12 such partitioned dataset pairs were used (Table 5.1). The range of event date matches that can logically occur for these pairs, assuming accurate reporting of dates, is illustrated schematically in Appendix 2.

Constrained E matching was carried out in two stages. Initial matches for each partitioned dataset pair were selected using 1:1 probability matching via the matching software *Websphere®* (previously known as *Automatch®* and *Integrity®*). In this stage, relatively broad match criteria were used to identify possible matches between RAC and hospital partitioned datasets. While the matching was probabilistic, at least partial matches were required on each of date of birth, sex, region and event dates. Some variation was allowed in exactness of match, particularly for event dates where allowable date differences between the RAC and hospital events were specified according to the types of events being matched (see Table 5.1). In addition, some variation in date of birth (matching on two out of day, month and year of birth) and region was considered.

Table 5.1: Dataset partitioning and event date match rules for E linkage constrained match selection

<table>
<tr><th>Partition code</th><th>Link priority</th><th>Hospital mode of separation</th><th>RAC type</th><th>RAC exit:hospital entry date matching</th><th>Hospital exit:RAC entry date matching</th></tr>
<tr><td>8ADM</td><td>11</td><td>To death, using person region[(a)]</td><td>Admissions</td><td rowspan="3">. .</td><td rowspan="5">|RAC entry–Hospital exit| ≤2 days
Or
|RAC pre-entry–Hospital exit| ≤2 days</td></tr>
<tr><td>9ADM</td><td>8</td><td>To usual residence, using person region</td><td>Admissions</td></tr>
<tr><td>0ADM</td><td>7</td><td>To other, using person region</td><td>Admissions</td></tr>
<tr><td>H9ADM</td><td>10</td><td>To usual residence, using hospital region</td><td>Admissions with ACAT in hospital</td><td rowspan="2">{also Hospital entry ≤ACAT date ≤ Hospital exit}</td></tr>
<tr><td>H0ADM</td><td>9</td><td>To other, using hospital region</td><td>Admissions with ACAT in hospital</td></tr>
<tr><td>9SOC</td><td>12</td><td>All</td><td>Social leave</td><td>RAC exit ≤Hospital entry</td><td>Hospital exit ≤RAC entry</td></tr>
<tr><td>NST8H</td><td>2</td><td>With non-statistical admission, to death</td><td>Hospital leave</td><td rowspan="3">|Hospital entry–RAC exit| = 0, 1, 2 days</td><td rowspan="3">[(b)](RAC entry–Hospital exit) ≤2 days</td></tr>
<tr><td>NST9H</td><td>1</td><td>With non-statistical admission, to usual residence</td><td>Hospital leave</td></tr>
<tr><td>NST0H</td><td>3</td><td>With non-statistical admission, to other</td><td>Hospital leave</td></tr>
<tr><td>ST8H</td><td>5</td><td>With statistical admission, to death</td><td>Hospital leave</td><td rowspan="3">(Hospital entry–RAC exit) ≥–2</td><td>[(b)](RAC entry–Hospital exit) ≤2 days</td></tr>
<tr><td>ST9H</td><td>4</td><td>With statistical admission, to usual residence</td><td>Hospital leave</td><td rowspan="2">[(c)](RAC entry–Hospital exit) = 0, 1, 2 days</td></tr>
<tr><td>ST0H</td><td>6</td><td>With statistical admission, to other</td><td>Hospital leave</td></tr>
</table>

(a) Region used throughout the matching depends on the particular strategy.

(b) Allows for discharge to hospital (negative difference).

(c) Stricter rules on end-date because it is difficult to allow consistently for death in hospital for a statistical admission, and to compensate for the poorer match event start date.

Note: See Appendix 2 for greater detail on the dataset partitioning and for diagrams illustrating possible event matches. Table A2.1 describes the categories available for hospital mode of separation.

Probabilistic matching generally involves two types of data items: blocking and match variables, with match variables being compared only within categories of blocking variables thereby greatly reducing the number of pair-wise comparisons needed to identify matches. Matches based on data of varying accuracy are identified by changing the blocking variables and adjusting the match variables appropriately. The extent of variation allowed in the *Websphere*® matching process for this project is shown in Table 5.2 which summarises the general approach taken to blocking and match variable specifications, with the particular specifications depending on the two datasets being compared (as per Table 5.1).

In the current project two types of constrained E linkage were considered. The first (CSLA) used statistical local area (SLA) group to identify region of usual residence, where an SLA group is that set of SLAs that overlap a postcode and two postcodes are said to match if they have a common SLA in their SLA groups; for example, in Figure 5.1 postcodes 1 and 2 match on SLA group, as both intersect with SLAs 2 and 4, but postcodes 1 and 3 do not. The second constrained matching strategy (CPC) used straight postcode when matching on region. Table 5.2 contains specifications for both the CSLA and CPC approaches.

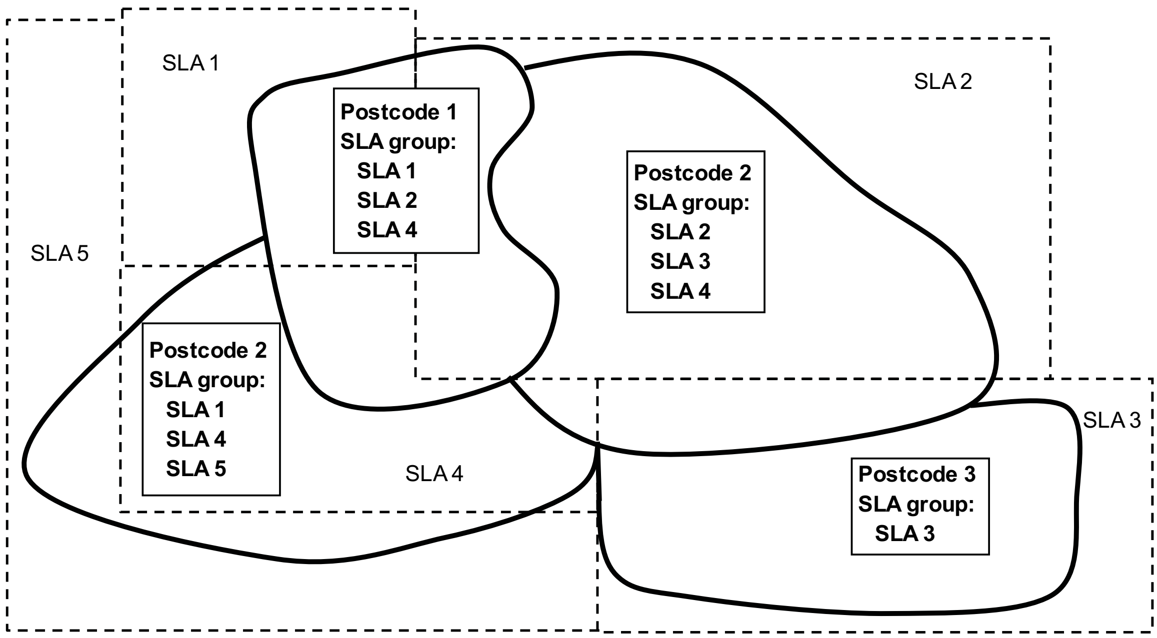


Figure 5.1: Illustrating the SLA group of a postcode

Table 5.2: Blocking and matching variables used in *Websphere*® passes for two constrained E linkage strategies

	Successive *Websphere*® matching passes						
	1	2	3	4	5	6	7
	Exact match	1-sided event date(s) variation	Year of birth variation	Month *or* day of date of birth variation	2-sided event date variation	Exact match using first 3 postcode digits	Exact match using first 2 postcode digits
Strategy	**Blocking variables**						
CSLA	SLA sex DOB [date][(a)]	SLA sex DOB	SLA sex day of birth month of birth [date][(a)]	SLA sex year of birth [date][(a)]	SLA sex DOB	3-digit postcode sex DOB [date][(a)]	2-digit postcode sex DOB [date][(a)]
CPC	postcode sex DOB [date][(a)]	3-digit postcode sex DOB	3-digit postcode sex day of birth month of birth [date][(a)]	3-digit postcode sex year of birth [date][(a)]	3-digit postcode sex DOB	. .	2-digit postcode sex DOB [date][(a)]

(a) Dates are only used as blocking variables if an exact date match is appropriate for the particular partitioned dataset pair being linked. This depends on whether RAC leave events or admissions are being compared, and whether a death/discharge to hospital is involved (see Table 5.1 and diagrams in Appendix 2).

Note: 3-digit postcode indicates that the first three digits of the postcode were used for region matching; 2-digit postcode indicates that the first two digits of the postcode were used for region matching.

In the second stage of match selection, the results from matching within the 12 partitioned dataset pairs were checked for compliance with the rules for acceptable variation (as specified in Table 5.1), and possible discharges to hospital and deaths in hospital while on hospital leave were identified. The 12 datasets were then combined.

Combining the linked sets may result in many-to-many matches. This is because there is overlap across the partitioned dataset pairs; for example, all RAC admissions are compared with the three datasets containing hospital separations to death, separations to usual residence/other and other separations. These duplicate links are reduced to a 1:1 match using priority ratings that rank matches based on RAC event type, reliability of region information and hospital separation mode. Overall, links to RAC hospital leave are given top priority, then RAC admissions and finally RAC social leave. Note that RAC social leave is given low priority under E linkage because of the relatively soft event date data available for these matches, and because people who go on social leave are likely to be the more robust RAC residents and so less likely to use hospital. Detailed match specifications and priorities for constrained E linkage are given in Appendix 1.

5.3 Linkage process: basic E matching

Basic E matching, as used in the initial feasibility study (AIHW 2003), involves linking using date of birth, sex and region of residence (as in the constrained strategy) but only the possible transfer date, that is, the hospital exit /RAC entry date. Consequently, it treats all events the same way when matching. As the matching uses quite a limited range of data, it may not always provide optimal links when a person (or persons with the same demographic data) has two RAC events on the same day (for example, a person may have a return from leave and a RAC admission on the same day).

In this study, basic E linkage essentially involved deterministic matching using sex, date of birth, transfer date, and region of usual residence. Only small differences in the match dates were allowed, with the RAC entry required to be at most 2 days after the hospital exit for it to be considered a possible match. If there were multiple possible links for a particular event, the 'best' match was that with the smallest difference in event dates. The strategy did not allow matches where the RAC (re-)entry was before hospital separation. As in the feasibility study, the region of usual residence used in the current project was the SLA group of a postcode.

In summary, for this project three sets of matches were established using the event-based strategies:

- constrained matching within SLA group (CSLA)
- constrained matching within postcode (CPC)
- basic matching within SLA group (BSESLA).

Comparisons of the three resulting match sets with the N matches are discussed below. Initial comparisons enabled the E linkage constrained linkage strategies to be refined to improve their accuracy. The methods for identifying possible refinements and comparisons with the resulting match sets are also presented.

6 Initial comparisons and refining the constrained E linkage strategies

6.1 Methods

When linking records four outcomes are possible: a true link, a true non-link, a false link (false positive) and a missed link (false negative). In the current analysis, the N linkage provided the reference standard, and so the status of E links (that is, whether a link was a true link, a true non-link, a false link or a missed link) was determined by comparing the E links with the N links (Figure 6.1).

	N links	**N non-links**
Linked by E linkage (E links)	E true links	E false links
Not linked by E linkage (E non-links)	E missed links	E true non-links

Figure 6.1: Classification of E links when compared with N links

Four key measures are commonly used when comparing matches (Medical University of South Carolina 2001). Using terminology originating in epidemiology and clinical trials, these are:

- Positive predictive value (PPV): the percentage of E links that are true links
  = E true links/(E true links + E false links)
  = E true links/E links
- Sensitivity: the percentage of all links that are identified by the E linkage strategy
  = E true links/(E true links + E missed links)
  = E true links/N links
- Specificity: the percentage of all non-links that are not linked by the E linkage strategy
  = E true non-links/(E true non-links + E false links)
  = E true non-links/N non-links
- Negative predictive value: the percentage of E non-links that are true non-links
  = E true non-links/(E true non-links + E missed links)
  = E true non-links/E non-links.

For this project, the number of non-links depends on whether hospital separations or RAC entry events are used as the baseline. Accordingly, both the specificity and negative predictive value will change depending on the perspective taken (that is, whether we look 'from hospital', or 'into residential care'). Consequently, the linkage comparisons presented in this report concentrate on the PPV and sensitivity of the E linkage strategies.

When comparing linkage strategies we can easily identify whether a record has been linked and whether it should have been linked. However, in the current context it is also important whether an event record in one dataset matches with the correct event record in the other dataset. That is, as well as knowing whether or not an event should have been linked, we need to know when two strategies result in an event in one dataset being linked to the same or different events in the second dataset. Overall, there are six types of possible link concordances, and these are illustrated in Figure 6.2. In the following analysis, situations where matches of types D, E and F in Figure 6.2 should be considered true links or false links are discussed.

Direct comparison of the N and E links presents the possibility of refining the E linkage constrained strategy to improve its PPV without greatly affecting sensitivity. An approach for identifying such refinements is presented below. These refinements are not project specific, being based on including or excluding particular matching strategies (that is, *Websphere*® passes), and so can be applied to linkage projects where comparisons with a reference standard are not possible.

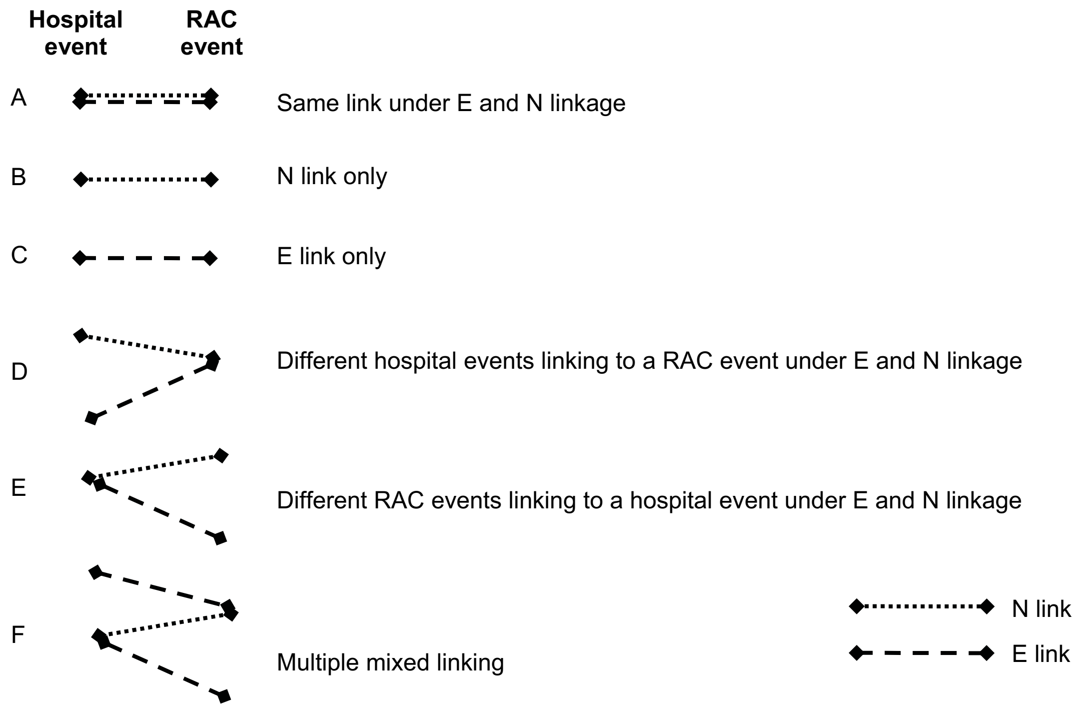


Figure 6.2: Types of link concordances between N and E links

6.2 Constrained matching within SLA group (CSLA)

Links were first derived using constrained matching with SLA group as the measure of region. Preliminary analysis indicated that matches that allow differences in both date of birth and event date often result in erroneous matches when compared with N matches (see Table A3.1.). Therefore, such variations were not considered; however, matching that allowed for variation in the match region was investigated (see Table 6.3).

Using SLA group as the primary measure of region resulted in 7,802 matches (called CSLA matches), excluding 136 links identified as a RAC discharge to hospital and a further 448 identified as deaths in hospital. Of the deaths in hospital, 429 links were for RAC hospital leave and 19 were for RAC social leave events. For N linkage, the corresponding figures were: 575 links to discharges to hospital (includes links to hospital episodes with a statistical separation not included in the E linkage strategy for movements from hospital to RAC), 464 links for deaths in hospital, and 8,106 links for other events (see Table 4.1). Detailed comparisons of the two sets of links involving movement from hospital to RAC are discussed below and presented in Tables 6.1 and 6.3.

Identifying true matches

From Table 6.1 it can be seen that all link types illustrated in Figure 6.2 occur; that is, identical links, N-only links, CSLA-only links and mixed links (types 31 and 32 in Table 6.1). Of the latter, there were 47 links between events for different people under N and CSLA linkage, and 75 between different events for the same person (including one set of three links where both these outcomes happened—compare link type F in Figure 6.2). The resulting 121 links related to 87 hospital separations and 93 RAC events.

Table 6.1: Comparing N and CSLA links: distinct link pairs

	Match to RAC event					
Match to hospital event	**1. RAC event links to same hospital event**	**2. RAC event links under N linkage only**	**31. RAC event links to same person, but diff. hospital event**	**32. RAC event links to hospital events for diff. people**	**4. RAC event links under CSLA only**	**Total**
1. Hospital event links to same RAC event	7,431	. .	. .	. .	. .	7,431
2. Hospital event links under N linkage only	. .	615	8	18	. .	641
31. Hospital event links to same person, but different RAC event	. .	27	1	1	29	58
32. Hospital event links to RAC events for different people	. .	5	. .	1	4	10
4. Hospital event links under CSLA only	. .	. .	9	18	310	337
Total	**7,431**	**647**	**18**	**38**	**343**	**8,477**

Note: The table counts distinct link pairs, so that mixed matches (types 31 and 32, corresponding to types D and E in Figure 6.2) are counted as two distinct link pairs—one under N linkage and one under E linkage.

Manually examining the mixed links where one event on one dataset matched with different (close) events for the same person in the other, in nine cases (18 links) the N link was preferred over the CSLA link, in 24 (48 links) the CSLA link was preferred over the N link, and in four cases (eight links) both were equally acceptable. In a majority of cases, different events were linked because of slight differences in the priorities used in the two strategies to choose between competing matches. Reasons why an N link was identified as the preferred link over the CSLA link included:

- an N match to social leave providing a better match than a CSLA match to an admission to a new RAC facility following the end of the leave period (priorities used in the CSLA strategy lead to matches to admissions being chosen over matches to social leave)
- an N match to hospital leave preferred over a CSLA match to a RAC re-admission at the end of the hospital leave (missed via CSLA due to poor postcode matching or poor event start dates).

Reasons why a CSLA link was preferred over the N link included:

- a CSLA match to hospital leave preferred over an N match to a RAC admission (priorities in N matching lead to choosing the admission that occurred just before the hospital leave)
- a CSLA closer match on hospital leave dates (with N linkage choosing the hospital leave just after relevant hospital leave).

Other reasons why different links were made related to apparent errors in the datasets, such as poor event date matches, and apparently contiguous hospital episodes without a statistical discharge for the first or a statistical admission for the second episode.

From the above discussion, it can be seen that the different events linked for the same people were generally very close, with the CSLA link preferred over the N link more often than vice versa (76% of CSLA links, or 28 out of 37, to the same person were the preferred link). On the other hand, for links to different people, the N link was much more likely than the CSLA link to be correct because of its use of name and address data in the matching process.

Taking the above into consideration, if a record in one dataset was matched by both N and CSLA linkages, and CSLA matched to an event for the same person as the N match in the second dataset, then for the purposes of match comparison the CSLA match was classified as a true link—that is, was said to be 'the same as' the N match. If, on the other hand, the CSLA-linked event was for a different person, then the CSLA match was classified as a false link. Using this categorisation, overall 95.6% (7,460) of the 7,802 CSLA matches for hospital events were true links, including 28 links for hospital events to different RAC events for the same person (that is, PPV = 95.6%). Also, these 7,460 matches accounted for 92.0% of the 8,106 N matches (sensitivity). Looking at the links from the RAC data perspective, 95.4% (7,440) of the 7,802 CSLA matches for RAC events were the same as the N matches, including 18 links to hospital events for the same person (sensitivity = 95.6%). In the remaining discussion, links are looked at from the hospital perspective. Very similar results would be obtained if a RAC perspective were to be used.

Refining CSLA matches

A match procedure in *Websphere®* used to compare a particular dataset pair can include up to seven different match passes (that is, seven separate match specifications—see Tables 5.1 and 5.2). The extent to which E matches are the same as N matches (that is, true links) can therefore be examined within both the procedures and passes used in the *Websphere®*

matching. These rates will vary depending on the particular datasets being compared, the nature of the passes and the measure of geography used. Poorly performing procedures and/or passes can then be identified and excluded.

When refining the E linkage we want to keep those match passes that lead to an acceptably high likelihood of making a true match and drop those that have an unacceptably high likelihood of making a false match. At the same time, we don't want to drop passes whose exclusion will result in missing a high number of true links. One way of identifying an acceptable strategy is to estimate the range of the actual PPV for a pass given the number of matches and true links observed under that strategy.

Assuming a binomial distribution of match outcomes (true versus false), a 95% one-sided confidence interval for the underlying PPV (as a rate) for a match strategy is approximated by

$$\text{PPV} \geq \text{PPV}' - 1.645\sqrt{\frac{\text{PPV}'(1-\text{PPV}')}{\text{Number of observed matches}}}$$

where PPV' is the observed positive predictive value (as a rate). From the point of view of ensuring a low false match rate, we would want this lower bound to be as high as possible; in particular, it should be above 0.5 so that it is highly likely that the strategy results in more true matches than false matches. Remembering that since we can be 95% sure of the underlying PPV being above this lower bound, setting a cut-off of 0.6 for this confidence limit means that, even if the actual PPV were as low as this, we would be making almost 50% more true matches than false matches under the strategy (60% versus 40%).

Figure 6.3 shows that the lower bound for the one-sided 95% confidence interval rises with PPV' and the total number of matches achieved. We can see that we can use the above rule to make decisions about true match rates based on quite small numbers of matches. For example, we can be highly confident that if we observe a PPV' of at least 0.8 out of 20 or more matches then the underlying PPV for that strategy will be more than 0.6. Even with as few as 10 matches, an observed PPV' of over 0.9 (that is, 9 or more out of the 10 E matches the same as the N match) implies an underlying PPV of more than 0.6. However, for an observed true match rate of 0.7, more than 70 matches need to be observed to be confident that the underlying PPV is over 0.6.

The above approximation for the 95% confidence interval does not hold very well when there are only very small numbers of matches.[1] In this case we can use the binomial distribution itself to aid decision making (see Table 6.2). For example, if there are five matches then the probability of observing three or more true matches is just over 30% if the underlying PPV is 0.4, and over 80% if the underlying PPV is 0.8. Consequently, observing a PPV' of 0.6 (3 out of 5) does not provide strong evidence of a true match rate of more than 50%. On the other hand, the probability of observing all five as true matches is only 1% if the underlying PPV is 0.4, so that observing a PPV' of 1.0 (5 out of 5) indicates that it is highly likely that the underlying PPV is greater than 0.4 when there are only five observed matches in all.

---

1 For statistical purposes, when estimating the proportion of items with a particular property there is said to be a small number of items if we observe five or fewer either with or without the required property—depending on whether the rate is below or above 0.5 (Hoel 1971:82). In the current situation, with an observed PPV' greater than 0.5, 'very small' implies that we have observed five or fewer CSLA matches that are different from N matches (that is, false matches).

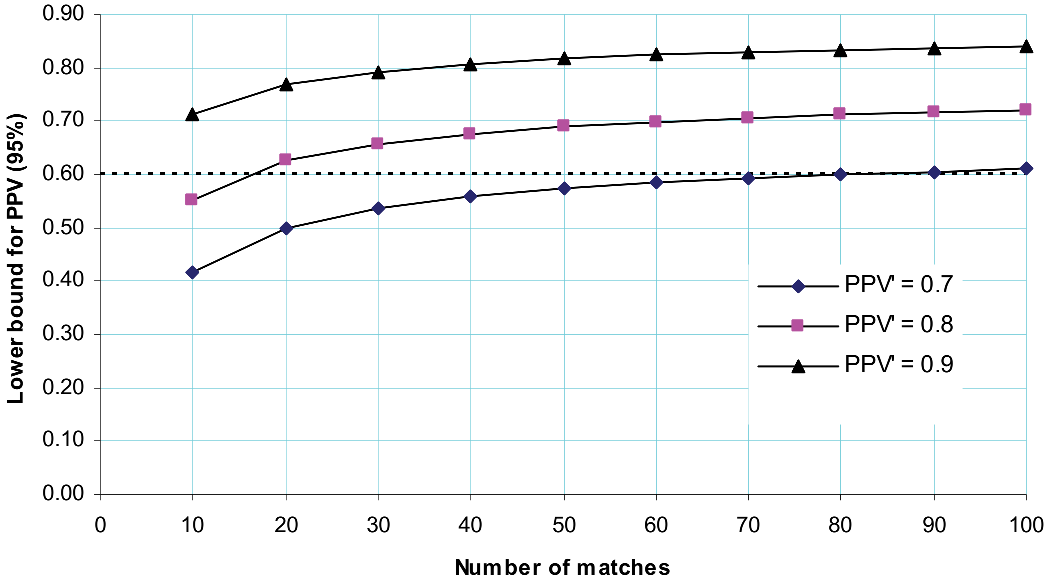


Figure 6.3: Lower bound of underlying PPV, by observed PPV' and number of matches achieved under E linkage

Table 6.2: Binomial distribution of matches, by number of matches and underlying PPV

Number n of true matches	Total number of matches: 10		5		2	
Underlying PPV = 0.4						
n	**Observed PPV'**	**Prob ≥n true matches**	**Observed PPV'**	**Prob ≥n true matches**	**Observed PPV'**	**Prob ≥n true matches**
0	0	1	0	1	0	1
1	0.1	0.994	0.2	0.922	0.5	0.640
2	0.2	0.954	0.4	0.663	1	0.160
3	0.3	0.833	0.6	0.317	. .	. .
4	0.4	0.618	0.8	0.087	. .	. .
5	0.5	0.367	1	0.010	. .	. .
6	0.6	0.166	. .	. .	. .	. .
7	0.7	0.055	. .	. .	. .	. .
8	0.8	0.012	. .	. .	. .	. .
9	0.9	0.002	. .	. .	. .	. .
10	1	0.000	. .	. .	. .	. .
					. .	. .

(continued)

Table 6.2 (continued): Binomial distribution of matches, by number of matches and underlying PPV

Number n of true matches	Total number of matches					
	10		5		2	
Underlying PPV = 0.7						
n	Observed PPV'	Prob ≥n true matches	Observed PPV'	Prob ≥n true matches	Observed PPV'	Prob ≥n true matches
0	0	1	0	1	0	1
1	0.1	1.000	0.2	0.998	0.5	0.910
2	0.2	1.000	0.4	0.969	1	0.490
3	0.3	0.998	0.6	0.837	. .	. .
4	0.4	0.989	0.8	0.528	. .	. .
5	0.5	0.953	1	0.168	. .	. .
6	0.6	0.850	. .	. .	. .	. .
7	0.7	0.650	. .	. .	. .	. .
8	0.8	0.383	. .	. .	. .	. .
9	0.9	0.149	. .	. .	. .	. .
10	1	0.028	. .	. .	. .	. .

The observed PPV' rates for the passes used in the *Websphere®* procedures for CSLA matching are shown in Table 6.3. These rates fall roughly into three groups: those with very high values (PPV' above 0.95), those with moderate to high values (PPV' between 0.75 and 0.95), and those with very poor values (PPV' under 0.33). The 95% one-sided confidence limit, as described above, can be used to identify *Websphere®* passes with moderate to high PPVs that should be retained, using the criterion of being highly confident that the underlying PPV was at least 0.6. This cut-off suggested that just one of the passes with a moderate to high PPV' should be excluded, along with those with a very poor observed PPV'. Applying these exclusion rules, the CSLA matching algorithm was refined by:

1. Not allowing date of birth variation when matching between RAC admissions and hospital discharge to usual residence: both CSL9ADM and CSLH9ADM had poor PPV's for passes 3 and 4 in which date of birth variations were allowed.
2. Not allowing event date variation when matching between RAC admissions and hospital discharge to usual residence using the hospital postcode: CSLH9ADM pass 2 had a PPV' of 75% which was not significantly above 60% (based on 16 matches), and pass 5 of the same procedure had no true links out of just three matches.

Apart from these passes, there were a number with fewer than 10 matches but with a 100% concordance with N matches. In this situation, it is not always clear that a decision can be made concerning the utility of the *Websphere®* pass. For example, when there are just two matches, the probability of observing one or two true matches is 64% even when the underlying PPV is as low as 0.4 (Table 6.2). However, even in this case there is only a 16% chance of both being correct (that is, 2 out of 2). Furthermore, because the numbers are so small, any false matches resulting from these passes would have minimal effect on the matched dataset. In the current analysis, those procedures linking RAC admissions to hospital separations other than to usual residence (CSLH0ADM) were retained, as the hospital separation mode and type of RAC event correspond (two matches each for passes 3 and 4, and seven matches for pass 5). Also, links between hospital separations to RAC hospital leave events were retained (CSLST0H), as a hospital event had been recorded on the RAC data (two matches each for passes 2 and 3, and four matches for pass 4.). In addition, although very few cases were matched in individual procedures when using the first three and two digits of postcode (except in CSL0ADM), the overall PPV' for these passes was high, suggesting that they should be retained. These latter passes all used exact event date data. (Note: variation in regional matching is discussed in more detail later.)

Excluding the six passes identified as poor performers noticeably improved the PPV' of the two procedures involved: from 76% to 93% for CSL9ADM, and from 69% to 100% for CSLH9ADM. Looking across all matches, the number of links that were CSLA-only (false matches, from both the hospital and RAC perspectives) dropped from 310 links before refinement to 160 after refinement, and the number of mixed links to events for different people dropped from 10 to 4 from the hospital perspective and from 38 to 14 from the RAC perspective (compare Table 6.1 with Table 6.4, hospital perspective in rows, RAC perspective in columns). Such refinement also has a cost in terms of true links, and the number of CSLA links that were identical to an N link dropped by under 1% (by 42 to 7,389), with a correspondingly small increase in the number identified only by the N linkage (CSLA false negatives). Taken altogether, refining CSLA linkage resulted in dropping 207 matches including 42 true matches, leading to a total of 7,595 CSLA matches. As a result, overall the PPV for the CSLA strategy rose from 95.6% to 97.7% (7,418 out of 7,595; Table 6.3), and the sensitivity dropped marginally from 92.0% to 91.5% of the N matches (7,418 out of 8,106).

Table 6.3: Summary of CSLA matches and positive predictive value using N links as the reference standard, by *Websphere*® procedure and pass

	Pass							
	1	2	3	4	5	6	7	
Procedure	**Exact match within SLA group**	**1-sided event date variation**	**YOB variation**	**Variation in month *or* day in DOB**	**2-sided event date variation**	**Exact match within first 3 postcode digits**	**Exact match within first 2 postcode digits**	**All links**
				Number of matches				
Initial matching algorithm, to deaths in hospital (but not RAC discharges to hospital)								
CSL8ADM	—	1	1	4	4	5	6	21
CSLNST8H	267	49	7	10	6	—	4	343
CSLST8H	64	11	1	—	—	—	—	76
Initial matching algorithm, excluding matches to deaths in hospital								
CSL0ADM	1,237	21	19	47	33	54	207	1,618
CSL9ADM	506	32	52	129	18	18	72	827
CSL9SOC	154	. .	. .	. .	. .	. .	. .	154
CSLH0ADM	34	17	2	2	7	5	16	83
CSLH9ADM	11	16	—	7	3	1	4	42
CSLNST0H	520	106	14	19	—	2	9	670
CSLNST9H	2,983	627	45	97	—	12	25	3,789
CSLST0H	157	2	2	4	—	4	4	173
CSLST9H	394	11	13	19	—	4	5	446
Total (all links, excluding deaths in hospital)	**5,996**	**832**	**147**	**324**	**61**	**100**	**342**	**7,802**
Refined matching algorithm								
CSL9ADM	506	32	. .	. .	18	18	72	646
CSLH9ADM	11	. .	. .	. .	. .	1	4	16
Total (all links, excluding deaths in hospital)	**5,996**	**816**	**95**	**188**	**58**	**100**	**342**	**7,595**

(continued)

Table 6.3 (continued): Summary of CSLA matches and positive predictive value using N links as the reference standard, by *Websphere*® procedure and pass

	Pass								
	1	2	3	4	5	6	7		
Procedure	Exact match within SLA group	1-sided event date variation	YOB variation	Variation in month *or* day in DOB	2-sided event date variation	Exact match within first 3 postcode digits	Exact match within first 2 postcode digits	All links	
	PPV (per cent CSLA links)								Number
Initial matching algorithm, to deaths in hospital (excluding discharges to hospital)									
CSL8ADM	. .	—	—	—	—	—	—	—	21
CSLNST8H	99.3	100.0	**100.0	*100.0	**100.0	. .	**100.0	99.4	343
CSLST8H	100.0	*90.9	**100.0	. .	. .	. .	. .	98.7	76
Initial matching algorithm, excluding matches to deaths in hospital									
CSL0ADM	97.3	81.0	*79.0	95.7	100.0	96.3	95.2	96.5	1,618
CSL9ADM	96.4	75.0	[(a)] ***32.7***	[(a)] ***9.3***	*77.8	*94.4	81.9	76.3	827
CSL9SOC	95.5	. .	. .	. .	. .	. .	. .	95.5	154
CSLH0ADM	97.1	88.2	**100.0	**100.0	**100.0	**100.0	*100.0	96.4	83
CSLH9ADM	*100.0	[(a)] ****75.0***	***. .***	[(a)] *****14.3***	[(a)] *****0.0***	**100.0	**100.0	69.0	42
CSLNST0H	99.4	100.0	*100.0	*100.0	. .	**100.0	**100.0	99.6	670
CSLNST9H	98.7	98.1	100.0	96.9	. .	*100.0	96.0	98.5	3,789
CSLST0H	99.4	**100.0	**100.0	**100.0	. .	**100.0	**100.0	99.4	173
CSLST9H	99.0	*90.9	*100.0	*84.2	. .	**100.0	**100.0	98.2	446
Total PPV (%)	**98.2**	**96.3**	**73.5**	**59.6**	**88.5**	**97.0**	**93.0**	**95.6**	**. .**
Total (all links, excluding deaths in hospital)	**5,996**	**832**	**147**	**324**	**61**	**100**	**342**	**7,460**	**7,802**
Refined matching algorithm									
CSL9ADM	96.4	75.0	. .	. .	*77.8	*94.4	81.9	93.2	646
CSLH9ADM	*100.0	. .	. .	. .	. .	100.0	100.0	100.0	16
Total PPV (%)	**98.2**	**96.7**	**95.8**	**95.7**	**93.1**	**97.0**	**93.0**	**97.7**	**. .**
Total (all links, excluding deaths in hospital)	**5,996**	**816**	**95**	**188**	**58**	**100**	**342**	**7,418**	**7,595**

(a) PPV not statistically above 60%, using one-sided 95% confidence interval; in bold italics to indicate dropped from refined algorithm.

* Based on 10–19 matches.

** Based on fewer than 10 matches.

Notes

1. Table counts links to hospital records, with true links including links to different RAC events for the same person.
2. Table excludes links to discharges to hospital while on RAC leave.
3. Preliminary analysis showed that allowing discrepancies in both date of birth and event dates leads to a high number of mismatches when compared with N links. Such passes were therefore dropped and replaced with 2- and 3-digit postcode matches.

Table 6.4: Comparing N and CSLA refined links: distinct link pairs

Match to hospital event	Match to RAC event					
	1. RAC event links to same hospital event	2. RAC event links under N linkage only	31. RAC event links to same person, but diff. hospital event	32. RAC event links to hospital events for diff. people	4. RAC event links under CSLA only	Total
1. Hospital event links to same RAC event	7,389	. .	. .	. .	. .	7,389
2. Hospital event links under N linkage only	. .	671	8	7	. .	686
31. Hospital event links to same person, but different RAC event	. .	28	1	. .	29	58
32. Hospital event links to RAC events for different people	. .	2	. .	1	1	4
4. Hospital event links under CSLA only	. .	. .	9	6	160	175
Total	7,389	701	18	14	190	8,312

Note: The table counts distinct link pairs, so that mixed matches (types 31 and 32, corresponding to types D and E in Figure 6.2) are counted as two distinct link pairs—one under N linkage and one under E linkage.

False negatives and false positives: why CSLA missed links and made false links

The reasons for the CSLA strategy missing some of the links identified by N linkage, and for making links under CSLA that were not identified by N linkage are of interest for two reasons. First, analysis may further identify ways to improve the E linkage strategy, and second, any patterns in the missed and false links could indicate whether there are likely to be biases in the CSLA linked dataset. A range of investigations into these issues were carried out, and the results are summarised below. Detailed results and discussion are given in Appendix 4.

False negatives, or missed links

N-only links were examined to determine the reason(s) that matches were missed by the CSLA strategy. The comparison was based on the refined CSLA strategy and excluded any N-only links that had a related link under CSLA (that is, a link to a different event for the same person—discussed previously). Overall, there were 671 links made via N linkage and not by the CSLA strategy. Over half of these related to hospital leave (53%), 46% were for RAC admissions and the remaining 1.5% were links between hospital events and RAC social leave (Table 6.5).

Table 6.5: CSLA missed links, using N links as the reference standard

	Hospital separation mode						
RAC event type	To RAC	To other health	Left against medical advice/statistical discharge from leave/unknown	Died	To usual res./other	Total	
			Number				%
Admissions	101	60	18	0	129	308	45.9
Hospital leave	35	34	16	4	264	353	52.6
Social leave	2	0	0	0	8	10	1.5
Total	**138**	**94**	**34**	**4**	**401**	**671**	**100.0**

Note: Table excludes mixed CSLA–N links, that is, types D, E and F in Figure 6.2.

Several measures were looked at to identify reasons for missing matches when using CSLA matching:

- Date of birth differences: dates of birth on the RAC and hospital datasets are not sufficiently similar for CSLA matching if they differ by two or more elements, or if the years of birth differ by 8 or more years.
- Different sex on the two datasets: unacceptable in the E linkage strategy.
- Possible SLA group difference, identified by different postcodes on the two datasets.
- Poor hospital-to-RAC date match, measured by (RAC in date – hospital out date). This is relevant for all RAC events, and is unacceptable for CSLA links if it is negative or more than 2 days.
- Poor RAC-to-hospital date match, measured by (hospital in date – RAC out date). This is relevant for RAC leave events only, and is unacceptable for CSLA links if it is negative or more than 2 days.

Around 13% of missed links appeared to have sufficiently matching data on the two datasets to enable a CSLA link to be identified. However, nearly half of these were matches that had been dropped when refining the CSLA strategy to exclude poorly performing *Websphere®* passes and procedures. The remainder all contained one element of mismatch which could have caused their exclusion via the *Websphere®* match selection weighting algorithms.

For N-only links to RAC admissions, poor region matching was the main reason for missing these matches under the CSLA strategy. This was a less important, but still significant, reason for missing matches among RAC leave events. Because of the very limited geographic data available on the hospital and RAC datasets for use in E linkage, matches missed due to poor region matching on the two datasets cannot be retrieved via adjusting the CSLA strategy.

Missed matches to RAC leave events were primarily the result of poorly matching event dates for related hospital and RAC events. The large gaps between the recorded dates for the linked events in these missed matches (commonly 3 or more days) indicate that it would not be possible to adjust the CSLA matching strategy to allow capture of these matches without risking the introduction of large numbers of false matches. A relatively small proportion of missed links had poorly matching information on both region and event dates—under 10% of all missed links (54 out of 671).

False matches

Using N matches as the reference standard, CSLA-only matches represent false matches made by the strategy. As when looking at missed matches, CSLA-only links that had a related link under N linkage were excluded from the analysis. Excluding these mixed matches, there were 160 CSLA-only matches when using the refined CSLA strategy (Table 6.6). The majority of these were for admissions (100, or 63%), and one-third were for RAC hospital leave events (53); the remaining seven related to RAC social leave.

To ensure that CSLA false links were not the result of person links being missed by N linkage, the reliability of the N linkage person links was re-examined by reviewing HiL's person links for the 160 CSLA-only links. Investigations by HiL of the client links implied by the CSLA-only links led to HiL identifying just two additional person links (out of a possible 160) between hospital and RAC clients that had previously been missed.

Among false links to RAC admissions, under half had the appropriate mode of separation (that is, 'to residential aged care') for the hospital event (42 out of 100; Table 6.6). However, a mismatch between the hospital mode of separation and the type of RAC event to which the hospital event was linked does not provide a basis for identifying false links as, in general, there seems to be considerable confusion in the hospital data when reporting the discharge destination for people being admitted to RAC. For example, among all CSLA links to RAC admissions, 1,701 had other health care establishment or RAC facility as their recorded post-hospital destination compared with 662 who were reported as going to their usual residence (which is what the RAC facility becomes if it is a permanent admission).

Table 6.6: CSLA-only links (number)

	Hospital separation mode			
RAC event type	**To RAC**	**To other health**	**To usual residence/other**	**Total**
Admission	42	15	43	100
Hospital leave	3	—	50	53
Social leave	—	—	7	7
Total	**45**	**15**	**100**	**160**

Note: Table excludes mixed CSLA–N links, that is, types D, E and F in Figure 6.2.

A high proportion of CSLA-only links had exact matches on both date of birth and event dates. This suggests that many of the false matches were caused by similar hospital and RAC activity by similar people (in terms of date of birth and sex) living in a particular region. In addition, comparisons indicate that one of the most effective ways of reducing the number of false matches made under the CSLA strategy would be to reduce the size of the geographic region used in matching. Analysis of CSLA-only links indicated that this would have a greater effect on links to RAC admissions than on links to RAC leave. However, narrowing the geographic matching criteria could result in dropping many more true links than false links (see passes 6 and 7 in Table 6.3). The effect of varying geographic region on the CSLA linkage strategy is discussed below.

Varying geographic region

The overall PPV when using the first 3 or the first 2 digits of the postcode to define the region used for matching in the CSLA strategy was high (over 90%, see Table 6.3), suggesting that this part of the match strategy should be retained. However, examination of CSLA false links points to reducing the match region size as one of the most effective means for lowering the number of false matches. In addition, theoretical analysis of the E linkage process indicates that false match rates increase with population size, and can get unacceptably high when matching within groups as large as 40,000 (AIHW: Karmel 2004). Furthermore, investigation into the distribution of the older population across 2- and 3-digit postcodes shows that using 2-digit postcodes could result in many matches being undertaken within large population groups (over 40,000) for some jurisdictions (see Tables A5.9 and A5.10).

Looking at matches using just the Western Australian data, there was limited evidence of false match rates increasing with population size: in the examples presented in Table 6.7, results for CSL9ADM (linking hospital discharges to 'usual residence' to RAC admissions) suggest some decline in the PPV with increasing population size, while the results for CSLNST9H (linking hospital discharges to 'usual residence' to RAC hospital leave) did not. However, random variation in match rates may be masking the effect to some extent. For example, even with an underlying PPV of 98%, within 1,000 matches we would expect to observe PPV rates of between 97% and 99% (95% confidence interval). In general, in the current project of using constrained E linkage within Western Australia, the number of matches within sex, region size and *Websphere*® procedure was always less than 1,000, and commonly less than 250. Among 250 matches, an underlying PPV of 95% would result in observed PPVs of 92–98% (95% confidence interval), so observed differences such as most of those seen in Table 6.7 are not statistically significant.

Table 6.7: CSLA positive predictive values using N links as the reference standard, for selected *Websphere*® procedures (refined) for women, by SLA group population size (% CSLA links)

SLA group women 65+ years	CSLNST9H PPV	CSLNST9H All matches using CSLA (refined)	CLS9ADM PPV	CLS9ADM All matches using CSLA (refined)	All procedures PPV	All procedures All matches using CSLA (refined)
	Per cent	Number	Per cent	Number	Per cent	Number
<2,500	98.0	646	97.3	110	97.8	1,187
2,501–5,000	98.8	810	97.8	92	98.4	1,440
5,001–7,500	97.8	324	94.9	59	97.4	646
7,501–10,000	98.9	278	90.6	53	98.1	585
10,001–15,000	98.5	390	83.3	36	97.1	687
15,001–20,000	100.0	133	90.5	21	98.1	263
20,001–30,000	100.0	2	66.7	3	87.5	8
Total (%)	**98.5**	. .	**94.1**	. .	97.9	. .
Total (number for women)	**2,544**	**2,583**	**352**	**374**	**4,713**	**4,816**

Note: The procedure CSLNST9H for women results in the largest number of matches when using CSLA matching. Matches using 2- and 3-digit postcodes (but not matching on SLA group) are not included.

On the other hand, PPVs for passes combined across CSLA procedures indicate that links made using 2-digit postcode, but not 3-digit postcode or SLA group, were more likely to be erroneous than those matching completely on 3-digit postcode. For example, links achieved using exact date and 3-digit postcode matching had a PPV of 97.0% while those achieved using exact date and only 2-digit postcode had a PPV of 93.0% (see passes 6 and 7 in

Table 6.3). While this difference is not statistically significant at the 95% confidence level, there is further evidence that broadening the region reduces the PPV when postcode itself is used as the underlying matching region, rather than SLA group: links achieved using exact date and postcode matching had a PPV of 98.4% while those achieved using exact date and only 2-digit postcode had a PPV of 94.7% (see passes 1 and 7 in Table 6.10). Given the number of matches made (5,559 and 509), such a difference is statistically significant above the 99% confidence level.

This last result in conjunction with the theoretical analysis of the E matching process indicates that false match rates could get unacceptably high when matching within large populations. Consequently, given the risk of biases introduced by differential PPVs for different jurisdictions due to the large variation in population size for 2-digit postcodes (see Section A5.3), matching within 2-digit postcodes should not be adopted for national analysis.

Excluding matches made using 2-digit postcodes, rather than SLA group, results in 342 matches being dropped from the refined links, leaving a total of 7,253 CSLA matches (Table 6.8). Of the matches dropped, 299 were for links to RAC admissions and 43 were for links to RAC hospital leave. Because these matches had a relatively low PPV (93%; see pass 7 in Table 6.3), excluding them led to a small rise in the PPV—from 97.7% to 97.9% of CSLA matches (including a small number of matches to different RAC events for the same person; Table 6.9). The effect on the sensitivity of the CSLA strategy, however, was more pronounced, with sensitivity dropping from 91.5% to 87.5% of the N matches (7,100 out of 8,106). These results suggest that for states with generally less populous 2-digit postcodes (Australian Capital Territory, Northern Territory, Tasmania and Western Australia—see Appendix 5), excluding 2-digit postcode matching could result in dropping too many true links for little gain in the PPV of the strategy. Consequently, retention of 2-digit postcode matching should be considered for some state-based analyses, depending on the distribution of the population across 2-digit postcodes.

Table 6.8: Summary of refined CSLA linkage using N links as the reference standard, excluding 2-digit postcode matching, by *Websphere®* pass (% CSLA links)

	Pass						All	
	1	2	3	4	5	6	%	Total
1. and 31. Hospital event links to same RAC event, or event for same person	98.2	96.7	95.8	95.7	93.1	97.0	97.9	7,100
32. Hospital event links to RAC events for different people	—	0.1	—	—	—	—	—	2
4. Hospital event links under CSLA only	1.8	3.2	4.2	4.3	6.9	3.0	2.1	151
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	7,253
Total (% match)	**82.7**	**11.3**	**1.3**	**2.6**	**0.8**	**1.4**	**100.0**	**. .**
All links (number)	**5,996**	**816**	**95**	**188**	**58**	**100**	**. .**	**7,253**

Table 6.9: Comparing N and refined CSLA links excluding 2-digit postcode matching: distinct link pairs

Match to hospital event	Match to RAC event					
	1. RAC event links to same hospital event	2. RAC event links under N linkage only	31. RAC event links to same person, but diff. hospital event	32. RAC event links to hospital events for diff. people	4. RAC event links under CSLA only	Total
1. Hospital event links to same RAC event	7,072	. .	. .	. .	. .	7,072
2. Hospital event links under N linkage only	. .	989	8	7	. .	1,004
31. Hospital event links to same person, but different RAC event	. .	27	1	. .	28	56
32. Hospital event links to RAC events for different people	. .	2	. .	1	1	4
4. Hospital event links under CSLA only	. .	. .	9	6	136	151
Total	**7,072**	**1,018**	**18**	**14**	**165**	**8,287**

Note: The table counts distinct link pairs, so that mixed matches (types 31 and 32, corresponding to types D and E in Figure 6.2) are counted as two distinct link pairs—one under N linkage and one under E linkage.

6.3 Constrained matching within postcode (CPC)

Linking within SLA group can allow quite geographically separated areas to be considered a match (see Figure 5.1). In addition, using SLA group rather than postcode adds considerable complexity to preparing the data for the matching process. Therefore, using postcode as the matching region was also investigated.

Since postcodes are smaller than their related SLA groups, we would expect to get fewer linked records (called CPC matches) when using straight postcode as the primary measure of region as opposed to SLA group. However, incorporating matching using only the first 3 or 2 digits of the postcode in some of the match passes allows matching on broader regions (see Table 6.10). As many SLA group links also match on 3- or 2-digit postcode (illustrated in Table A4.6), there should be strong similarities between the two approaches.

After excluding 582 CPC links to RAC leave which ended in either discharge to hospital (44) or death in hospital (447), or both (91), CPC matching resulted in 7,781 matches compared with 7,802 for (unrefined) CSLA. Broad comparisons of CPC and N links are discussed below and presented in Tables 6.10 and 6.11. Differences between the CPC and CSLA links are then examined.

Overall, 95.7% (7,450) of the 7,781 CPC matches were true links (including 28 links for hospital events to different RAC events for the same person). The sensitivity of the strategy was 91.8% (7,450 out of 8,106).

The PPVs for the various passes for the *Websphere*® procedures used for CPC matching are shown in Table 6.10. As for CSLA links, these rates fall roughly into three groups: those with very high PPVs (above 95%), those with moderate to high PPVs (65–95%), and those with very poor PPVs (under 30%). Again, using 95% one-sided confidence intervals to identify poorly performing *Websphere*® passes, the CPC matching algorithm was refined by:

1. Not allowing date of birth variation when matching between RAC admissions and hospital discharge to usual residence: both CPC9ADM and CPCH9ADM had poor true match rates for passes 3 and 4 in which date of birth variations were allowed.
2. Not allowing two-sided event date variation when matching between RAC admissions and hospital discharge to usual residence using the hospital postcode (CSLH9ADM pass 5). (Note that all event date variation was dropped for the corresponding procedure for CSLA. This difference in approach was due to a slightly higher PPV under CPC, caused by linking within 3-digit postcode rather than SLA group.)
3. Not allowing 3- or 2-digit postcode matching when linking to RAC social leave (CPC9SOC passes 2 and 3).

Excluding these seven passes resulted in dropping 194 matches, including 36 true matches (when compared with N links), leading to a total of 7,587 CPC matches. As a consequence, the overall PPV rose from 95.7% to 97.8% (7,418 out of 7,587; Table 6.10), and the sensitivity decreased marginally from 91.9% to 91.5% (7,418 out of 8,106).

Table 6.10: Summary of CPC matches and positive predictive value using N links as the reference standard, by *Websphere®* procedure and pass

	Pass						
	1	2	3	4	5	7	
		Using first 3 digits of postcode					
Procedure	**Exact match within postcode**	**1-sided event date variation**	**YOB variation**	**Variation in month *or* day in DOB**	**2-sided event date variation**	**Exact match within first 2 postcode digits**	**All links**
			Number of matches				
Initial matching algorithm, to deaths in hospital (but not RAC discharges to hospital)							
CPC8ADM	Dropped in light of CSLA results						
CPCNST8H	260	52	7	10	6	7	342
CPCST8H	62	11	1	—	—	2	76
Initial matching algorithm, excluding matches to deaths in hospital							
CPC0ADM	998	217	18	46	37	300	1,616
CPC9ADM	431	92	61	109	16	105	814
CPC9SOC	147	(a)***3***	(b)***11***	. .	. .	. .	161
CPCH0ADM	26	33	1	1	4	19	84
CPCH9ADM	7	17	3	6	1	6	40
CPCNST0H	503	119	14	19	—	14	669
CPCNST9H	2,910	679	46	96	—	49	3,780
CPCST0H	153	9	2	3	—	5	172
CPCST9H	384	18	13	19	—	11	445
Total (all links, excluding deaths in hospital)	**5,559**	**1,187**	**169**	**299**	**58**	**509**	**7,781**
Refined matching algorithm							
CPC9ADM	431	92	. .	. .	16	105	644
CPC9SOC	147	. .	. .	. .	. .	. .	147
CPCH9ADM	7	17	. .	. .	. .	6	30
Total (all links, excluding deaths in hospital)	**5,559**	**1,184**	**94**	**184**	**57**	**509**	**7,587**

(continued)

Table 6.10 (continued): Summary of CPC matches and positive predictive value using N links as the reference standard, by *Websphere®* procedure and pass

	Pass							
	1	2	3	4	5	7		
		Using first 3 digits of postcode						
Procedure	Exact match within postcode	1-sided event date variation	YOB variation	Variation in month *or* day in DOB	2-sided event date variation	Exact match within first 2 postcode digits	All links	
	PPV (per cent CPC links)							Number
Initial matching algorithm, to deaths in hospital (but not RAC discharges to hospital)								
CPC8ADM	Dropped following CSLA results							
CPCNST8H	99.6	98.1	**100.0	*100.0	**100.0	**100.0	99.4	342
CPCST8H	100.0	*90.9	**100.0	. .	. .	**100.0	98.7	76
Initial matching algorithm, excluding matches to deaths in hospital								
CPC0ADM	96.9	96.3	77.8	93.5	97.3	96.7	96.5	1,616
CPC9ADM	97.0	88.0	(c)***27.9***	(c)***11.9***	*93.8	85.7	77.9	814
CPC9SOC	99.3	(c)(a)*****66.7***	(c)(b)****0.0***	. .	. .	. .	91.9	161
CPCH0ADM	96.2	97.0	**100.0	**100.0	**100.0	*100.0	97.6	84
CPCH9ADM	**100.0	*82.4	(c)*****0.0***	(c)*****0.0***	(c)*****0.0***	**100.0	67.5	40
CPCNST0H	99.4	100.0	*100.0	*100.0	. .	*100.0	99.6	669
CPCNST9H	98.8	98.1	100.0	95.8	. .	95.9	98.6	3,780
CPCST0H	99.4	**100.0	**100.0	**100.0	. .	**100.0	99.4	172
CPCST9H	99.0	*94.4	*100.0	*84.2	. .	*100.0	98.2	445
Total PPV (%)	**98.4**	**96.8**	**63.3**	**62.5**	**94.8**	**94.7**	**95.6**	**. .**
Total (all links, exc. deaths in hospital)	**5,559**	**1,187**	**169**	**299**	**58**	**509**	**7,450**	**7,781**
Refined matching algorithm								
CPC9ADM	97.0	88.0	. .	. .	*93.8	85.7	93.8	644
CPC9SOC	99.3	. .	. .	. .	. .	. .	99.3	147
CPCH9ADM	**100.0	*82.4	. .	. .	. .	**100.0	90.0	30
Total PPV (%)	**98.4**	**96.9**	**95.7**	**94.6**	**96.5**	**94.7**	**97.8**	**. .**
Total (all links, exc. deaths in hospital)	**5,559**	**1,184**	**94**	**184**	**57**	**509**	**7,418**	**7,587**
Excluding 2-digit postcode matching								
Total PPV (%)	**98.4**	**96.9**	**95.7**	**94.6**	**96.5**	**. .**	**98.0**	**. .**
Total (all links, exc. deaths in hospital)	**5,559**	**1,184**	**94**	**184**	**57**	**. .**	**6,936**	**7,078**

(a) Cover match, using 3-digit postcode.

(b) Cover match, using 2-digit postcode and full date of birth.

(c) PPV not statistically above 60%, using one-sided 95% confidence interval; in bold italics to indicate dropped from refined algorithm.

* Based on 10–19 matches.

** Based on fewer than 10 matches.

Note: Table counts links to hospital records, with true links including links to different RAC events for the same person. Links to discharges to hospital while on RAC leave are excluded.

Dropping matches achieved by using the first 2 digits of postcode had a greater effect on postcode-based matching than on matching done primarily using SLA groups. This is because SLA groups commonly cover more than one postcode so that in some cases matching by SLA group will equate to matching by the first 2 digits of postcode. Overall, excluding links made using 2-digit postcode resulted in dropping 509 CPC matches (including 482 true links), leaving 7,078 CPC matches of which 6,936 were true links (Table 6.11). This exclusion had minimal effect on the PPV of the strategy, but reduced the sensitivity by 6 percentage points—from 91.5% down to 85.6%.

Table 6.11: Comparing N and CPC links: distinct link pairs by matching strategies

	Initial	Refined	Refined, excluding 2-digit postcode
1. Hospital event links to same RAC event	7,422	7,390	6,910
2. Hospital event links under N linkage only	651	684	1,166
31. Hospital event links to same person, but different RAC event	56	56	52
32. Hospital event links to RAC events for different people	10	8	8
4. Hospital event links under CPC only	326	165	138
Total	**8,465**	**8,303**	**8,274**

Note: The table counts distinct link pairs, so that mixed matches (types 31 and 32, corresponding to types D and E in Figure 6.2) are counted as two distinct link pairs—one under N linkage and one under E linkage.

Comparing CPC with CSLA

As stated above, the initial CPC strategy resulted in 7,781 matches compared with 7,802 for CSLA—just 0.3% fewer matches. Of the CPC matches, 98.5% (or 7,664) were the same as those made under CSLA (Table 6.12). Differences in the matches are primarily seen in the stage (that is, *Websphere®* pass) at which links were made as, unlike the CSLA strategy, the CPC strategy included relaxing the regional match requirement at pass 2 (compare Tables 6.3 and 6.10).

Overall, when comparing the initial (unrefined) CSLA and CPC strategies, 136 of the matches to hospital events were made only using CSLA linkage and 115 were made only using CPC matching (Table 6.12). Many of the differences between CPC and CSLA matches were the result of date of birth differences in the RAC and hospital data (87 and 94 of the CPC-only and CSLA-only links to hospital events, respectively). Only six of the CSLA-only links were for links based on exact date, date of birth and SLA matching (that is, pass 1); two of the different matches for both strategies result from 2-digit postcode matching.

Excluding CSLA and CPC match strategies with poor PPVs more than halved the number of matches made only under either CSLA or CPC ('refined matches' column in Table 6.12). As a consequence, 99.3% (7,537) of the remaining 7,587 CPC matches were the same as CSLA matches (including links to different events for the same person), and these accounted for 99.2% of all 7,595 CSLA links.

Table 6.12: Comparing initial CSLA and initial CPC links: distinct link pairs

	Match to RAC event							
Match to hospital event	1. RAC event links to same hospital event	2. RAC event links under CSLA only	31. RAC event links to same person, but diff. hospital event	32. RAC event links to hospital events for diff. people	4. RAC event links under CPC only	Total	(a)Total for refined matches	(b)Total for refined matches, and exc. pass 7
1. Hospital event links to same RAC event	7,664	. .	. .	. .	. .	7,664	7,537	7,030
2. Hospital event links under CSLA only	. .	123	2	11	. .	136	57	222
32. Hospital event links to RAC events for different people	. .	2	. .	—	2	4	2	2
4. Hospital event links under CPC only	. .	. .	2	47	102	115	49	47
Total	**7,664**	**125**	**4**	**7,301**	**104**	**7,919**	**7,645**	**7,301**

(a) Excludes CSLA and CPC passes with poor correspondence with N matches.

(b) Excludes CSLA and CPC passes with poor correspondence with N matches; also excludes matches made using 2-digit postcode information.

Note: The table counts distinct link pairs, so that mixed matches (types 31 and 32, corresponding to types D and E in Figure 6.2) are counted as two distinct link pairs—one under CSLA linkage and one under CPC linkage.

After dropping matches made using 2-digit postcode information from the refined matching strategies, 99.3% (7,030) of the remaining 7,078 CPC matches were identical to a CSLA match, and these accounted for 96.9% of the 7,253 CSLA matches. The number of matches made only under CSLA increased to 222 while those made using CPC only dropped to 47. These shifts reflect the greater number of matches made via 2-digit postcode under CPC when compared with CSLA linkage.

Because of the high concordance between CSLA and CPC matches, factors leading to CPC missed and false matches (when compared with N matches) were not examined.

6.4 Basic matching within SLA group (BSESLA)

Under basic E linkage, the same linkage process is used for all hospital and RAC events, irrespective of their admission and separation characteristics. As a consequence, match data are limited to date of birth, sex, region of usual residence and a single movement date (that is, hospital separation and RAC entry date). Basic matching using SLA group as the match region (BSESLA) was used in the initial feasibility study of hospital–RAC event-based linkage (AIHW 2003). It is therefore important to compare the results from this approach with those obtained using the person-based linkage. It is also useful to compare the effectiveness of the BSESLA strategy with that of the constrained approaches to measure the gains (if any) in moving from a simple strategy to the more complex constrained one.

Using BSESLA linkage resulted in 6,693 matches, compared with 8,106 for N linkage, and 7,253 using refined constrained SLA group matching without 2-digit postcode (Tables 6.13

and 6.14). Overall, the PPV of the BSESLA hospital matches was 97.7%, including 37 links to different events for the same person. These 6,539 true matches accounted for 80.7% of the N matches (sensitivity).

Table 6.13: Comparing N and BSESLA links: distinct link pairs

	Match to RAC event					
Match to hospital event	**1. RAC event links to same hospital event**	**2. RAC event links under N linkage only**	**31. RAC event links to same person, but diff. hospital event**	**32. RAC event links to hospital events for diff. people**	**4. RAC event links under BSESLA only**	**Total**
1. Hospital event links to same RAC event	6,502	. .	. .	. .	. .	6,502
2. Hospital event links under N linkage only	. .	1,547	9	9	. .	1,565
31. Hospital event links to same person, but different RAC event	. .	36	1	. .	37	74
32. Hospital event links to RAC events for different people	. .	2	. .	1	1	4
4. Hospital event links under BSESLA only	. .	. .	10	8	134	152
Total	**6,502**	**1,585**	**20**	**18**	**172**	**8,297**

Notes

1. Table excludes identified deaths in hospital and discharges to hospital in the N matches, and BSESLA-identified deaths in hospital. Analysis indicated that *Websphere*® BSESLA passes allowing variation in date of birth with or without variation in event date led to PPVs of under 55%. These passes were therefore excluded.
2. The table counts distinct link pairs, so that mixed matches (types 31 and 32, corresponding to types D and E in Figure 6.2) are counted as two distinct link pairs—one under N linkage and one under E linkage.

Given the high PPV of BSESLA matching, the big difference between this and the constrained match strategies is in the number of matches made: BSESLA had nearly 8% fewer matches than CSLA matching excluding 2-digit postcode links (6,693 versus 7,253). Whether there are also differences in the types of events being linked (for example, RAC admissions compared with RAC leave) is examined in detail later, but a brief examination shows that N links missed by the BSESLA strategy were more likely to be to RAC admissions than those missed by CSLA—52% of missed links compared with 46% (compare Tables 6.5 and 6.15).

When comparing the BSESLA links with the refined CSLA links excluding those made using 2-digit postcode, nearly 98% of the BSESLA links were the same (or for the same person) as a CSLA link (Table 6.16). Not surprisingly, this concordance was higher among links based on exact event date matching than among matches made allowing some event date variation (99% versus 77%). Interestingly, while 77% of the BSESLA links made using event date variation corresponded with a CSLA link, 88% corresponded with an N link (a difference, however, of just 24 links).

Table 6.14: Summary of BSESLA matches using N links as the reference standard, by *Websphere*® pass (% BSESLA links)

	Pass			
	1 (exact match)	2 (with 1-sided event date variation)	Total	
1. and 31. Hospital event links to same RAC event or event for same person under N linkage and BSESLA (PPV)	98.1	87.6	97.7	6,539
32. Hospital event links to different RAC events for different people	—	—	—	2
4. Hospital event links under BSESLA only	1.9	12.4	2.3	152
Total (%)	**100.0**	**100.0**	**100.0**	**. .**
Total (number)	**6,468**	**225**	**. .**	**6,693**

Note: Table excludes identified deaths in hospital and discharges to hospital. Preliminary analysis indicated that *Websphere*® passes allowing variation in date of birth with or without variation in event date led to PPVs of under 55%. These passes were therefore excluded.

Table 6.15: BSESLA missed links, using N links as the reference standard

	Hospital separation mode						
RAC event type	To RAC	To other health care establishment	Left against medical advice/ statistical discharge from leave/unknown	Died	To usual residence/other	Total	
Admission	393	170	28	—	217	808	52.2
Hospital leave	69	68	23	4	562	726	46.9
Social leave	1	—	—	—	12	13	0.8
Total	**463**	**238**	**51**	**4**	**791**	**1,547**	**100.0**

Note: Table excludes mixed BSESLA–N links, that is, types D, E and F in Figure 6.2.

Table 6.16: Summary of BSESLA same-match rates compared with refined CSLA match pairs excluding 2-digit postcode matching, by *Websphere*® pass (% BSESLA links)

	Pass			
	1 (exact match)	2 (with 1-sided event date variation)	Total	
1. and 31. Hospital event links to same RAC event or event for same person under CSLA and BSESLA	98.5	76.9	97.8	6,544
32. Hospital event links to different RAC events for different people	—	—	—	2
4. Hospital event links under BSESLA only	1.5	23.1	2.2	147
Total (%)	**100.0**	**100.0**	**100.0**	**. .**
Total (number)	**6,468**	**225**	**. .**	**6,693**

6.5 Summary of E linking

From the above discussion it can be seen that over 95% of the links from all three E linkage strategies were the same as (or close to) links achieved by N linkage (Table 6.17). For both the SLA- and postcode-based constrained linkage approaches, removing match passes that led to unacceptable false match rates improved their PPV by 2 percentage points. However, refining the strategies in this way also reduced the total number of true matches, and consequently reduced marginally the sensitivity—down by around half a percentage point. Both the basic linkage strategy and refined constrained strategies had PPVs of around 98%; consequently 2% of the E links did not correspond to an N link. The constrained E linkages resulted in over 5% more links than the basic unconstrained strategy.

Table 6.17: Comparing E linkage strategies to N linkage

	CSLA			CPC			
	Initial	Refined	Refined, not 2-digit postcode	Initial	Refined	Refined, not 2-digit postcode	BSESLA
Total E matches (A)	7,802	7,595	7,253	7,781	7,587	7,078	6,693
E true match (B)	7,460	7,418	7,100	7,450	7,418	6,936	6,539
Missed N links	646	688	1,006	656	688	1,170	1,567
False E links	342	177	153	331	169	142	154
PPV (B/A%)	95.6	97.7	97.9	95.7	97.8	98.0	97.7
False matches ((A–B)/A%)	4.4	2.3	2.1	4.3	2.2	2.0	2.3
Sensitivity (B/8,106%)	92.0	91.5	87.6	91.9	91.5	85.6	80.7
Relative size (A/8,106%)	96.2	93.7	89.5	96.0	93.4	87.3	82.3

Note: Table is based on comparing the RAC event(s) linked to a particular hospital episode (with true links including links to different RAC events for the same person). For this reason, the number of links are slightly different from those presented in Tables 6.1, 6.4, 6.9, 6.11 and 6.13 which compare all link pairs (as per Figure 6.2).

While 2-digit postcode may be used for matching in the less populous states, it cannot be used either for the larger states or for national analyses because of the increased risk of false matches. Excluding such matching for Western Australia decreased the number of CSLA links by 5% and the number of CPC links by 7%. After excluding 2-digit postcode matching, the PPV increased only marginally—by less than half a percentage point for both CSLA and CPC. However, the effect on the sensitivity of the strategies was greater, dropping by 4 percentage points for CSLA (to 88%) and by 6 percentage points for CPC matching (to 86%). Overall, if 2-digit postcode matching is excluded, using postcode as the basis for regional matching rather than SLA group reduced the number of achieved links (true or false) by 2.4% (7,078 versus 7,253).

The three E linkage strategies resulted in highly overlapping sets of links. However, the differential rates of missed matches for the constrained and basic E linkage strategies may indicate biases in the datasets that could affect analyses. This aspect of the various linkage strategies is examined in the following sections.

7 Match strategy: efficiency comparisons

From the above discussion it can be seen that there are five E linkage strategies that could be used in different situations—depending on the regional distribution of the population under study and the human resources available to undertake the data linkage:

- $CSLA_s$—constrained SLA group E linkage strategy, not including matching within expanded regions defined by the first 2 digits of postcode (termed 2-digit postcode matching)
- $CSLA_f$—constrained SLA group E linkage strategy, including 2-digit postcode matching
- CPC_s—constrained postcode E linkage strategy, not including 2-digit postcode matching
- CPC_f—constrained postcode E linkage strategy, including 2-digit postcode matching
- BSESLA—basic SLA group E linkage strategy.

These five strategies result in very similar positive predictive values but a range of sensitivities (Table 7.1).

Table 7.1: Positive predictive value and sensitivity of E linkage strategies, using N linkage as the reference standard

Match strategy	(a)True links (A)	False links (B)	Missed links (C)	Total links (D = A + B)	PPV (A/D)	Sensitivity (A/F)
	Number				Per cent	
N linkage(b)	(F) 8,106	. .	. .	. .	. .	. .
$CSLA_s$	7,100	153	1,006	7,253	97.9	87.6
$CSLA_f$	7,418	177	688	7,595	97.7	91.5
CPC_s	6,936	142	1,170	7,078	98.0	85.6
CPC_f	7,418	169	688	7,587	97.8	91.5
BSESLA	6,539	154	1,567	6,693	97.7	80.7

(a) E matches to a different event for the same person linked by N linkage are included as true links.

(b) Manual inspection indicates that for a small number of links the event match chosen by the N linkage strategy was not the preferred link. In particular, for 18 matches (0.2% of N links) the preferred link was to a RAC hospital leave event rather than the chosen admission event (for the same person).

If the results from all the E linkage strategies are equally representative of movements of people from hospital to residential aged care, the results in Table 7.1 indicate that if 2-digit postcode matching can be included there is little to choose between the $CSLA_f$ and CPC_f strategies. In this case, since straight postcode matching is much easier both to understand and carry out than SLA group matching, the CPC_f strategy would be preferred over the $CSLA_f$ strategy. However, if 2-digit postcode matching cannot be included, the $CSLA_s$ strategy would be preferred because of its greater sensitivity while having almost the same PPV as CPC_s matching. Because of its relatively poor sensitivity, BSESLA matching would only be used if there were insufficient resources to undertake the more complex constrained matching. Analysis in Section 5 suggests that in this case basic matching using 3-digit postcode could provide similar results without introducing the complication of SLA group matching (see passes 1 and 2 of Table 6.10).

The similar total positive predictive values but varying sensitivities of the E linkage strategies raise the question of whether match dataset differences are in size only, as the consistent PPVs suggest, or whether there are some underlying distributional differences in the matches made under the various strategies. In this section, the representative nature of the linked datasets is examined in terms of the variables used in the E linkage strategies. How any identified differences might impact on possible analyses is examined in detail in Section 8.

7.1 Links to RAC hospital leave

Matches to RAC hospital leave provides an opportunity to compare strategies, as the RAC administrative system has recorded an absence from residential aged care specifically because the resident has gone to hospital. However, it must be remembered that, due to the exclusion of hospital separations ending with a statistical discharge from the E linkage hospital linkage set, all E linkage strategies may miss links to RAC hospital leave that ended with the person being discharged to hospital during a hospital episode reported as a statistical discharge (that is, the person subsequently changed care type within the hospital or was transferred to another hospital).

The constrained strategies linked just over 80% of recorded RAC hospital leave events, with the basic strategy matching somewhat fewer (69%) (Table 7.2). While less than 100% identification was expected for the E linkage strategies, this was not the case for the N linkage strategy which included all hospital separations irrespective of separation mode. Consequently, given that the RAC facility had recorded a visit to hospital, it would be expected that the person-based matching would result in matches for nearly all recorded RAC hospital leave events. Nevertheless, even after identifying links to same-day hospital episodes (153 N links, unable to be identified by E linkage due to exclusion of same-day episodes from the hospital linkage dataset), the N linkage strategy appears to have missed around 290[2] links between hospital episodes and RAC hospital leave—or 4.2% of all RAC hospital leave events. This suggests that some or all of the following may have occurred:

- The N strategy did not identify some people who really did use both RAC and hospital services, possibly due to insufficient or erroneous demographic details. The lack of event dates on the RAC records for WADLS linkage may have contributed to this.
- The RAC or hospital administrative data systems contained errors in event types or event dates preventing appropriate selection of those event records from within a series of linked person records.
- Some records may indicate that a person left an aged care facility to go to hospital but in fact they were never actually admitted to hospital but were treated in an emergency department and so do not appear on the hospital morbidity dataset.

Of these possibilities, the last seems the most likely, with RAC residents arriving at hospital emergency departments and then either dying or being treated and returned to their RAC

---

2 Note: 20 of these were for cases where N linkage linked to the RAC admission just before the RAC hospital leave event (see Table 4.1 note 3).

facility without being admitted. As the WADLS also contains links from 2001 to emergency presentations, it was possible to investigate this further.

To estimate the relative frequency of each of these scenarios, all WADLS-linked records for a sample of 60 of the RAC cases that had a RAC hospital leave episode but no matching hospital admission record were extracted and scrutinised. A close examination of all the WADLS-linked events for emergency, hospital admission and death records for these RAC clients revealed that 16 had a matching emergency record; 10 had a series of connecting hospital events that extended beyond the RAC dates; 16 had a matching hospital admission record that must have been missed during the event selection process (see footnote 2); and 17 were true 'missed links'. As emergency records did not cover the whole state for the entire 12-month study period (beginning in teaching hospitals in 2001), it is likely that a number of the 'unlinked' group were in fact seen at an emergency department without subsequently being admitted into hospital.

Table 7.2: Matches to RAC hospital leave, by linkage strategy (number)

Match strategy	Linked to same-day hospital episode	Discharge to hospital	Death in hospital	Retained matches	Total	Per cent
N linkage[(a)]	153	[(b)]575	464	5,370	6,562	95.8
$CSLA_s$	[(c)] . .	[(d)]137	[(d)]448	5,035	5,620	82.1
$CSLA_f$	[(c)] . .	[(d)]137	[(d)]452	5,078	5,667	82.7
CPC_s	[(c)] . .	[(d)]136	[(d)]451	4,987	5,574	81.4
CPC_f	[(c)] . .	[(d)]136	[(d)]451	5,066	5,653	82.5
BSESLA	[(c)] . .	[(e)] . .	[(e)] . .	4,707	4,707	68.7
Number of events	. .	. .	. .	. .	**[(f)]6,849**	**100.0**

(a) Manual inspection indicates that for a small number of links the event match chosen by the N linkage strategy was not the preferred link. In particular, for 18 matches (0.2% of N links) the preferred link was to a RAC hospital leave event rather than the chosen admission event (for the same person).

(b) Includes links to some hospital episodes (with statistical separations) not available to the AIHW for linkage.

(c) Same-day hospital episodes were excluded from the hospital linkage set used for E linkage.

(d) May include a very small number of RAC leave events (<5) without at least some days in the period 1 July 2000–30 June 2001.

(e) Linkage strategy did not identify these.

(f) RAC hospital leave events ending in the period 1 July 2000–30 June 2001, including deaths in hospital and discharges to hospital.

The above results for RAC hospital leave indicate that there may be some distributional differences between the results of the N and E link strategies. These are examined further below by considering the PPV and sensitivity rates within a number of variables.

7.2 Positive predictive value

RAC event type

RAC events are classified as either a permanent or respite admission or as a hospital or social leave event. These RAC event types were used to define the partitioned datasets for the constrained E linkage strategies (see Table 5.1). In addition, analysis of movements from hospital to residential aged care is likely to be carried out separately for the different event types. Not surprisingly, given the strong similarities between the different E linkage strategies, the PPVs were comparable for the different E linkage strategies within all RAC event types except RAC social leave (Table 7.3). For social leave, the PPV for the BSESLA strategy was significantly below that for the other strategies (99% confidence).[3] This is not surprising given that a single event date was used in the matching and the only expectation was that the hospital separation date was before or equal to the RAC re-entry date. Considering that the reliability of these matches was relatively low and that the number of matches was small, this raises the question of whether social leave should be considered for matching when using the basic linkage strategy.

While all five E linkage strategies considered had similar PPV profiles across RAC event types, within a strategy the PPV varies by RAC event type. In particular, matches to RAC permanent admissions had significantly lower PPVs than those for both respite admissions and hospital leave—around 95% compared with 97–99%. While it is understandable that the PPV would be better for hospital leave than admissions (due to more event date information being used when matching) it is unclear why matches to respite admissions should be more reliable than matches to permanent admissions. It is possible that information on the address of usual residence (postcode) is more reliable for respite admissions because the patients are less likely to be in the process of changing their usual residence.

Table 7.3: Positive predictive value, by RAC event type

Match strategy	Permanent admission	Respite admission	Hospital leave	Social leave	Total
		Number of true matches by strategy			
CSLA$_s$	1,293	691	4,969	147	7,100
CSLA$_f$	1,524	736	5,011	147	7,418
CPC$_s$	1,198	669	4,923	146	6,936
CPC$_f$	1,531	741	5,000	146	7,418
BSESLA	1,124	629	4,638	148	6,539
		Total number of matches by strategy			
CSLA$_s$	1,360	704	5,035	154	7,253
CSLA$_f$	1,607	756	5,078	154	7,595
CPC$_s$	1,261	683	4,987	147	7,078
CPC$_f$	1,612	762	5,066	147	7,587
BSESLA	1,177	643	4,707	166	6,693
N linkage[(a)]	1,723	852	5,370	161	8,106
		PPV (per cent)			
CSLA$_s$	95.1	98.2	98.7	95.5	97.9
CSLA$_f$	94.8	97.4	98.7	95.5	97.7
CPC$_s$	95.0	98.0	98.7	99.3	98.0
CPC$_f$	95.0	97.2	98.7	99.3	97.8
BSESLA	95.5	97.8	98.5	89.2	97.7

(a) Analysis indicated that for a small number of links the event match chosen by the N linkage strategy was not the preferred link. In particular, for 18 matches (0.2% of N links) the preferred link was to a RAC hospital leave event rather than the chosen (earlier) admission event for the same person.

3 Significance of the difference between two PPVs was tested using the normal approximation for testing the difference between two independent proportions. Unless stated otherwise, 95% confidence intervals were used.

The PPV for admissions could be increased for the constrained strategies by allowing only exact event date and date of birth matching, that is, by dropping passes 2 to 5 when matching to admissions in CSLA matching (see Table 6.3), and by adjusting pass 2 and dropping passes 3 to 5 in CPC matching (see Table 6.10). However, this would also result in excluding quite a number of true matches (174 out of 198 matches for $CSLA_s$).

Hospital event types

Hospital event type is identified by both the mode of admission and mode of separation of a hospital episode. These two variables were used to partition the hospital data for constrained E linkage (see Table 5.1). Hospital mode of admission distinguishes between hospital episodes that start with and without a transfer within the hospital system (a statistical versus non-statistical admission). Under all E linkage strategies the reliability of the match varied with the mode of hospital admission, with links to hospital episodes starting with a statistical admission having a lower PPV than those to hospital episodes starting with a non-statistical admission (Table 7.4). Examination of rates within RAC event type suggests that this was caused by the relatively large number of links to RAC admissions for hospital events starting with a statistical admission: under $CSLA_s$, 58% of the links for hospital episodes starting with a statistical admission were to RAC admissions, compared with 21% of those for hospital episodes starting with a non-statistical admission (Table A6.1).

Table 7.4: Positive predictive value, by hospital mode of admission

Match strategy	Statistical	Non-statistical	Total
	Number of true matches by strategy		
$CSLA_s$	1,434	5,666	7,100
$CSLA_f$	1,571	5,847	7,418
CPC_s	1,371	5,565	6,936
CPC_f	1,566	5,852	7,418
BSESLA	1,285	5,254	6,539
	Total number of matches by strategy		
$CSLA_s$	1,489	5,764	7,253
$CSLA_f$	1,639	5,956	7,595
CPC_s	1,423	5,655	7,078
CPC_f	1,632	5,955	7,587
BSESLA	1,329	5,364	6,693
N linkage	1,773	6,333	8,106
	PPV (per cent)		
$CSLA_s$	96.3	98.3	97.9
$CSLA_f$	95.9	98.2	97.7
CPC_s	96.4	98.4	98.0
CPC_f	96.0	98.3	97.8
BSESLA	96.7	98.0	97.7

Hospital mode of separation gives the reported patient destination following the period in hospital. Looking at hospital mode of separation, there were only marginal differences (often not statistically different at the 95% level) between the PPVs for the E linkage strategies (Table 7.5). In addition, the rates for the three main destinations were very close—between 97.2% and 98.4% for all strategies. However, if RAC event type is also considered, there were distinct differences in the positive predictive values (see example for $CSLA_s$ in Table A6.2). In particular, and as already seen in Tables 6.3 and 6.10, at 90.5% the PPV for links between RAC permanent admissions and hospital episodes separating to usual residence was statistically lower than nearly all other PPVs—the exception being for the small number of links between RAC social leave and hospital separations to usual residence (see Table A6.2).

Table 7.5: Positive predictive value, by hospital mode of separation

Match strategy	To RAC	To other health	Left against medical advice/ statistical discharge from leave/unknown	Death	To usual residence/ other	Total
			Number of true matches by strategy			
$CSLA_s$	1,462	797	24	. .	4,817	7,100
$CSLA_f$	1,631	854	24	. .	4,909	7,418
CPC_s	1,392	762	24	. .	4,758	6,936
CPC_f	1,630	851	25	. .	4,912	7,418
BSESLA	1,303	711	6	—	4,519	6,539
			Total number of matches by strategy			
$CSLA_s$	1,502	810	24	. .	4,917	7,253
$CSLA_f$	1,678	870	24	. .	5,023	7,595
CPC_s	1,431	775	25	. .	4,847	7,078
CPC_f	1,676	867	26	. .	5,018	7,587
BSESLA	1,331	723	6	7	4,626	6,693
N linkage	1,773	949	59	4	5,321	8,106
			PPV (per cent)			
$CSLA_s$	97.3	98.4	100.0	. .	98.0	97.9
$CSLA_f$	97.2	98.2	100.0	. .	97.7	97.7
CPC_s	97.3	98.3	96.0	. .	98.2	98.0
CPC_f	97.3	98.2	96.2	. .	97.9	97.8
BSESLA	97.9	98.3	100.0	—	97.7	97.7

Client characteristics (age and sex)

Both date of birth and sex were used in the E linkage strategies to establish links, and these two variables are often used in research analysis. Different PPVs within client characteristics could indicate biases within the matched set that could affect subsequent analyses.

If less reliable data on date of birth are provided for people in some age groups compared with others, then we would expect to see differences in the PPV by age group. The results in Table 7.6 suggest that this could be the case, with the PPVs for 71–80 year olds being slightly below those for other age groups. However, this difference was not always statistically significant at the 95% level; for example, under $CSLA_f$, the same-match rate for people aged

71–80 years (96.8%) was significantly different from that for 81–90 year olds (98.0%, with 95% confidence) but not from that for 65–70 year olds (97.3%).

The difference in the PPVs for different age groups seems to result from the different proportions already in permanent RAC (that is, links to RAC hospital leave) and those newly entering permanent RAC (that is, links to permanent admissions). For example, for the $CSLA_s$ linkage under 65% of links for people aged 80 years or less were for RAC hospital leave compared with over 70% for the older age groups (Table A6.3). Consequently, while the observed PPV for 71–80 year olds entering permanent RAC (92.9%) was statistically different—with over 99% confidence—from that for 91–100 year olds (98%), it was not significantly different from the rates for either the 81–90 (95.6%) or 65–70 (94.7%) year olds, at the 95% confidence level. This suggests that the actual differences across age groups were less pronounced than they appear in the tables.

The PPV for events for men and women were very close, with the largest difference being one-fifth of a percentage point (that is, 0.2)—for the $CSLA_s$ and CPC_s strategies (Table 7.7).

Table 7.6: Positive predictive value, by year of birth (age in 2000)

Match strategy	<1900 (>100 yrs)	1900–09 (91–100 yrs)	1910–19 (81–90 yrs)	1920–29 (71–80 yrs)	1935–39 (65–70 yrs)	Total
			Number of true matches by strategy			
$CSLA_s$	22	1,231	3,571	1,896	380	7,100
$CSLA_f$	22	1,277	3,714	2,007	398	7,418
CPC_s	23	1,206	3,498	1,844	365	6,936
CPC_f	23	1,273	3,719	2,008	395	7,418
BSESLA	18	1,120	3,330	1,730	341	6,539
			Total number of matches by strategy			
$CSLA_s$	22	1,251	3,637	1,954	389	7,253
$CSLA_f$	22	1,299	3,791	2,074	409	7,595
CPC_s	23	1,227	3,564	1,893	371	7,078
CPC_f	23	1,296	3,796	2,069	403	7,587
BSESLA	18	1,144	3,399	1,787	345	6,693
N linkage	28	1,376	4,074	2,188	440	8,106
			PPV (per cent)			
$CSLA_s$	100.0	98.4	98.2	97.0	97.7	97.9
$CSLA_f$	100.0	98.3	98.0	96.8	97.3	97.7
CPC_s	100.0	98.3	98.2	97.4	98.4	98.0
CPC_f	100.0	98.2	98.0	97.1	98.0	97.8
BSESLA	100.0	97.9	98.0	96.8	98.8	97.7

Table 7.7: Positive predictive value, by sex

Match strategy	Males	Females	Total
	Number of true matches by strategy		
$CSLA_s$	2,282	4,818	7,100
$CSLA_f$	2,389	5,029	7,418
CPC_s	2,233	4,703	6,936
CPC_f	2,390	5,028	7,418
BSESLA	2,103	4,436	6,539
	Total number of matches by strategy		
$CSLA_s$	2,328	4,925	7,253
$CSLA_f$	2,447	5,148	7,595
CPC_s	2,276	4,802	7,078
CPC_f	2,447	5,140	7,587
BSESLA	2,154	4,539	6,693
N linkage	2,613	5,493	8,106
	PPV (per cent)		
$CSLA_s$	98.0	97.8	97.9
$CSLA_f$	97.6	97.7	97.7
CPC_s	98.1	97.9	98.0
CPC_f	97.7	97.8	97.8
BSESLA	97.6	97.7	97.7

Note: Sex as recorded on RAC data.

7.3 Sensitivity

While the reliability of matches is important, the rate at which matches are identified is also of consequence; for example, biases in the matched set could result if some types of matches are more likely to be identified than others. The sensitivity of the different strategies is measured by comparing the number of true matches made under a strategy with the total number of possible matches (that is, N matches).

Because in general the five E linkage strategies had very similar PPVs, the relative size of their sensitivities tends to reflect the total number of matches made under each strategy. Consequently, at almost 92% the constrained strategies using 2-digit postcode had significantly higher sensitivities than those excluding such matching (86% for CPC_s and 88% for $CSLA_s$),[4] and all constrained strategies had higher rates than the basic linkage strategy (sensitivity of 81%) (Table 7.1). However, within a strategy there may still be different sensitivities for different sub-populations.

---

4 As all sensitivities are derived using a common denominator, statistical significance between sensitivities for two strategies was tested using a simple confidence interval around either of the rates (normal approximation). Unless stated otherwise, 95% confidence intervals were used.

RAC event type

For all strategies, the sensitivity was lower for admissions than for leave events; furthermore, the sensitivity rates were lower for permanent admissions than for respite admissions when 2-digit postcode matching was excluded (Table 7.8).[5] At 66%, matches made to permanent admissions under BSESLA had the lowest sensitivity within RAC event type. Changing to constrained SLA matching increased this to 75%—significantly above the 70% achieved using postcode-based constrained matching. For respite admissions, the sensitivity for BSESLA matches was 74%, significantly below the 81% and 79% obtained for constrained SLA- and postcode-based matching, respectively (without 2-digit postcode matching).

For Western Australia, allowing 2-digit postcode matching increased the sensitivity of the constrained strategies by 5 to 10 percentage points for respite admissions and by 10 to 20 percentage points for permanent admissions, without greatly affecting the positive predictive value of the links (see Table 7.3). If 2-digit postcode matching can be used, it results in much closer identification rates for permanent and respite admissions than if it has to be excluded. Hospital staff recording the address of the receiving RAC facility as the patient's usual address could explain this effect of using 2-digit postcode, with people tending to move to RAC facilities near to their previous residence. This may have occurred for some establishments as no specific instructions for recording residential address, separation mode and discharge destination were specified for hospitals providing data for the Western Australian Hospital Morbidity Data System until 2002.

Table 7.8: Sensitivity of strategy, by RAC event type (true E matches as % N links)

Match strategy	Permanent admission	Respite admission	Hospital leave	Social leave	Total	
	Sensitivity (per cent)					Number
$CSLA_s$	75.0	81.1	92.5	91.3	87.6	7,100
$CSLA_f$	88.5	86.4	93.3	91.3	91.5	7,418
CPC_s	69.5	78.5	91.7	90.7	85.6	6,936
CPC_f	88.9	87.0	93.1	90.7	91.5	7,418
BSESLA	65.2	73.8	86.4	91.9	80.7	6,539
N linkage (number)	1,723	852	5,370	161	. .	8,106

Source: Table 7.3.

Hospital event types

As expected from the differing PPVs for the two hospital modes of admission, the E linkage strategies were less efficient at identifying links to hospital episodes with a statistical admission compared with those with a non-statistical admission (Table 7.9). This difference was apparent both with and without 2-digit postcode matching, although the gap was smaller when 2-digit postcode matching was used (a gap of around 4 percentage points compared with 10).

5 Significance of the difference between sensitivities for links within different subgroups was tested using the normal approximation for testing the difference between two independent proportions. Unless stated otherwise, 95% confidence intervals were used.

Looking at the type of RAC events linking to hospital episodes with statistical and non-statistical admissions, as with the PPV the difference in linkage sensitivities appears to be caused by the relatively large number of links to RAC admissions for hospital events starting with a statistical admission (Table A6.4). In addition, it is interesting to note that—at least for $CSLA_s$ linkage—while there were differences between the sensitivity of links to permanent and respite admissions within hospital admission mode, there was no significant difference for RAC admissions across hospital mode of admission. That is, the sensitivity of the $CSLA_s$ linkage was the same for links between RAC permanent admissions and hospital episodes with either a statistical or non-statistical admission, with a similar result for respite admissions. On the other hand, there were significant differences between the sensitivity of the $CSLA_s$ linkage strategy for links between RAC hospital leave and hospital episodes with statistical and non-statistical admission. This latter finding most likely reflects the exclusion of hospital episodes with a statistical discharge from the NHMD linkage set for E linkage and the resulting less accurate dates for matching to hospital episodes with a statistical admission—matching on event end date, with the RAC leave start date on or before the hospital episode start date.

Table 7.9: Sensitivity of strategy, by hospital mode of admission (true E matches as % N links)

Match strategy	Statistical	Non-statistical	Total	
	Sensitivity (per cent)			Number
$CSLA_s$	80.9	89.5	87.6	7,100
$CSLA_f$	88.6	92.3	91.5	7,418
CPC_s	77.3	87.9	85.6	6,936
CPC_f	88.3	92.4	91.5	7,418
BSESLA	72.5	83.0	80.7	6,539
N linkage (number)	1,773	6,333	. .	8,106

Source: Table 7.4.

When using 2-digit postcode matching ($CSLA_f$ and CPC_f), the sensitivities for the E linkage strategies were similar for the three main destinations after separation from hospital (90–92% for links to hospital episodes discharged to RAC, to usual residence, and to other health facility) (Table 7.10). However, after excluding such matching the sensitivity for links to hospital episodes separating to usual residence was significantly higher than those for links to episodes with other destinations on discharge. For example, under $CSLA_s$ the sensitivity for people recorded as going to their usual residence was 91% compared with 83% for people reported as entering RAC and 84% for those going to another health facility. The reason for this difference between $CSLA_f$ and $CSLA_s$ is two-fold: first, as expected, hospital episodes with the patient recorded as returning to their usual residence were predominantly linked to RAC hospital leave events, and the E linkage strategies have relatively high PPV and sensitivities for these links (see Table A6.2); second, hospital episodes linking to RAC admissions were more commonly (and correctly) recorded as separating from hospital to go ‘to RAC’ than those linking to RAC leave events (see discussion on CSLA false matches in Section 5), and since 2-digit matching leads proportionally to more links to RAC admissions than leave events, allowing such matching increases the pick-up rate more for hospital separations to RAC. All five E linkage strategies were poor at picking up the small number of links to events identified in the hospital data as a death in hospital (but with a date of death after hospital separation) or left against medical advice.

Table 7.10: Sensitivity of strategy, by hospital mode of separation (true E matches as % N links)

Match strategy	To RAC	To other health care estab.	Left against medical advice/statistical discharge from leave/unknown	Died	To usual residence/ other	Total	
			Sensitivity (per cent)				Number
CSLA$_s$	82.5	84.0	40.7	. .	90.5	87.6	7,100
CSLA$_f$	92.0	90.0	40.7	. .	92.3	91.5	7,418
CPC$_s$	78.5	80.3	40.7	. .	89.4	85.6	6,936
CPC$_f$	91.9	89.7	42.4	. .	92.3	91.5	7,418
BSESLA	73.5	74.9	10.2	—	84.9	80.7	6,539
N linkage (number)	1,773	949	59	4	5,321	. .	8,106

Source: Table 7.5.

Client characteristics (age and sex)

In general, the E linkage strategies tended to have a slightly higher sensitivity for older than younger patients, with the exception of centenarians for whom the sensitivity was relatively low for the small number of links (28) made by N linkage (Table 7.11). For example, the sensitivity of the CSLA$_s$ linkage for 91–100 year olds was 89.5% compared with 86.4% for 65–70 year olds and 78.6% for centenarians. This pattern is less pronounced for strategies that employed 2-digit postcode linking. The reason for this trend could be that older people are more likely to be in RAC, and so their links are more likely than those for younger people to relate to RAC hospital leave than to a RAC admission.

Table 7.11: Sensitivity of strategy, by year of birth (true E matches as % N links)

Match strategy	<1900 (>100 yrs)	1900–09 (91–100 yrs)	1910–19 (81–90 yrs)	1920–29 (71–80 yrs)	1935–39 (65–70 yrs)	Total	
			Sensitivity (per cent)				Number
CSLA$_s$	78.6	89.5	87.7	86.7	86.4	87.6	7,100
CSLA$_f$	78.6	92.8	91.2	91.7	90.5	91.5	7,418
CPC$_s$	82.1	87.6	85.9	84.3	83.0	85.6	6,936
CPC$_f$	82.1	92.5	91.3	91.8	89.8	91.5	7,418
BSESLA	64.3	81.4	81.7	79.1	77.5	80.7	6,539
N linkage (number)	28	1,376	4,074	2,188	440	. .	8,106

Source: Table 7.6.

As when looking only at RAC event type, there were significant differences (95% confidence) between the linkage sensitivities for links to the different RAC event types within age groups; that is, the sensitivity of CSLA$_s$ links to permanent admissions for 81–90 year olds was different from the sensitivity of links to respite admissions for this age group (Table A6.5). However, in general there were not significant differences in the sensitivity of links to particular RAC event types across age groups. For example, the sensitivity of links to permanent admissions for 91–100 year olds was not statistically different from that for 65–70 year olds.

The pick-up rate of links was very similar for men and women (Table 7.12).

Table 7.12: Sensitivity of strategy, by sex (true E matches as % N links)

Match strategy	Males	Females	Total	
	Sensitivity (per cent)			Number
$CSLA_s$	87.3	87.7	87.6	7,100
$CSLA_f$	91.4	91.6	91.5	7,418
CPC_s	85.5	85.6	85.6	6,936
CPC_f	91.5	91.5	91.5	7,418
BSESLA	80.5	80.8	80.7	6,539
N linkage (number)	2,613	5,493	. .	8,106

Source: Table 7.7.

Which N links are missed?

The above discussion indicates that there may be some distributional differences between the E linkage and N linkage matched sets. Whether differential propensities to link within different groups of events are likely to affect analyses using the matched sets was investigated further for a wide range of variables using logistic regression.

Logistic regression was undertaken to examine whether some N links were more likely to be missed than others, in terms of the data available on the characteristics of hospital and RAC events. To limit the number of comparisons made, analysis was restricted to two E linkages: $CSLA_s$ linkage as the preferred E strategy for national analyses, and BSESLA linkage as the strategy with the lowest sensitivity.

Model fitting was carried out using a stepwise forward process (via *SAS®*) but allowing removal of effects at each step. For this investigative analysis, interaction effects were not considered. Model fitting stopped—using a cut-off of 5% significance level for the associated Chi square statistic—when no further effect could be added to the model or if the effect just entered into the model was the only effect removed in the subsequent backward elimination. The dependent variable modelled was whether an N link had been missed as a link under the particular E linkage strategy so that selection of a variable for the model by the fitting process indicated an association between that variable and a propensity to miss links. The 29 independent variables included in the modelling process are listed in Table 7.13. In addition to variables being used as they were recorded in the data collections, a number of binomial variables were derived to allow specific issues to be examined (for example, whether the presence of dementia affected the likelihood of missing N links via E linkage). Also, univariate analysis showed that for some variables only one or two categories seemed to have an extreme sensitivity value, and so further binomial variables were derived to reflect this to see if in fact only the presence or absence of one category was important in determining the propensity to miss a link (see Table A6.6 for sensitivities).

Because of their fundamental differences in the nature of the transition, most analyses of movements from hospital to RAC would logically require separate examination of movements into permanent admissions, respite admissions and RAC leave returns. Given this, and the dominant effect of RAC event type on both the PPV and sensitivity within categories of other variables (as seen above), logistic regressions were fitted within RAC event type. However, because of the small number of links to RAC social leave, RAC social and hospital leave events were combined. In addition, when modelling admissions into RAC, convergence problems led to the exclusion of variables relating to the region of residence of the patient/resident. Note, however, that these variables are very highly associated with the region of the RAC facility.

The variables selected for models fitted for $CSLA_s$ linkage are indicated in Table 7.13. For the model for permanent RAC admissions, days of leave from hospital and region of the RAC facility had the most significant effects, with hospital separation mode also affecting the propensity of $CSLA_s$ linkage to miss N links. N-linked hospital events with more than a week of leave were more likely to be missed than those with no days of leave reported, and N links for people moving to RAC in either the South West or Goldfields regions were less likely to be missed than links for people going into facilities in Perth. For hospital separation mode, the effect centred on cases more likely to have poor destination data, that is, people reported by the hospital as returning to their usual residence, and difficult or unusual cases (20 N links to episodes reported ending as left against medical advice, statistical discharge on leave or unknown reason, or a link to an episode identified in the hospital records as ending with death). N links for these groups were more likely to be missed than those for episodes with the patient reported by the hospital as moving to RAC.

For $CSLA_s$ links to admissions into respite RAC, the stepwise logistic regression process did not include any variables in the model, suggesting that there are unlikely to be any biases resulting from different propensities to identify links in the linked dataset for respite events (Table 7.13).

Two-thirds of the N links were for leave events. The model for the propensity of $CSLA_s$ linkage to miss these events included five variables: use of leave during the hospital episode, remoteness of the receiving RAC facility, hospital separation mode, hospital care type and major diagnostic category (MDC) (Table 7.13). As for the model for permanent RAC admissions, N links to hospital episodes containing a leave period (totalling 49) were more likely to be missed than those that did not. The small number of links for people returning to RAC facilities in remote regions (15 N links) were also more likely to be missed than others, as were the 33 N links to hospital episodes reported with unusual hospital separation modes (left against medical advice, statistical discharge on leave or unknown reason, or a link to an episode identified in the hospital records as ending with death). Among hospital care types, N links to rehabilitation episodes (384 N links) were less likely than those for acute care episodes to be missed by the $CSLA_s$ linkage while those for psychogeriatric care episodes (111 links) were more likely. Finally, links for episodes with their MDC in the category 'Factors influencing health status and other contacts with health services' were more likely to be missed via $CSLA_s$ than those for other episodes. In the Western Australian hospital data, common diagnoses in this group (ICD-10-AM codes starting with Z) included 'Person awaiting admission to adequate facility elsewhere', 'Holiday care relief', 'Healthy person accompanying sick person' and 'Other boarder in health care facility'.

When deriving models for the BSESLA linkage, again the logistic regression fitted an intercept-only model for links to RAC respite admissions (see Table A6.7). For the models for RAC permanent admissions and leave events, more variables were included in the models for the BSESLA links than for the $CSLA_s$ links, including marital status (for permanent admissions) and age (for leave events). Some of the other selected variables related to data quality, while others related to characteristics of the hospital episode.

Table 7.13: Results from logistic regressions for modelling $CSLA_s$-missed link status among N links, within RAC event type, 2000–01

Variable	Model for RAC permanent admissions	Model for RAC respite admissions	Model for RAC leave events
Intercept	**	***	**
Sex	-	-	-
Age at RAC (re-) entry	-	-	-
Marital status (on RAC)	-	-	-
Country of birth (on RAC)	-	-	-
Hospital separation mode	**	-	***
Unusual/unknown hospital separation mode(a)	-	-	-
Hospital care type	-	-	**
Hospital admission mode	-	-	-
Hospital length of stay (categorised)	-	-	-
Number of episodes of leave from hospital (categorised)	-	-	***
Days of leave from hospital (categorised)	***	-	-
Principal diagnosis	-	-	-
Dementia as principal diagnosis(a)	-	-	-
Dementia as any diagnosis(a)	-	-	-
Any diagnoses include external causes(a)	-	-	-
Any diagnoses include 'Factors influencing health status etc.'(a)	-	-	-
Major diagnostic category (MDC)	-	-	-
MDC is 'Factors influencing health status and other contacts with health services'(a)	-	-	***
RAC event type	. .	. .	-
Location of ACAT assessment	-	-	-
RCS appraisal category on RAC (re-) entry	-	-	-
RAC outlet postcode region	***	-	-
RAC outlet in very remote region(a)	-	-	***
RAC facility postcode in Goldfields region(a)	-	-	-
Locality (DoHA) of RAC facility	-	-	-
Person postcode region (in RAC data)	(b)	(b)	-
Poor quality person postcode data(a)	(b)	(b)	-
Person postcode in Goldfields region(a)	(b)	(b)	-
Quality of SLA data in hospital record	-	-	-

(a) Derived binomial variable.

(b) Variable excluded from model due to complete separation of data points. Note that person postcode is highly associated with RAC facility postcode for linked records.

- Not statistically significant at 0.05 level.

* Significant at 0.01–<0.05 level.

** Significant at 0.0001–<0.01 level.

*** Significant at <0.0001 level.

Overall, only a small number of variables were found to influence the propensity of $CSLA_s$ to miss N links. While demographic variables do not appear in the $CSLA_s$ models, there could be some regional and/or hospital episode differences in the profile of N and E links for RAC permanent admissions and RAC leave events. These differences are likely to be greater for BSESLA than for $CSLA_s$ links.

7.4 Summary

The PPV and, in particular, the sensitivity of E links varied with RAC event type. Differences were also apparent for categories within a range of other variables. However, much of this variation can be explained by the RAC event profile of the events within particular categories. When modelling the propensity of E linkage to miss N links within RAC event type, only a small number of variables were selected, with more variables being identified for the less sensitive BSESLA linkage than for $CSLA_s$ linkage.

The dominant effect of RAC event type on the efficiency of E linkage is largely mitigated by the logical requirement of separate examination of movement into permanent admissions, respite admissions and RAC leave returns. Furthermore, given that the E linkage strategies are much more likely to miss N links than make false matches (see Table 7.1), the above analyses indicate that E links provide a good basis for examining the demographic profile of people undertaking various types of transitions. There is some evidence, however, that there could be some regional and/or hospital episode differences in the profile of N and E links for RAC permanent admissions and RAC leave events. Such differences are likely to be greater for basic than for constrained E links.

Noting the above, the question then arises as to whether differences in the profiles of N and E links affect the utility of the E-linked datasets for looking at movements between sectors. When undertaking analysis of transitions, it is the combined effect of missed N links and false E links that determines the overall utility of a linked dataset. The next section explores whether differences between the complete N- and E-linked datasets are—for practical analytic purposes—unimportant, having minimal effect on analyses likely to be of interest when examining the hospital–RAC interface.

8 Match strategy: analysis comparisons

In general, when undertaking analyses of linked data there is no reference standard for the links and so it is not possible to identify the false matches. As a result, all identified matches—both true and false—are used. Consequently, knowledge of biases in the complete match set is important. The above analysis of strategies' PPV, sensitivity and mixed matches indicates that there could be some biases. Whether these differences will affect analyses in practice is examined below, first by considering simple distributions across the match sets and second, by looking at a number of analytical examples.

8.1 Examination of distributional differences

The distributions of linked records from the N linkage and five E linkage strategies were derived for a range of demographic and region variables (Tables 8.1 and 8.2). Visual comparisons between the distributions for the E linkage strategies and the N linkage match set suggest very small distributional differences, with the exception of distributions across RAC event type. For this variable, comparisons show that the basic E linkage and the constrained $CSLA_s$ and CPC_s strategies resulted in relatively few links to RAC admissions, with $CSLA_s$ linkage most closely reflecting the N linkage distribution out of these three. However, allowing 2-digit postcode matching when using constrained linkage strategies increased the percentage of links to admissions, shifting the distribution of links by RAC event type closer to that for the N links.

Table 8.1: Distribution of matches, by age and sex, marital status, country of birth and RAC event type, by linkage strategy (per cent)

		[(a)]N linkage	$CSLA_s$	$CSLA_f$	CPC_s	CPC_f	BSESLA
Sex[(b)]	**Age**[(b)]						
Males	65–69	2.1	2.0	2.1	2.0	2.1	1.9
	70–79	10.1	10.2	10.2	10.2	10.2	10.0
	80–89	14.5	14.3	14.5	14.3	14.5	14.7
	90–99	5.4	5.5	5.4	5.6	5.4	5.5
	≥100	0.1	0.1	0.1	0.1	0.1	0.1
	Total	*32.3*	*32.1*	*32.2*	*32.2*	*32.2*	*32.2*
Females	65–69	1.7	1.7	1.7	1.6	1.7	1.7
	70–79	12.3	12.2	12.4	12.0	12.4	12.0
	80–89	35.4	35.4	35.2	35.7	35.2	35.6
	90–99	17.9	18.1	18.1	18.1	18.1	18.1
	≥100	0.5	0.5	0.5	0.5	0.5	0.5
	Total	*67.8*	*67.9*	*67.8*	*67.8*	*67.8*	*67.8*
	Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Marital status[(b)]	Never married	6.7	6.6	6.6	6.6	6.6	6.6
	De facto	0.3	0.3	0.3	0.3	0.3	0.3
	Married	21.2	21.2	21.3	21.2	21.4	21.7
	Separated	1.6	1.6	1.7	1.7	1.7	1.6
	Divorced	5.0	5.1	5.1	5.0	5.1	5.0
	Widowed	63.4	63.2	63.2	63.3	63.2	62.9
	Unknown	1.9	2.0	1.9	2.0	1.8	1.9
	Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Country of birth[(b)]	Australia	62.7	62.6	62.5	62.5	62.5	62.8
	NZ/Oceania	0.9	0.9	0.9	1.0	0.9	1.0
	UK/Ireland	22.0	22.4	22.3	22.5	22.4	22.5
	Europe	8.9	8.7	8.8	8.6	8.8	8.4
	Asia	3.1	3.1	3.1	3.1	3.1	3.0
	Other/missing	2.4	2.3	2.3	2.4	2.4	2.3
	Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
RAC event type	Permanent admission	21.3	18.8	21.2	17.8	21.3	17.6
	Respite admission	10.5	9.7	10.0	9.7	10.0	9.6
	Hospital leave	66.3	69.4	66.9	70.5	66.8	70.3
	Social leave	2.0	2.1	2.0	2.1	1.9	2.5
	Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Total (number)		**8,106**	**7,253**	**7,595**	**7,078**	**7,587**	**6,693**

(a) Manual inspection indicates that for a small number of links the event match chosen by the N linkage strategy was not the preferred link. In particular, for 18 matches (0.2% of N links) the preferred link was to a RAC hospital leave event rather than the chosen admission event (for the same person).

(b) From RAC data.

Table 8.2: Distribution of matches, by three region variables and linkage strategy

		[a]N linkage	$CSLA_s$	$CSLA_f$	CPC_s	CPC_f	BSESLA
Person postcode region before event[b]	Perth	80.9	80.5	81.1	80.1	81.1	80.8
	South West	8.6	9.0	8.8	9.1	8.8	8.8
	Great Southern	3.7	3.9	3.8	4.0	3.8	4.0
	Goldfields	2.6	2.7	2.6	2.8	2.6	2.5
	Central Coastal/ Murchison	2.2	2.3	2.2	2.4	2.2	2.4
	North	1.6	1.5	1.5	1.5	1.4	1.3
	PO Box	0.1	0.0	0.0	0.0	0.0	0.0
	Not WA	0.2	0.1	0.1	0.1	0.1	0.1
	Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
DoHA locality of RAC facility[b]	Major city	80.7	80.1	80.7	79.8	80.7	80.3
	Inner regional	8.5	8.9	8.7	9.0	8.7	8.9
	Outer regional	8.6	9.0	8.7	9.3	8.8	9.0
	Remote	1.6	1.6	1.6	1.7	1.6	1.5
	Very remote	0.5	0.4	0.4	0.3	0.3	0.3
	Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
RAC facility postcode region[b]	South West	81.5	80.9	81.5	80.5	81.4	81.1
	Great Southern	8.4	8.8	8.5	8.9	8.5	8.6
	Goldfields	3.9	4.0	3.9	4.1	3.9	4.1
	Central Coastal/ Murchison	2.6	2.7	2.6	2.7	2.6	2.5
	North	2.2	2.3	2.2	2.3	2.2	2.4
	South West	1.6	1.4	1.4	1.4	1.4	1.3
	Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Total (number)		**8,106**	**7,253**	**7,595**	**7,078**	**7,587**	**6,693**

(a) Manual inspection indicates that for a small number of links the event match chosen by the N linkage strategy was not the preferred link. In particular, for 18 matches (0.2% of N links) the preferred link was to a RAC hospital leave event rather than the chosen admission event (for the same person).

(b) From RAC data.

While there are standard goodness of fit tests for comparing distributions, there are two limitations in using them in this project. First, as sample sizes increase, the closer the distributions need to be in order not to be statistically significantly different. While this is to be expected (and desirable from the point of view of statistical tests), for large sample sizes many tests (such as the Chi squared test) are, in practice, always statistically significant. Second, when comparing match sets the two distributions being compared are based on highly overlapping datasets, bringing the validity of standard tests into question.

In the current context, we are actually more interested in whether any differences in linkages make any practical difference in terms of interpretation of analyses and application of results. Therefore, two possible tests for gauging practical differences between distributions across categories have been used:

(a) Average absolute relative difference between distributions (for categorical data):

$$R = \frac{\sum_{j=1}^{k} | p_{1j} - p_j | / p_j}{k} \times 100\ \%$$

(b) Average absolute difference between distributions (for categorical data):

$$A = \frac{\sum_{j=1}^{k} | p_{1j} - p_j |}{k}$$

where p_{1j} is the proportion in category j for test linkage 1

p_j is the proportion in category j for the reference linkage

k is the number of categories.

Smaller values of R and A for one match set rather than another (when both are compared against the same reference set) suggest that the former's distribution is closer to the standard than the latter's. The two measures of distributional difference are presented for the E linkage strategies, when compared with the N linkage match set, for a range of variables in Table 8.3.

In general, the average relative differences R for E linkage distributions when compared with that of the N linkage match set were smaller for the strategies that incorporated 2-digit postcode matching than for those that did not, and the constrained strategies performed better on this measure than the basic strategy (with the exception of the sex by age distributions). In addition, using SLA group rather than postcode in the constrained strategies commonly resulted in smaller average relative differences.

When deriving R, all categories are treated equally with respect to the relative difference in the proportion in a category; consequently, small differences in the numbers falling into the rarer categories can have a large effect on R. This effect is seen in the relatively high values of R for the region variables, in particular for the RAC service locality and person postcode variables. For these two variables, small differences in the number of matches for people going to RAC services in very remote regions (for locality) and in matches for people coming from outside Western Australia or using a post office box number for their address resulted in large relative differences in the proportion in the category—relative changes of 30% or more. These large changes in one or two rare categories dominated the corresponding average relative differences.

As for R, when comparing E linkage match set distributions with those for the N linkage, the average absolute distributional difference A was smaller for the strategies that used 2-digit postcode matching than for others; in addition, constrained methods using SLA group performed better than those using postcode. However, the basic strategy BSESLA had slightly smaller average absolute differences than $CSLA_s$ (and CPC_s) for the region measures but not for the other demographic variables examined.

While A is not affected by small differences in the number of matches in rare categories in the same way as R, it is affected by small relative differences in common categories. For example, the relatively small differences in the proportion of matches to RAC hospital leave (a 6% difference, or less, across all E linkage match sets) provided the largest contributions to the average absolute difference for distributions across RAC event type for all E linkage strategies—up to 4 percentage points depending on the strategy.

Table 8.3: Measures of distributional differences compared with N linkage match set distribution, by linkage strategy (per cent per category)

Distribution[(a)]	$CSLA_s$	$CSLA_f$	CPC_s	CPC_f	BSESLA
	Average relative difference *R* (per cent per category)				
Sex by age	3.6	3.6	4.2	3.6	3.4
Marital status	2.2	0.9	2.1	1.3	2.3
Country of birth	1.4	1.2	2.1	1.2	3.0
Person postcode region before event	18.4	17.9	19.7	18.3	22.4
DoHA locality of RAC facility	8.3	7.4	10.5	9.1	10.7
RAC facility postcode region	4.9	2.8	6.6	3.2	6.9
RAC event type	7.7	2.2	8.8	2.0	14.2
	Average absolute difference *A* (percentage points per category)				
Sex by age	0.082	0.074	0.141	0.077	0.138
Marital status	0.073	0.083	0.071	0.103	0.170
Country of birth	0.128	0.113	0.185	0.122	0.212
Person postcode region before event	0.180	0.099	0.268	0.108	0.178
DoHA locality of RAC facility	0.318	0.108	0.454	0.128	0.294
RAC facility postcode region	0.257	0.075	0.375	0.080	0.243
RAC event type	1.653	0.328	2.150	0.263	2.285

(a) Using categories as in Tables 8.1 and 8.2.

In summary, in general the distributions of links across a range of variables were similar for the N and E links. The E link distributions from the strategies that used 2-digit postcode matching tended to match those for N links more closely than those from other strategies. In addition, constrained methods using SLA group performed better in this respect than those using postcode.

8.2 Examination of analytical differences

Many analyses of movement between hospital and RAC will want to compare people who move between the two sectors with those who do not. In this case, both linked and unlinked records are examined. Examples of such analyses are given below. Three broad groups of analysis are considered: movement from hospital, movement into RAC, and an example looking at a specific issue—dementia. In all cases, analysis was carried out taking into account the type of transition into RAC, thereby removing one of the greatest sources of possible bias identified in the analyses in Section 7. Comparisons were limited to the three E linkage strategies $CSLA_s$, CPC_s and BSESLA. The 2-digit postcode E linkage strategies have not been considered as they cannot be used for national analyses.

Movement from hospital

After a period in hospital, a patient may return home (in either RAC or in the community), go to RAC for respite care or to live permanently, or die. A number of the patterns relating to these transitions are examined below.

Destination after hospital

The different sensitivities of the E linkage strategies are reflected in the absolute numbers of people making the various transitions following a period in hospital (Table 8.4). All unlinked hospital separations, other than those ending in death, are classified as going 'To community/other', and as the E linkage strategies result in many more missed than false links, more separations are allocated to this residual category for all the E linkage strategies than for N linkage. Consequently, the E linkage strategies underestimate the number of hospital separations related to RAC entries. Of the three strategies examined, this bias was least for $CSLA_s$ and greatest for BSESLA. Given the high sensitivity of the 2-digit postcode strategies ($CSLA_f$ and CPC_f), the biases for these would be less.

Despite the bias in flow numbers caused by missed links, the proportional distributions of the E linkage strategies across the possible post-hospital destinations were very similar to that for N linkage, but with N linkage having a slightly higher proportion of separations identified as moves from hospital to RAC as either an admission or return from leave: 9.5% compared with 8.5% for $CSLA_s$, 8.4% for CPC_s and 7.9% for BSESLA. As expected from their relative sensitivities (see Table 7.8), the distribution for $CSLA_s$ most closely resembled that for N linkage, with that for BSESLA being the most removed. However, for all linkage strategies the relative sizes of the movement groups were similar. In particular, $CSLA_s$ linkage, like N linkage, indicates that:

- More than twice as many transitions from hospital to RAC are due to people in RAC having episodes of hospitalisation, rather than the result of people being admitted into RAC.
- Around twice as many hospital separations result in movement into permanent RAC than into respite RAC.

Similar results were seen for the other two strategies, although for CPC_s and BSESLA the ratio of moves into respite RAC to those into permanent RAC was slightly higher.

Table 8.4: Destination of people aged 65 years and over, by movement type, age and sex, by linkage strategy,[(a)] separations, 2000–01

Movement type within linkage strategy	Males			Females			
	65–79	80+	All	65–79	80+	All	All
N linkage				**Number**			
Return to permanent RAC[(b)]	635	1,052	1,687	800	3,037	3,837	5,524
To permanent RAC[(b)]	235	363	598	265	837	1,102	1,700
To respite RAC[(b)]	153	152	305	169	375	544	849
To community/other[(c)]	25,753	9,597	35,350	23,803	13,266	37,069	72,419
Died in hospital[(c)]	1,134	961	2,095	772	1,042	1,814	3,909
All	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**84,401**
$CSLA_s$							
Return to permanent RAC[(b)]	594	991	1,585	747	2,839	3,586	5,171
To permanent RAC[(b)]	184	278	462	215	648	863	1,325
To respite RAC[(b)]	130	125	255	134	310	444	699
To community/other[(c)]	25,868	9,769	35,637	23,940	13,716	37,656	73,293
Died in hospital[(c)]	1,134	962	2,096	773	1,044	1,817	3,913
All	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**84,401**
CPC_s							
Return to permanent RAC[(b)]	591	978	1,569	728	2,818	3,546	5,115
To permanent RAC[(b)]	175	264	439	191	598	789	1,228
To respite RAC[(b)]	123	120	243	131	304	435	678
To community/other[(c)]	25,887	9,801	35,688	23,986	13,793	37,779	73,467
Died in hospital[(c)]	1,134	962	2,096	773	1,044	1,817	3,913
All	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**84,401**
BSESLA							
Return to permanent RAC[(b)]	547	946	1,493	700	2,656	3,356	4,849
To permanent RAC[(b)]	151	249	400	180	567	747	1,147
To respite RAC[(b)]	116	119	235	124	279	403	638
To community/other[(c)]	25,963	9,851	35,814	24,033	14,014	38,047	73,861
Died in hospital[(c)]	1,133	960	2,093	772	1,041	1,813	3,906
All	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**84,401**

(continued)

Table 8.4 (continued): Destination of people aged 65 years and over, by movement type, age and sex, by linkage strategy,[a] separations, 2000–01

Movement type within linkage strategy	Males			Females				
	65–79	80+	All	65–79	80+	All	All	N
N linkage				**Per cent**				
Return to permanent RAC[b]	2.3	8.7	4.2	3.1	16.4	8.6	6.5	5,524
To permanent RAC[b]	0.8	3.0	1.5	1.0	4.5	2.5	2.0	1,700
To respite RAC[b]	0.5	1.3	0.8	0.7	2.0	1.2	1.0	849
To community/other[c]	92.3	79.2	88.3	92.2	71.5	83.6	85.8	72,419
Died in hospital[c]	4.1	7.9	5.2	3.0	5.6	4.1	4.6	3,909
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*84,401*
$CSLA_s$								
Return to permanent RAC[b]	2.1	8.2	4.0	2.9	15.3	8.1	6.1	5,171
To permanent RAC[b]	0.7	2.3	1.2	0.8	3.5	1.9	1.6	1,325
To respite RAC[b]	0.5	1.0	0.6	0.5	1.7	1.0	0.8	699
To community/other[c]	92.7	80.6	89.0	92.8	73.9	84.9	86.8	73,293
Died in hospital[c]	4.1	7.9	5.2	3.0	5.6	4.1	4.6	3,913
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*84,401*
CPC_s								
Return to permanent RAC[b]	2.1	8.1	3.9	2.8	15.2	8.0	6.1	5,115
To permanent RAC[b]	0.6	2.2	1.1	0.7	3.2	1.8	1.5	1,228
To respite RAC[b]	0.4	1.0	0.6	0.5	1.6	1.0	0.8	678
To community/other[c]	92.8	80.8	89.1	92.9	74.3	85.2	87.0	73,467
Died in hospital[c]	4.1	7.9	5.2	3.0	5.6	4.1	4.6	3,913
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*84,401*
BSESLA								
Return to permanent RAC[b]	2.0	7.8	3.7	2.7	14.3	7.6	5.7	4,849
To permanent RAC[b]	0.5	2.1	1.0	0.7	3.1	1.7	1.4	1,147
To respite RAC[b]	0.4	1.0	0.6	0.5	1.5	0.9	0.8	638
To community/other[c]	93.0	81.2	89.5	93.1	75.5	85.8	87.5	73,861
Died in hospital[c]	4.1	7.9	5.2	3.0	5.6	4.1	4.6	3,906
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	. .
Total	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**. .**	**84,401**

(a) For all strategies, links to a permanent admission on the same or next day as the end of a period of hospital leave for the same person have been reassigned as linking to the hospital leave. This affected 17 links to permanent admissions for N linkage, 19 links for both $CSLA_s$ and CPC_s linkage and 14 links for BSESLA; in addition, in one case for N linkage and one case for BSESLA linkage this resulted in dropping a different link for the hospital leave episode. However, 18 N links to an earlier RAC admission rather than the preferred hospital leave event have not been reassigned (as per Table 4.1).

(b) Based on linked hospital and residential aged care records. Same-day and next day re-admissions to permanent RAC are treated as transfers and so have been combined into a single period of care when identifying returns to RAC after hospital leave. Links to RAC hospital and social leave are both classified as returns to RAC.

(c) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

Note: Age is as at time of hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals.

Destination patterns for particular subgroups (defined by age, sex and region) were compared using results from the E and N linkages. Looking across age and sex, the E linkage strategies showed similar patterns of use when compared with the N linkage results. For example, both strategies showed that:

- Fewer people aged 80 years and over returned to the community than people aged 65 to 79 years, with this difference being much more marked for women than men.
- Admission into RAC, both permanently and for respite care, was more common for older people than younger people, and was generally higher for women than men.

While the general patterns were the same, some of the small differences noticeable under N linkage between specific age-by-sex categories within the smaller movement groups were not apparent in the E linkages, indicating that when using E linkage, detailed statements about such groups should be avoided.

In the previous section, region was identified as one distribution possibly affected by bias in the E linkage results. The post-hospital destination of people by their region prior to hospitalisation is presented in Table 8.5. Once again, the E linkage and N linkage strategies had similar patterns:

- People living in Perth prior to hospitalisation are more likely to already be RAC residents than those in other regions.
- Entry into permanent RAC following hospitalisation is less likely for people from the Central Coastal/Murchison and North regions than those from other regions.
- People from Perth or the South West region are less likely to go into residential respite care than those from elsewhere.

Taking a different perspective and looking at hospital care type within destination group, the distribution of hospital care types within post-hospital destination was similar for all linkage strategies (Table 8.8), although the use of rehabilitation among hospitalised RAC residents was slightly underestimated by the E linkage strategies, as was the use of psychogeriatric care among people moving between the care sectors. These differences were generally less for $CSLA_s$ than for the other two E strategies.

Table 8.5: Destination of people aged 65 years and over, by movement type and postcode region at hospital admission, by linkage strategy, separations, 2000–01

Movement type within linkage strategy	Perth	South West	Great Southern	Goldfields	Central Coast./ Murch-ison	North	PO box/not WA/ unknown	All
N linkage				**Number**				
Return to permanent RAC[(a)]	4,561	450	187	133	119	71	3	5,524
To permanent RAC[(a)]	1,438	141	49	32	22	13	5	1,700
To respite RAC[(a)]	547	96	73	45	37	49	2	849
To community/other[(b)]	50,897	9,287	4,349	2,393	2,696	2,239	558	72,419
Died in hospital[(b)]	2,903	461	208	121	131	74	11	3,909
All	**60,346**	**10,435**	**4,866**	**2,724**	**3,005**	**2,446**	**579**	**84,401**
CSLA$_s$								
Return to permanent RAC[(a)]	4,264	432	174	127	115	59	—	5,171
To permanent RAC[(a)]	1,092	130	44	29	19	10	1	1,325
To respite RAC[(a)]	450	79	64	38	35	32	1	699
To community/other[(b)]	51,633	9,333	4,376	2,409	2,705	2,271	566	73,293
Died in hospital[(b)]	2,907	461	208	121	131	74	11	3,913
All	**60,346**	**10,435**	**4,866**	**2,724**	**3,005**	**2,446**	**579**	**84,401**
CPC$_s$								
Return to permanent RAC[(a)]	4,212	431	174	127	115	56	—	5,115
To permanent RAC[(a)]	1,001	123	44	29	18	10	3	1,228
To respite RAC[(a)]	427	79	67	37	35	32	1	678
To community/other[(b)]	51,799	9,341	4,373	2,410	2,706	2,274	564	73,467
Died in hospital[(b)]	2,907	461	208	121	131	74	11	3,913
All	**60,346**	**10,435**	**4,866**	**2,724**	**3,005**	**2,446**	**579**	**84,401**
BSESLA								
Return to permanent RAC[(a)]	4,029	389	168	102	110	51	—	4,849
To permanent RAC[(a)]	944	113	40	28	17	5	—	1,147
To respite RAC[(a)]	398	77	62	37	36	28	—	638
To community/other[(b)]	52,072	9,395	4,390	2,437	2,711	2,288	568	73,861
Died in hospital[(b)]	2,903	461	206	120	131	74	11	3,906
All	**60,346**	**10,435**	**4,866**	**2,724**	**3,005**	**2,446**	**579**	**84,401**

(continued)

Table 8.5 (continued): Destination of people aged 65 years and over to hospital, by movement type and postcode region at hospital admission, by linkage strategy, separations, 2000–01

Movement type within linkage strategy	Perth	South West	Great South-ern	Gold-fields	Central Coast./ Murch-ison	North	PO box/not WA/ unk.	All	
N linkage					**Per cent**				
Return to permanent RAC[(a)]	7.6	4.3	3.8	4.9	4.0	2.9	0.5	6.5	5,524
To permanent RAC[(a)]	2.4	1.4	1.0	1.2	0.7	0.5	0.9	2.0	1,700
To respite RAC[(a)]	0.9	0.9	1.5	1.7	1.2	2.0	0.3	1.0	849
To community/other[(b)]	84.3	89.0	89.4	87.8	89.7	91.5	96.4	85.8	72,419
Died in hospital[(b)]	4.8	4.4	4.3	4.4	4.4	3.0	1.9	4.6	3,909
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*84,401*
$CSLA_s$									
Return to permanent RAC[(a)]	7.1	4.1	3.6	4.7	3.8	2.4	—	6.1	5,171
To permanent RAC[(a)]	1.8	1.2	0.9	1.1	0.6	0.4	0.2	1.6	1,325
To respite RAC[(a)]	0.7	0.8	1.3	1.4	1.2	1.3	0.2	0.8	699
To community/other[(b)]	85.6	89.4	89.9	88.4	90.0	92.8	97.8	86.8	73,293
Died in hospital[(b)]	4.8	4.4	4.3	4.4	4.4	3.0	1.9	4.6	3,913
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*84,401*
CPC_s									
Return to permanent RAC[(a)]	7.0	4.1	3.6	4.7	3.8	2.3	—	6.1	5,115
To permanent RAC[(a)]	1.7	1.2	0.9	1.1	0.6	0.4	0.5	1.5	1,228
To respite RAC[(a)]	0.7	0.8	1.4	1.4	1.2	1.3	0.2	0.8	678
To community/other[(b)]	85.8	89.5	89.9	88.5	90.0	93.0	97.4	87.0	73,467
Died in hospital[(b)]	4.8	4.4	4.3	4.4	4.4	3.0	1.9	4.6	3,913
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*84,401*
BSESLA									
Return to permanent RAC[(a)]	6.7	3.7	3.5	3.7	3.7	2.1	—	5.7	4,849
To permanent RAC[(a)]	1.6	1.1	0.8	1.0	0.6	0.2	—	1.4	1,147
To respite RAC[(a)]	0.7	0.7	1.3	1.4	1.2	1.1	—	0.8	638
To community/other[(b)]	86.3	90.0	90.2	89.5	90.2	93.5	98.1	87.5	73,861
Died in hospital[(b)]	4.8	4.4	4.2	4.4	4.4	3.0	1.9	4.6	3,906
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*. .*
Total	**60,346**	**10,435**	**4,866**	**2,724**	**3,005**	**2,446**	**579**	**. .**	**84,401**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

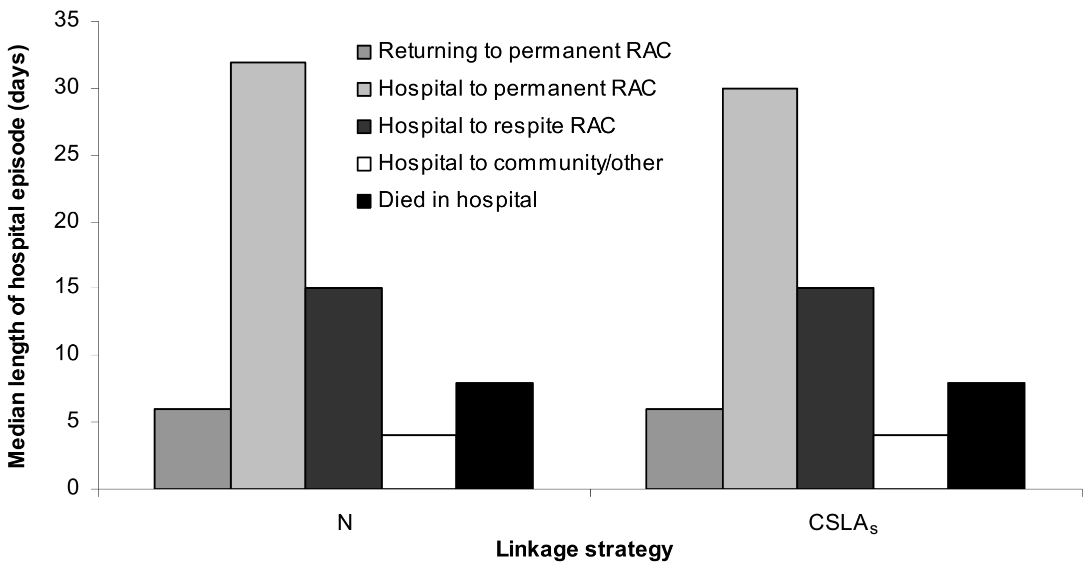
(b) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

Note: Age is as at time of hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals.

Length of stay in hospital

Another area of interest is length of stay in hospital, and whether this varies depending on the destination of the patient at the end of their period in hospital.[6] Comparing median length of stay by post-hospital destination for the N and E linkage strategies shows that all strategies gave very similar results:

- People who return to the community have the shortest median hospital stay, followed by those who are returning to permanent RAC (Figure 8.1, Table 8.6).
- People moving into permanent RAC have much longer median stays than other people—seven to eight times as long as those who go back to the community and twice as long as those who go into respite RAC.
- Median length of stay does not change with age and sex among those returning to RAC, but there is some variation among the other movement groups.
- Median length of stay varies with the region of the patient prior to their hospitalisation (Table 8.7).
- Within post-hospital destination, length of stay varies with the type of care received during the hospital episode (Table 8.9): for all strategies rehabilitation, palliation and psychogeriatric care patients had higher median stays than others.



Source: Table 8.6.

Figure 8.1: Median length of stay, by post-hospital destination for N and $CSLA_s$ linkage strategies

When comparing median length of stay, it needs to be remembered that distribution percentiles for small numbers of cases are sensitive to individual values, particularly when the variable of interest can vary widely (see Tables 8.7 and 8.9). Therefore, as with all

---

6 Length of stay refers only to the selected hospital event and so is not necessarily the total time spent in hospital.

analyses, the statistical accuracy of estimates needs to be considered before making statements about differences.

Table 8.6: Hospital separations for people aged 65 years and over: median length of stay, by movement type, age at separation and sex, by linkage strategy, 2000–01

	Males			Females			Total	
Movement type	**65–79**	**80+**	**All**	**65–79**	**80+**	**All**	**All**	**Number**
Returning to permanent RAC[(a)]				**Median (days)**				
N linkage	6	6	6	6	6	6	6	5,524
$CSLA_s$	5.5	6	6	6	6	6	6	5,171
CPC_s	6	6	6	6	6	6	6	5,115
BSESLA	5	6	6	6	6	6	6	4,849
Hospital to permanent RAC[(a)]								
N linkage	36	32	33	34	30	31	32	1,700
$CSLA_s$	33	32	32	34	28	29	30	1,325
CPC_s	33	30	31	34	28	29	29	1,228
BSESLA	36	29	32	34	27	29	30	1,147
Hospital to respite RAC[(a)]								
N linkage	17	15	16	12	16	14.5	15	849
$CSLA_s$	18	15	16	13	15	15	15	699
CPC_s	18	15	16	13	15	14	15	678
BSESLA	16	15	15	13	15	14	15	638
Hospital to community/other[(b)]								
N linkage	3	5	4	4	5	4	4	72,419
$CSLA_s$	3	5	4	4	5	4	4	73,293
CPC_s	3	5	4	4	5	4	4	73,467
BSESLA	3	5	4	4	5	4	4	73,861
Died in hospital[(b)]								
N linkage	7	8	8	8	8	8	8	3,909
$CSLA_s$	7	8	8	8	8	8	8	3,913
CPC_s	7	8	8	8	8	8	8	3,913
BSESLA	7	8	8	8	8	8	8	3,906
All separations	**4**	**5**	**4**	**4**	**6**	**5**	**4**	**. .**
All separations (number)	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**. .**	**84,401**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

Note: Age is as at time of hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals. Length of stay excludes days on leave from hospital.

Table 8.7: Hospital separations for people aged 65 years and over: median length of stay, by movement type and postcode region at admission to hospital, by linkage strategy, 2000–01

Movement type	Perth	South West	Great Southern	Goldfields	Central Coastal/ Murch-ison	North	PO box/ not WA/ unk.	All	Number
Returning to permanent RAC(a)				**Median (days)**					
N linkage	6	5	4	4	6	4	n.p.	6	5,524
$CSLA_s$	6	5	4	4	7	4	n.p.	6	5,171
CPC_s	6	5	4	4	7	4	n.p.	6	5,115
BSESLA	6	5	4	4	6	4	n.p.	6	4,849
Hospital to permanent RAC(a)									
N linkage	33	23	17	17	*28.5	*12	n.p.	32	1,700
$CSLA_s$	32	25	18.5	16	*39	*11	n.p.	30	1,325
CPC_s	31	25	18.5	16	*44.5	*9	n.p.	29	1,228
BSESLA	32	25	18.5	16	*34	n.p.	n.p.	30	1,147
Hospital to respite RAC(a)									
N linkage	17	15.5	12	8	*17	*7	n.p.	15	849
$CSLA_s$	17	15	12	9	*18	*6.5	n.p.	15	699
CPC_s	16	15	12	9	*18	*6.5	n.p.	15	678
BSESLA	16.5	15	12	9	*17.5	*6.5	n.p.	15	638
Hospital to community(b)									
N linkage	4	4	4	4	4	3	3	4	72,419
$CSLA_s$	4	4	4	4	4	3	3	4	73,293
CPC_s	4	4	4	4	4	3	3	4	73,467
BSESLA	4	4	4	4	4	3	3	4	73,861
Died in hospital(b)									
N linkage	8	7	6	8	7	7	*4	8	3,909
$CSLA_s$	8	7	6	8	7	7	*4	8	3,913
CPC_s	8	7	6	8	7	7	*4	8	3,913
BSESLA	8	7	6	8.5	7	7	*4	8	3,906
All separations	5	4	4	4	4	3	3	4	. .
All separations (number)	**60,346**	**10,435**	**4,866**	**2,724**	**3,005**	**2,446**	**579**	**. .**	**84,401**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

n.p. Not publishable as based on fewer than 10 cases (see Table 8.5).

* Based on 50 or fewer records (see Table 8.5).

Note: Age is as at time of hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals. Length of stay excludes days on leave from hospital.

Table 8.8: Hospital separations for people aged 65 years and over, by movement type and type of care, by linkage strategy, 2000–01

Movement type	Acute care	Rehabili-tation	Palliative care	Psycho-geriatric/ other	Non-acute care	Total	
Returning to permanent RAC[(a)]			**Per cent**			**Number**	
N linkage	90.0	7.1	0.4	2.0	0.5	100.0	5,524
CSLAs	90.8	6.6	0.4	1.7	0.5	100.0	5,171
CPCs	90.8	6.6	0.4	1.7	0.5	100.0	5,115
BSESLA	90.7	6.8	0.3	1.7	0.4	100.0	4,849
Hospital to permanent RAC[(a)]							
N linkage	47.9	20.7	1.4	6.5	23.5	100.0	1,700
CSLAs	47.5	21.5	1.4	5.1	24.4	100.0	1,325
CPCs	47.6	21.7	1.3	4.8	24.7	100.0	1,228
BSESLA	48.7	21.1	1.5	4.9	23.8	100.0	1,147
Hospital to respite RAC[(a)]							
N linkage	68.2	17.6	1.9	2.4	10.0	100.0	849
CSLAs	67.2	18.5	1.4	2.0	10.9	100.0	699
CPCs	67.8	17.8	1.5	1.9	10.9	100.0	678
BSESLA	67.1	18.5	1.4	1.7	11.3	100.0	638
Hospital to community/other[(b)]							
N linkage	93.4	3.3	0.8	0.4	2.2	100.0	72,419
CSLAs	93.1	3.4	0.8	0.5	2.3	100.0	73,293
CPCs	93.0	3.4	0.8	0.5	2.3	100.0	73,467
BSESLA	92.9	3.5	0.8	0.5	2.4	100.0	73,861
Died in hospital[(b)]							
N linkage	71.9	2.8	22.3	0.6	2.3	100.0	3,909
CSLAs	71.9	2.8	22.3	0.6	2.3	100.0	3,913
CPCs	71.9	2.8	22.3	0.6	2.3	100.0	3,913
BSESLA	71.9	2.8	22.3	0.6	2.3	100.0	3,906
All separations	**91.0**	**4.0**	**1.8**	**0.6**	**2.6**	**100.0**	**. .**
All separations (number)	**76,819**	**3,358**	**1,491**	**528**	**2,205**	**. .**	**84,401**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

Note: Age is as at time of hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals.

Table 8.9: Hospital separations for people aged 65 years and over: median length of stay, by movement type and type of care, by linkage strategy, 2000–01

Movement type	Acute care	Rehabili-tation	Palliative care	Psycho-geriatric/ other	Non-acute care	Total	
Returning to permanent RAC[(a)]			**Median (days)**				**Number**
N linkage	5	16	*7	28	*7.5	6	5,524
CSLA$_s$	5	16	*6.5	27	*6.5	6	5,171
CPC$_s$	5	16	*6.5	26.5	*6.5	6	5,115
BSESLA	5	16	*6.5	26	*6.5	6	4,849
Hospital to permanent RAC[(a)]							
N linkage	27	37	*39	90	27	32	1,700
CSLA$_s$	26	38	*40	90.5	25	30	1,325
CPC$_s$	25.5	37	*36	78	24	29	1,228
BSESLA	26	36	*43	92.5	26	30	1,147
Hospital to respite RAC[(a)]							
N linkage	12	24	*24	*31.5	17	15	849
CSLA$_s$	12	24	*24.5	*28	18	15	699
CPC$_s$	12	24	*24.5	*27	17.5	15	678
BSESLA	12	24	*23	*29	17.5	15	638
Hospital to community/other[(b)]							
N linkage	4	15	9	26	3	4	72,419
CSLA$_s$	4	16	9	30	3	4	73,293
CPC$_s$	4	16	9	30	3	4	73,467
BSESLA	4	16	9	30	3	4	73,861
Died in hospital[(b)]							
N linkage	7	23	7	*44	30	8	3,909
CSLA$_s$	7	23	7	*44	30	8	3,913
CPC$_s$	7	23	7	*44	30	8	3,913
BSESLA	7	22	7	*44	30	8	3,906
All separations	**4**	**17**	**8**	**32**	**5**	**4**	. .
All separations (number)	**76,819**	**3,358**	**1,491**	**528**	**2,205**	. .	**84,401**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

* Based on 50 or fewer records (see Table 8.8).

Note: Age is as at time of hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals. Length of stay excludes days on leave from hospital.

Diagnoses

Diagnosis is a central element in the hospital data and it is important to know if the E linked data reflect the underlying distribution across diagnoses within movement groups. While there are some minor differences in the distributions, as would be expected over so many categories, the general look of the distributions is similar (Table 8.10). For example, looking at the distribution of cases within movement groups by principal diagnosis (categorised into 19 groups), in all but one case the E linkage strategies identified the same diagnoses as the N linkage strategy as the top three reasons for hospitalisation:

- Among people returning to RAC, cardiovascular disease, diseases of the respiratory system, and injuries and poisoning were the three most common principal diagnosis chapters identified by all linkage strategies.
- Among people entering permanent RAC, factors influencing health status, awaiting admission elsewhere and mental disorders were identified as the three most common principal diagnosis chapters by all linkage strategies except BSESLA. For BSESLA, slightly more of these linked cases related to cardiovascular disease than mental disorders.
- For all strategies, factors influencing health status was the most common principal diagnosis chapter among patients moving to respite RAC, followed by cardiovascular disease and injury and poisoning.
- For people who return to the community, cardiovascular disease, neoplasms and diseases and disorders of the digestive system were identified as the most common principal diagnosis chapters by all strategies.

Table 8.10: Hospital separations, by principal diagnosis and movement type, by linkage strategy, 2000–01 (per cent)

Principal diagnosis chapter (ICD-10-AM)	N linkage	$CSLA_s$	CPC_s	BSESLA
Returning to permanent RAC[(a)]				
Certain infectious & parasitic diseases (A00–B99)	1.4	1.4	1.4	1.4
Neoplasms (C00–D48)	5.4	5.6	5.6	5.4
Blood & blood-forming organs and immunological disorders (D50–D89)	2.2	2.4	2.4	2.4
Endocrine, nutritional and metabolic diseases(E00–E90)	2.8	2.8	2.8	2.8
Mental and behavioural disorders (F00–F99)	3.6	3.3	3.3	3.3
Diseases of the nervous system & sense organs (G00–G99)	3.5	3.5	3.5	3.3
Diseases of the eye and adnexa (H00–H59)	2.1	2.2	2.2	2.1
Diseases of the ear and mastoid process (H60–H95)	0.2	0.2	0.2	0.2
Cardiovascular disease (I00–I99)	14.6	14.9	14.9	15.0
Diseases of the respiratory system (J00–J99)	13.0	12.9	13.0	13.2
Diseases of the digestive system (K00–K93)	9.3	9.6	9.6	9.6
Diseases of the skin & subcutaneous tissue (L00–L99)	3.2	3.2	3.2	3.2
Diseases of the musculoskeletal system and connective tissue (M00–M99)	4.1	4.0	3.9	3.8
Diseases of the genitourinary system (N00–N99)	6.2	6.3	6.3	6.3
Congenital malformations (Q00–Q99)	—	—	—	—
Symptoms, sign & ill-defined conditions (R00–R99)	7.3	7.2	7.2	7.2
Injury & poisoning (S00–T98)	12.5	12.8	12.9	12.8
Awaiting admission to adequate facility elsewhere (Z75.1)	0.3	0.3	0.3	0.2
Factors influencing health status and contact with health services (other Z codes)	8.1	7.5	7.5	7.8
Total	*100.0*	*100.0*	*100.0*	*100.0*
Total (number)	*5,524*	*5,171*	*5,115*	*4,849*
Hospital to permanent RAC[(a)]				
Certain infectious & parasitic diseases (A00–B99)	0.6	0.7	0.7	0.5
Neoplasms (C00–D48)	3.5	3.4	3.1	3.2
Blood & blood-forming organs and immunological disorders (D50–D89)	0.3	0.4	0.4	0.4
Endocrine, nutritional and metabolic diseases(E00–E90)	1.5	1.9	1.9	1.9
Mental and behavioural disorders (F00–F99)	11.6	11.3	11.2	10.9
Diseases of the nervous system & sense organs (G00–G99)	7.6	6.9	6.9	7.1
Diseases of the eye and adnexa (H00–H59)	0.2	0.1	0.1	0.1
Diseases of the ear and mastoid process (H60–H95)	—	—	—	—
Cardiovascular disease (I00–I99)	10.6	10.8	11.1	11.9
Diseases of the respiratory system (J00–J99)	4.2	4.4	4.2	4.3
Diseases of the digestive system (K00–K93)	2.0	1.6	1.7	1.7
Diseases of the skin & subcutaneous tissue (L00–L99)	0.9	1.0	1.0	1.0
Diseases of the musculoskeletal system and connective tissue (M00–M99)	3.0	3.2	3.3	3.2
Diseases of the genitourinary system (N00–N99)	3.1	3.0	2.8	3.1
Congenital malformations (Q00–Q99)	—	—	—	—
Symptoms, sign & ill-defined conditions (R00–R99)	4.4	4.8	4.8	4.7
Injury & poisoning (S00–T98)	6.6	6.1	6.2	6.2
Awaiting admission to adequate facility elsewhere (Z75.1)	14.6	15.1	14.7	14.6
Factors influencing health status and contact with health services (other Z codes)	25.2	25.4	26.0	25.2
Total	*100.0*	*100.0*	*100.0*	*100.0*
Total (number)	*1,700*	*1,325*	*1,228*	*1,147*

(continued)

Table 8.10 (continued): Hospital separations, by principal diagnosis and movement type, by linkage strategy, 2000–01 (per cent)

Principal diagnosis chapter (ICD-10-AM)	N linkage	$CSLA_s$	CPC_s	BSESLA
Hospital to respite RAC[(a)]				
Certain infectious & parasitic diseases (A00–B99)	1.1	1.0	1.0	0.8
Neoplasms (C00–D48)	5.7	5.6	5.6	5.5
Blood & blood-forming organs and immunological disorders (D50–D89)	0.2	0.3	0.4	0.3
Endocrine, nutritional and metabolic diseases(E00–E90)	2.2	2.6	2.4	2.4
Mental and behavioural disorders (F00–F99)	7.8	7.4	7.4	7.4
Diseases of the nervous system & sense organs (G00–G99)	6.4	6.6	6.8	6.0
Diseases of the eye and adnexa (H00–H59)	0.5	0.3	0.3	0.5
Diseases of the ear and mastoid process (H60–H95)	0.2	0.3	0.3	0.2
Cardiovascular disease (I00–I99)	10.7	11.0	11.2	11.0
Diseases of the respiratory system (J00–J99)	7.2	7.2	7.4	7.5
Diseases of the digestive system (K00–K93)	2.2	1.7	1.8	2.2
Diseases of the skin & subcutaneous tissue (L00–L99)	1.9	2.0	1.9	2.0
Diseases of the musculoskeletal system and connective tissue (M00–M99)	4.8	5.0	5.2	5.3
Diseases of the genitourinary system (N00–N99)	3.4	2.9	2.8	2.8
Congenital malformations (Q00–Q99)	—	—	—	—
Symptoms, sign & ill-defined conditions (R00–R99)	7.1	6.7	6.8	6.4
Injury & poisoning (S00–T98)	10.1	9.9	9.9	10.0
Awaiting admission to adequate facility elsewhere (Z75.1)	5.1	5.6	5.5	5.6
Factors influencing health status and contact with health services (other Z codes)	23.4	24.0	23.5	24.1
Total	*100.0*	*100.0*	*100.0*	*100.0*
Total (number)	*849*	*699*	*678*	*638*
Hospital to community/other[(b)]				
Certain infectious & parasitic diseases (A00–B99)	1.1	1.1	1.1	1.1
Neoplasms (C00–D48)	10.8	10.7	10.7	10.7
Blood & blood-forming organs and immunological disorders (D50–D89)	1.2	1.2	1.2	1.2
Endocrine, nutritional and metabolic diseases(E00–E90)	2.5	2.5	2.5	2.5
Mental and behavioural disorders (F00–F99)	1.8	1.9	1.9	2.0
Diseases of the nervous system & sense organs (G00–G99)	2.3	2.3	2.3	2.4
Diseases of the eye and adnexa (H00–H59)	3.6	3.6	3.6	3.6
Diseases of the ear and mastoid process (H60–H95)	0.5	0.5	0.5	0.5
Cardiovascular disease (I00–I99)	18.3	18.2	18.2	18.1
Diseases of the respiratory system (J00–J99)	8.4	8.4	8.4	8.4
Diseases of the digestive system (K00–K93)	10.5	10.4	10.4	10.4
Diseases of the skin & subcutaneous tissue (L00–L99)	1.7	1.7	1.7	1.7
Diseases of the musculoskeletal system and connective tissue (M00–M99)	10.0	9.9	9.9	9.8
Diseases of the genitourinary system (N00–N99)	6.3	6.2	6.2	6.2
Congenital malformations (Q00–Q99)	0.1	0.1	0.1	0.1
Symptoms, sign & ill-defined conditions (R00–R99)	6.9	6.8	6.8	6.8
Injury & poisoning (S00–T98)	7.0	7.1	7.1	7.1
Awaiting admission to adequate facility elsewhere (Z75.1)	0.2	0.2	0.3	0.3
Factors influencing health status and contact with health services (other Z codes)	7.0	7.1	7.2	7.2
Total	*100.0*	*100.0*	*100.0*	*100.0*
Total (number)	*72,419*	*73,293*	*73,467*	*73,861*

(continued)

Table 8.10 (continued): Hospital separations, by principal diagnosis and movement type, by linkage strategy, 2000–01 (per cent)

Principal diagnosis chapter (ICD-10-AM)	N linkage	$CSLA_s$	CPC_s	BSESLA
Died in hospital[(b)]				
Certain infectious & parasitic diseases (A00–B99)	2.9	2.9	2.9	2.9
Neoplasms (C00–D48)	29.2	29.2	29.2	29.3
Blood & blood-forming organs and immunological disorders (D50–D89)	0.8	0.8	0.8	0.8
Endocrine, nutritional and metabolic diseases(E00–E90)	2.1	2.1	2.1	2.2
Mental and behavioural disorders (F00–F99)	1.3	1.3	1.3	1.3
Diseases of the nervous system & sense organs (G00–G99)	1.8	1.8	1.8	1.8
Diseases of the eye and adnexa (H00–H59)	—	—	—	—
Diseases of the ear and mastoid process (H60–H95)	—	—	—	—
Cardiovascular disease (I00–I99)	24.1	24.0	24.0	24.1
Diseases of the respiratory system (J00–J99)	14.0	14.0	14.0	14.0
Diseases of the digestive system (K00–K93)	6.7	6.6	6.6	6.6
Diseases of the skin & subcutaneous tissue (L00–L99)	0.9	0.9	0.9	0.9
Diseases of the musculoskeletal system and connective tissue (M00–M99)	1.5	1.5	1.5	1.5
Diseases of the genitourinary system (N00–N99)	3.6	3.6	3.6	3.6
Congenital malformations (Q00–Q99)	0.1	0.1	0.1	0.1
Symptoms, sign & ill-defined conditions (R00–R99)	1.8	1.8	1.8	1.8
Injury & poisoning (S00–T98)	4.7	4.7	4.7	4.7
Awaiting admission to adequate facility elsewhere (Z75.1)	0.7	0.7	0.7	0.7
Factors influencing health status and contact with health services (other Z codes)	3.8	3.8	3.8	3.7
Total	*100.0*	*100.0*	*100.0*	*100.0*
Total (number)	*3,909*	*3,913*	*3,913*	*3,906*

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked hospital separations. Deaths are based on reported hospital mode of separation.

Notes

1. ICD-10-AM Ed. 1 is the international statistical classification of diseases and related health problems, 10th revision, Australian modification, first edition (NCCH 1998).
2. Age is as at time of hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals.

Movement into RAC

Movement into RAC can be the result of either people returning after a period of leave or due to a new admission or transfer into either permanent or respite RAC. Examples looking at these various groups are given below.

People returning to RAC from hospital leave

All four linkage strategies, including N linkage, apparently understate the number of residents who return to RAC after visiting hospital (Table 8.11). However, some of this discrepancy was caused by people going from RAC to hospital but either being recorded as same-day episodes in the hospital system (and so not included in any of the data linkage strategies) or never actually being admitted (see Section 7). This tendency not to link very short episodes of RAC hospital leave is reflected in the smaller proportions of leave of length 1 day identified among the leave returns by all linkage strategies compared with those identified as returns to RAC using just the RAC data. Overall, the distributions across length of leave were similar for all four linkage strategies, although N linkage underestimated the proportion of returns for single overnight stays slightly more than the E linkage strategies and overestimated the proportion of long periods of hospitalisation (over 2 weeks).

Table 8.11: Returns to permanent RAC after hospital leave for people aged 65 years and over, by length of leave, by linkage strategy,[(a)] 2000–01 (per cent)

Length of leave (days)	[(a)]RAC hospital leave identified through RAC data as ending with:			[(b)]Returns to RAC identified by strategy:			
	Death in hospital	Discharge to hospital	Return to RAC	N linkage	$CSLA_s$	CPC_s	BSESLA
1	14.5	0.7	12.7	9.7	10.1	10.2	9.9
2	9.7	1.8	9.1	8.9	9.3	9.4	9.1
3	8.3	1.8	9.7	9.7	10.2	10.2	10.1
4–6	18.5	6.7	21.4	21.9	22.5	22.5	22.4
7–13	25.2	14.4	24.1	25.1	25.0	25.0	25.1
14–20	10.3	16.4	9.9	10.5	10.2	10.1	10.6
≥21	13.5	58.3	13.1	14.1	12.7	12.7	12.8
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**496**	**611**	**5,717**	**5,354**	**5,008**	**4,961**	**4,674**
Per cent identified	. .	. .	**100.0**	**93.7**	**87.6**	**86.8**	**81.8**

(a) Classification is based on RAC data only. If the permanent RAC care period and hospital leave end on the same day (after allowing for transfers between RAC facilities), the leave period is classified as ending with death in hospital if the reason for discharge from RAC is recorded as 'death'; otherwise it is classified as a discharge to hospital. Other periods of hospital leave are assumed to end with the person returning to RAC. Note that the few cases where people transfer between permanent and respite residential care at the end of the hospital leave will be identified as being discharged to hospital.

(b) Based on linked hospital and RAC records. See also note (a) to Table 8.4 for additional information.

Note: Age is at time of return to RAC.

Source of admissions into RAC

Admissions into RAC may be either from the community or via hospital (Table 8.12). As when estimating the volume of hospital separations related to RAC entries, the inability of E linkage to identify a proportion of related hospital and RAC events means that these strategies underestimate the proportion of RAC entries coming from hospital. All unlinked records are necessarily classified as coming from the community.

The disparity between the N and E linkage distributions across source of admission is particularly noticeable for permanent admissions (Table 8.12), and was caused by the relatively low sensitivities of the E strategies when matching to permanent RAC admissions (see Table 7.8). Even so, all four strategies gave very similar distributions across age and sex within transition group (Table 8.13).

Table 8.12: Source of people aged 65 years and over admitted into RAC, by movement type, age and sex, by linkage strategy, RAC entry events, 2000–01

Movement type within linkage strategy	Males			Females				
	65–79	80+	All	65–79	80+	All	All	N
N linkage				Per cent of admissions				
From hospital to permanent RAC[(a)]	20.6	22.8	21.9	19.9	21.4	21.0	21.3	1,699
From hospital to respite RAC[(a)]	13.5	9.5	11.2	13.2	9.5	10.4	10.7	849
Transfer from permanent RAC[(b)]	16.7	19.5	18.3	20.0	22.2	21.7	20.5	1,636
Transfer from respite RAC[(b)]	1.5	1.3	1.4	1.2	1.0	1.1	1.2	92
From community into permanent RAC[(b)]	11.4	16.4	14.3	16.1	19.7	18.8	17.3	1,377
From community into respite RAC[(b)]	36.3	30.6	33.0	29.6	26.2	27.0	29.1	2,315
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*7,968*
CSLA$_s$								
From hospital to permanent RAC[(a)]	16.1	17.5	16.9	16.4	16.5	16.5	16.6	1,325
From hospital to respite RAC[(a)]	11.4	7.9	9.3	10.5	7.8	8.5	8.8	699
Transfer from permanent RAC[(b)]	16.8	19.6	18.5	19.9	22.3	21.7	20.6	1,639
Transfer from respite RAC[(b)]	1.5	1.3	1.4	1.3	1.0	1.1	1.2	94
From community into permanent RAC[(b)]	15.7	21.6	19.1	19.7	24.6	23.4	21.9	1,748
From community/other into respite RAC[(b)]	38.4	32.2	34.8	32.2	27.8	28.9	30.9	2,463
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*7,968*
CPC$_s$								
From hospital to permanent RAC[(a)]	15.4	16.6	16.1	14.5	15.2	15.1	15.4	1,228
From hospital to respite RAC[(a)]	10.7	7.6	8.9	10.2	7.7	8.3	8.5	678
Transfer from permanent RAC[(b)]	16.8	19.7	18.5	20.1	22.3	21.7	20.6	1,643
Transfer from respite RAC[(b)]	1.5	1.3	1.4	1.3	1.0	1.1	1.2	94
From community into permanent RAC[(b)]	16.5	22.4	19.9	21.4	25.9	24.8	23.1	1,841
From community into respite RAC[(b)]	39.1	32.5	35.2	32.4	28.0	29.1	31.2	2,484
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*7,968*
BSESLA								
Return to RAC from hospital[(a)]	13.2	15.7	14.7	13.7	14.5	14.3	14.4	1,147
From hospital to respite RAC[(a)]	10.2	7.5	8.6	9.7	7.0	7.7	8.0	638
Transfer from permanent RAC[(b)]	16.9	19.7	18.5	19.9	22.2	21.7	20.6	1,641
Transfer from respite RAC[(b)]	1.5	1.3	1.4	1.4	1.0	1.1	1.2	95
From community into permanent RAC[(b)]	18.6	23.3	21.3	22.4	26.7	25.6	24.1	1,924
From community into respite RAC[(b)]	39.6	32.6	35.5	32.9	28.6	29.6	31.7	2,523
All	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*7,968*
Total	**1,140**	**1,590**	**2,730**	**1,289**	**3,949**	**5,238**	**7,968**	**7,968**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked RAC events.

Note: Age is as at (re-) entry into RAC. Table excludes unlinked RAC hospital leave events.

As a consequence of lower sensitivities when matching to permanent RAC admissions, the relative proportions of permanent and respite admissions were biased towards respite admissions in the E linkage distributions. Because of the greater sensitivity of $CSLA_s$, these biases were less for this strategy than for CPC_s and BSESLA. Nevertheless, looking at patterns—rather than absolute flows—for admissions into RAC (Table 8.12), both N and E linkage exhibited the same patterns (in relative terms):

- More respite admissions into RAC come from the community for men than for women, while the reverse is true for permanent admissions.
- RAC transfer admissions from permanent care are more common for women than men.
- Among those admitted, older people are less likely to have been admitted for respite care than younger people, either from hospital or from the community.

Table 8.13: Age and sex distribution of people aged 65 years and over entering RAC, within movement type, by linkage strategy, RAC entry events, 2000–01

Movement type within linkage strategy	Males			Females			All	
	65–79	80+	All	65–79	80+	All	All	
N linkage				Per cent				Number
Return to RAC from hospital[(a)]	11.4	19.1	30.5	14.7	54.8	69.5	100.0	5,523
Return to RAC from social leave[(b)]	12.1	11.6	23.7	21.5	54.8	76.3	100.0	4,482
From hospital to permanent RAC[(a)]	13.8	21.3	35.1	15.1	49.7	64.9	100.0	1,699
From hospital to respite RAC[(a)]	18.1	17.8	35.9	20.0	44.1	64.1	100.0	849
Transfer from permanent RAC[(b)]	11.6	18.9	30.6	15.8	53.7	69.4	100.0	1,636
Transfer from respite RAC[(b)]	18.5	21.7	40.2	17.4	42.4	59.8	100.0	92
From community into permanent RAC[(b)]	9.4	19.0	28.4	15.0	56.6	71.6	100.0	1,377
From community into respite RAC[(b)]	17.9	21.0	38.9	16.5	44.7	61.1	100.0	2,315
All	**12.9**	**17.6**	**30.5**	**17.1**	**52.5**	**69.5**	**100.0**	**. .**
All (number)	**2,314**	**3,165**	**5,479**	**3,065**	**9,429**	**12,494**	**. .**	**17,973**
$CSLA_s$								
Return to RAC from hospital[(a)]	11.4	19.2	30.6	14.6	54.8	69.4	100.0	5,169
Return to RAC from social leave[(b)]	12.1	11.7	23.7	21.5	54.8	76.3	100.0	4,489
From hospital to permanent RAC[(a)]	13.9	21.0	34.9	16.0	49.1	65.1	100.0	1,325
From hospital to respite RAC[(a)]	18.6	17.9	36.5	19.3	44.2	63.5	100.0	699
Transfer from permanent RAC[(b)]	11.7	19.0	30.8	15.6	53.6	69.2	100.0	1,639
Transfer from respite RAC[(b)]	18.1	21.3	39.4	18.1	42.6	60.6	100.0	94
From community into permanent RAC[(b)]	10.2	19.6	29.9	14.5	55.6	70.1	100.0	1,748
From community into respite RAC[(b)]	17.8	20.8	38.6	16.8	44.6	61.4	100.0	2,463
All	**12.9**	**17.6**	**30.5**	**17.1**	**52.4**	**69.5**	**100.0**	**. .**
All (number)	**2,273**	**3,106**	**5,379**	**3,008**	**9,239**	**12,247**	**. .**	**17,626**

(continued)

Table 8.13 (continued): Age and sex distribution of people aged 65 years and over entering RAC, within movement type, by linkage strategy, RAC entry events, 2000–01

Movement type within linkage strategy	Males			Females			All	
	65–79	80+	All	65–79	80+	All		
N linkage				**Per cent**				**Number**
CPC$_s$								
Return to RAC from hospital[(a)]	11.5	19.2	30.6	14.4	55.0	69.4	100.0	5,113
Return to RAC from social leave[(b)]	12.1	11.6	23.7	21.5	54.8	76.3	100.0	4,496
From hospital to permanent RAC[(a)]	14.3	21.5	35.7	15.2	49.0	64.3	100.0	1,228
From hospital to respite RAC[(a)]	18.0	17.8	35.8	19.5	44.7	64.2	100.0	678
Transfer from permanent RAC[(b)]	11.7	19.1	30.7	15.8	53.5	69.3	100.0	1,643
Transfer from respite RAC[(b)]	18.1	21.3	39.4	18.1	42.6	60.6	100.0	94
From community into permanent RAC[(b)]	10.2	19.3	29.5	15.0	55.5	70.5	100.0	1,841
From community into respite RAC[(b)]	18.0	20.8	38.7	16.8	44.4	61.3	100.0	2,484
All	**12.9**	**17.6**	**30.5**	**17.0**	**52.5**	**69.5**	**100.0**	**. .**
All (number)	**2,270**	**3,093**	**5,363**	**2,993**	**9,221**	**12,214**	**. .**	**17,577**
BSESLA								
Return to RAC from hospital[(a)]	11.3	19.5	30.8	14.5	54.7	69.2	100.0	4,848
Return to RAC from social leave[(b)]	12.1	11.6	23.7	21.4	54.8	76.3	100.0	4,477
From hospital to permanent RAC[(a)]	13.1	21.8	34.9	15.3	49.8	65.1	100.0	1,147
From hospital to respite RAC[(a)]	18.2	18.7	36.8	19.6	43.6	63.2	100.0	638
Transfer from permanent RAC[(b)]	11.8	19.1	30.8	15.7	53.5	69.2	100.0	1,641
Transfer from respite RAC[(b)]	17.9	21.1	38.9	18.9	42.1	61.1	100.0	95
From community into permanent RAC[(b)]	11.0	19.2	30.2	15.0	54.7	69.8	100.0	1,924
From community into respite RAC[(b)]	17.9	20.5	38.4	16.8	44.7	61.6	100.0	2,523
All	**12.9**	**17.7**	**30.6**	**17.1**	**52.4**	**69.4**	**100.0**	**. .**
All (number)	**2,229**	**3,056**	**5,285**	**2,952**	**9,056**	**12,008**	**. .**	**17,293**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked RAC events.

Notes

1. Age is as at (re-) entry into RAC. Table excludes unlinked RAC hospital leave events.
2. Constituent parts may not sum to total due to rounding.

Survival after admission into permanent residential aged care

While the volume of admissions may be underestimated when using E linkage, the linked datasets may still provide a valuable source of information about other aspects of the hospital–RAC interface. One matter of interest is how long people remain in permanent RAC once they have been admitted. Examination of this issue is illustrated in Table 8.14 using data from the four linkage strategies. All four strategies showed that:

- A much greater proportion of people admitted from hospital die within 6 months of admission than those admitted from the community.
- Older men coming from hospital are more likely to die within 6 months of admission than younger men; however, such an age differential was not observed for women.
- Among people admitted from the community, older men are again more likely to die within 6 months than younger men, with some evidence that the reverse is true for women.

The differences between people admitted from the community and those admitted from hospital are less marked for the E linkage strategies than for N linkage. However, because of its greater sensitivity, this dampening effect of unidentified links in the 'from community' group is less for $CSLA_s$ than the other two E strategies.

It should also be noted that, for all strategies, these estimated death rates within age and sex groups are based on quite small numbers, and so have quite high relative standard errors. Consequently, some apparent differences may not be statistically significant. For example, Table 8.14 shows that around 100 men aged 65–79 years had their first admission from hospital into permanent RAC in the first half of 2000–01; of these, 6.1% had died by the end of the year. If this were the underlying death rate for 65–79 year old men admitted into RAC, for 100 such men admitted in any 6-month period we would expect to observe a death rate of between 1.4% and 10.8% (95% confidence interval, see Table A6.8).

Table 8.14: First permanent admission into RAC: proportion who died in the RAC service within 6 months of admission, by movement type, age and sex, by linkage strategy, first non-transfer permanent admissions 1 July–31 December 2000

Movement type within linkage strategy	Males			Females			All	
	65–79	80+	All	65–79	80+	All		
N linkage				**Number**				
From hospital[(a)]	122	152	274	125	279	404		678
From community[(b)]	98	208	306	163	549	712		1,018
All	**220**	**360**	**580**	**288**	**828**	**1,116**		**1,696**
From hospital[(a)]			**Per cent within age/sex group**					**Number**
N linkage	27.9	32.9	30.7	19.2	20.4	20.0	24.3	678
CSLA$_s$	28.3	37.0	33.0	20.4	20.7	20.6	25.9	514
CPC$_s$	26.1	38.1	32.7	19.1	20.2	19.9	25.3	482
BSESLA	31.3	37.4	34.7	19.0	19.4	19.3	25.8	454
From community[(b)]								
N linkage	6.1	14.9	12.1	8.0	5.8	6.3	8.1	1,018
CSLA$_s$	9.9	15.4	13.5	8.9	7.6	7.9	9.6	1,182
CPC$_s$	12.5	15.4	14.4	10.1	8.0	8.5	10.3	1,214
BSESLA	10.2	16.2	14.1	10.3	8.3	8.8	10.5	1,242
All	**18.2**	**22.5**	**20.9**	**12.8**	**10.7**	**11.3**	**14.6**	**. .**
Total	**220**	**360**	**580**	**288**	**828**	**1,116**	**. .**	**1,696**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked residential aged care records. For E linkage in particular these include missed links to permanent admissions (see discussion on sensitivity—Table 7.8).

Notes

1. Death rates do not include people who left the RAC prior to death, for example, those who died at home after leaving RAC or who died in hospital after discharge from RAC and who were recorded as discharged from RAC to hospital (see Table 7.2 for information on number of such discharges).
2. In a small number of cases, people move from permanent to respite care and then back to permanent care. The later re-admission into permanent care is included in this table.
3. Age is as at entry into RAC.

An example of examining an issue: dementia

Linked data can be used to examine particular issues of interest. One such issue is dementia: as the population ages the effects of the increasing numbers of people with dementia on the hospital and aged care systems become more and more important.

The pattern of prevalence of dementia as the principal diagnosis in the hospital separations for the various movement groups was similar for all linkage strategies (Table 8.15):

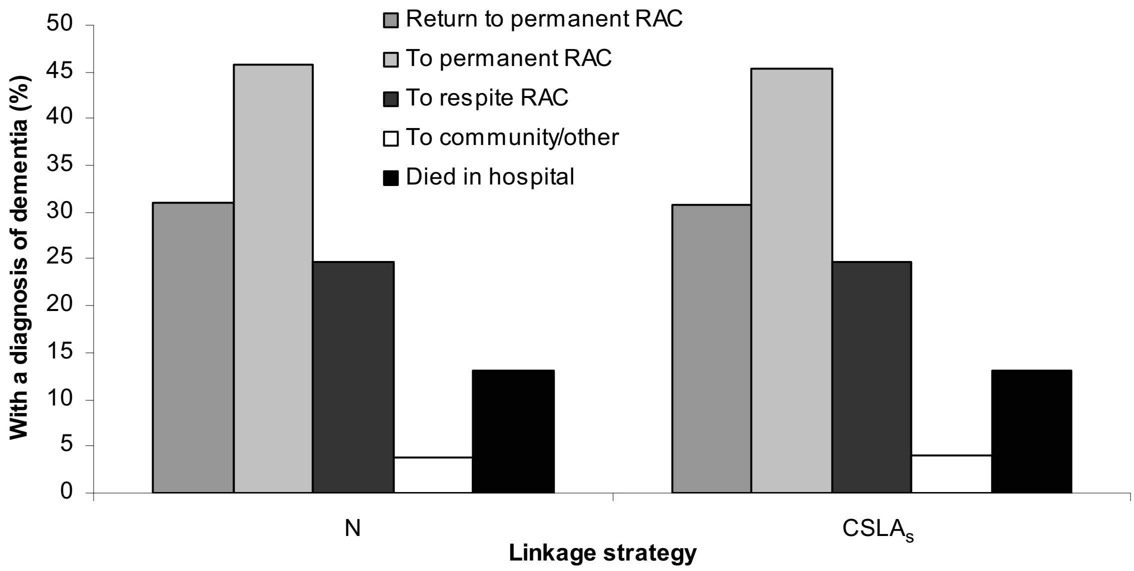
- People being admitted into permanent RAC via hospital have a much higher prevalence of dementia as the principal diagnosis than those returning to RAC.
- The prevalence tends to be greater among people being admitted permanently into RAC than those being admitted for respite care.
- Prevalence is very low among those returning to the community.

There also seem to be some differences by age and/or sex, but many of these were not statistically significant for any of the linkage strategies; for example, the rates of dementia for men aged 65–79 years and those aged 80 years and over who move into permanent RAC—14.9% versus 11.8% for N linkage—are not statistically significant (see Table A6.8).

Looking at dementia as either the principal diagnosis or as an additional diagnosis contributing to hospital cost (Table 8.16), the story is quite different. However, the N and E strategies again revealed the same story:

- Dementia is more commonly an additional than principal diagnosis.
- The prevalence of dementia is much higher among those groups going to RAC than among people going back to the community (Figure 8.2).
- The prevalence of dementia among people returning to RAC is between those for people being admitted permanently and those entering for respite care.

Again, there seem to be some differences by age and sex, but these may not be statistically significant (for example, the difference between rates of dementia for women aged 65–79 years and those aged 80 years and over who move into permanent RAC—44.5% versus 49.7% for N linkage—is not statistically significant).



Source: Table 8.15.

Figure 8.2: Prevalence of dementia diagnosis, by post-hospital destination for N and CSLA$_s$ linkage strategies

From the RAC perspective, all the linkage strategies indicated the same care needs profiles for the different origins of RAC entries (Table 8.17):

- Irrespective of movement group, people with dementia are more likely to have high care needs than others.
- The care needs differential between people with and without dementia varies considerably across the movement types, with the greatest absolute difference being for people returning to RAC after hospital leave.
- People returning to RAC following social leave have a similar care needs profile as those entering for respite care.
- People entering permanent RAC from hospital have the greatest proportion with high care needs.

Table 8.15: Hospital separations: proportion with a principal diagnosis of dementia, by movement type, age and sex, by linkage strategy, 2000–01

Movement type within linkage strategy	Males			Females				
	65–79	80+	All	65–79	80+	All	All	N
N linkage	**Per cent with principal diagnosis in hospital of dementia (within age/sex group)**							
Return to permanent RAC[(a)]	2.7	2.7	2.7	2.4	2.1	2.2	2.3	5,524
To permanent RAC[(a)]	14.9	11.8	13.0	15.8	11.4	12.4	12.6	1,700
To respite RAC[(a)]	8.5	8.6	8.5	3.6	6.7	5.7	6.7	849
To community/other[(b)]	0.3	0.8	0.4	0.3	1.0	0.6	0.5	72,419
Died in hospital[(b)]	0.7	2.0	1.3	0.8	2.1	1.5	1.4	3,909
$CSLA_s$								
Return to permanent RAC[(a)]	2.5	2.4	2.5	2.0	1.9	2.0	2.1	5,171
To permanent RAC[(a)]	13.6	9.7	11.3	16.7	10.5	12.1	11.8	1,325
To respite RAC[(a)]	8.5	8.8	8.6	4.5	6.8	6.1	7.0	699
To community/other[(b)]	0.4	1.0	0.5	0.4	1.2	0.7	0.6	73,293
Died in hospital[(b)]	0.7	2.0	1.3	0.8	2.1	1.5	1.4	3,913
CPC_s								
Return to permanent RAC[(a)]	2.5	2.4	2.4	2.1	2.0	2.0	2.1	5,115
To permanent RAC[(a)]	14.3	10.2	11.8	15.7	10.5	11.8	11.8	1,228
To respite RAC[(a)]	8.9	9.2	9.1	4.6	6.9	6.2	7.2	678
To community/other[(b)]	0.4	1.0	0.5	0.4	1.3	0.7	0.6	73,467
Died in hospital[(b)]	0.7	2.0	1.3	0.8	2.1	1.5	1.4	3,913
BSESLA								
Return to permanent RAC[(a)]	2.4	2.4	2.4	1.9	2.0	1.9	2.1	4,849
To permanent RAC[(a)]	11.9	10.4	11.0	16.1	10.1	11.5	11.3	1,147
To respite RAC[(a)]	8.6	9.2	8.9	3.2	6.8	5.7	6.9	638
To community/other[(b)]	0.4	1.0	0.6	0.4	1.3	0.8	0.7	73,861
Died in hospital[(b)]	0.7	2.0	1.3	0.8	2.0	1.5	1.4	3,906
Total	**0.5**	**1.5**	**0.8**	**0.6**	**1.8**	**1.1**	**1.0**	**. .**
Total (number)	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**. .**	**84,401**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information, and for numbers of cases contributing to cells.

(b) Unlinked hospital records.

Note: Age is as at hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals.

Table 8.16: Hospital separations: proportion with any diagnosis of dementia, by movement type, age and sex, by linkage strategy, 2000–01

Movement type within linkage strategy	Males			Females				
	65–79	80+	All	65–79	80+	All	All	N
N linkage	**Per cent with any hospital diagnosis of dementia (within age/sex group)**							
Return to permanent RAC[(a)]	25.5	33.8	30.7	23.5	33.2	31.2	31.0	5,524
To permanent RAC[(a)]	39.1	41.6	40.6	44.5	49.7	48.5	45.7	1,700
To respite RAC[(a)]	24.2	23.0	23.6	20.1	27.5	25.2	24.6	849
To community/other[(b)]	1.7	6.2	2.9	2.1	8.5	4.4	3.7	72,419
Died in hospital[(b)]	6.2	17.9	11.6	6.1	21.2	14.8	13.0	3,909
CSLA$_s$								
Return to permanent RAC[(a)]	25.6	33.9	30.8	22.5	32.8	30.6	30.7	5,171
To permanent RAC[(a)]	39.7	42.1	41.1	43.7	49.1	47.7	45.4	1,325
To respite RAC[(a)]	26.2	22.4	24.3	20.1	27.1	25.0	24.7	699
To community/other[(b)]	1.8	6.7	3.2	2.3	9.6	4.9	4.1	73,293
Died in hospital[(b)]	6.2	17.9	11.5	6.1	21.2	14.7	13.0	3,913
CPC$_s$								
Return to permanent RAC[(a)]	25.9	33.6	30.7	22.5	32.8	30.7	30.7	5,115
To permanent RAC[(a)]	41.1	42.8	42.1	42.4	48.3	46.9	45.2	1,228
To respite RAC[(a)]	26.0	22.5	24.3	20.6	26.6	24.8	24.6	678
To community/other[(b)]	1.8	6.8	3.2	2.3	9.8	5.1	4.2	73,467
Died in hospital[(b)]	6.2	17.9	11.5	6.1	21.2	14.7	13.0	3,913
BSESLA								
Return to permanent RAC[(a)]	25.2	33.6	30.5	21.6	33.2	30.8	30.7	4,849
To permanent RAC[(a)]	39.1	41.0	40.3	41.1	48.1	46.5	44.3	1,147
To respite RAC[(a)]	24.1	23.5	23.8	19.4	25.8	23.8	23.8	638
To community/other[(b)]	2.0	7.0	3.3	2.4	10.2	5.3	4.3	73,861
Died in hospital[(b)]	6.2	17.8	11.5	6.1	21.1	14.7	13.0	3,906
Total	**2.9**	**10.8**	**5.3**	**3.4**	**15.5**	**8.5**	**6.9**	**. .**
Total (number)	**27,910**	**12,125**	**40,035**	**25,809**	**18,557**	**44,366**	**. .**	**84,401**

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information, and for numbers of cases contributing to cells.

(b) Unlinked hospital records.

Note: Age is as at hospital admission. Table excludes same-day hospital episodes, statistical discharges and transfers to other hospitals.

Table 8.17: Residential aged care entries: movement type, by dementia type and care level, by linkage strategy, 2000–01

	Care level			Total		Care level			Total	
	High	Low	All			High	Low	All		
		Row %		Col. %	N		Row %		Col. %	N
Return from hospital[(a)]			N linkage					CSLA$_s$		
With dementia	64.5	35.5	100.0	31.0	1,711	64.9	35.1	100.0	30.7	1,583
Without dementia	33.6	66.4	100.0	69.0	3,802	33.3	66.7	100.0	69.3	3,575
All	*43.2*	*56.8*	*100.0*	*100.0*	*5,513*	*43.0*	*57.0*	*100.0*	*100.0*	*5,158*
Return from social leave[(b)]	22.3	77.7	100.0	. .	4,462	22.3	77.7	100.0	. .	4,469
Into permanent RAC from hospital[(a)]										
With dementia	83.1	16.9	100.0	45.9	769	83.9	16.1	100.0	45.9	598
Without dementia	72.6	27.4	100.0	54.1	905	74.5	25.5	100.0	54.1	705
All	*77.4*	*22.6*	*100.0*	*100.0*	*1,674*	*78.8*	*21.2*	*100.0*	*100.0*	*1,303*
Into respite RAC from hospital[(a)]										
With dementia	32.5	67.5	100.0	24.6	209	32.4	67.6	100.0	24.7	173
Without dementia	19.1	80.9	100.0	75.4	640	19.6	80.4	100.0	75.3	526
All	*22.4*	*77.6*	*100.0*	*100.0*	*849*	*22.7*	*77.3*	*100.0*	*100.0*	*699*
Transfer into permanent RAC[(b)]	60.1	39.9	100.0	. .	1,625	59.9	40.1	100.0	. .	1,627
Into permanent RAC from community[(b)]	39.4	60.6	100.0	. .	1,351	46.7	53.3	100.0	. .	1,720
Transfer into respite RAC[(b)]	34.8	65.2	100.0	. .	92	35.1	64.9	100.0	. .	94
Into respite RAC from community[(b)]	26.2	73.8	100.0	. .	2,315	25.9	74.1	100.0	. .	2,463
Return from hospital[(a)]			CPC$_s$					BSESLA		
With dementia	65.1	34.9	100.0	30.7	1,567	65.5	34.5	100.0	30.7	1,486
Without dementia	33.4	66.6	100.0	69.3	3,535	34.1	65.9	100.0	69.3	3,352
All	*43.1*	*56.9*	*100.0*	*100.0*	*5,102*	*43.7*	*56.3*	*100.0*	*100.0*	*4,838*
Return from social leave[(b)]	22.3	77.7	100.0	. .	4,476	22.2	77.8	100.0	. .	4,457
Into permanent RAC from hospital[(a)]										
With dementia	84.2	15.8	100.0	45.7	551	84.3	15.7	100.0	44.7	504
Without dementia	74.5	25.5	100.0	54.3	655	74.2	25.8	100.0	55.3	623
All	*78.9*	*21.1*	*100.0*	*100.0*	*1,206*	*78.7*	*21.3*	*100.0*	*100.0*	*1,127*
Into respite RAC from hospital[(a)]										
With dementia	32.9	67.1	100.0	24.6	167	33.6	66.4	100.0	23.8	152
Without dementia	19.8	80.2	100.0	75.4	511	19.3	80.7	100.0	76.2	486
All	*23.0*	*77.0*	*100.0*	*100.0*	*678*	*22.7*	*77.3*	*100.0*	*100.0*	*638*
Transfer into permanent RAC[(b)]	60.0	40.0	100.0	. .	1,631	60.1	39.9	100.0	. .	1,629
Into permanent RAC from community[(b)]	48.2	51.8	100.0	. .	1,813	49.6	50.4	100.0	. .	1,894
Transfer into respite RAC[(b)]	35.1	64.9	100.0	. .	94	35.8	64.2	100.0	. .	95
Into respite RAC from community[(b)]	25.8	74.2	100.0	. .	2,484	25.8	74.2	100.0	. .	2,523

(a) Based on linked hospital and residential aged care records. See also notes (a) and (b) to Table 8.4 for additional information.

(b) Unlinked RAC records (therefore no information on diagnoses).

Note: Age is as at (re-) entry into RAC. Diagnosis of dementia includes diagnoses of dementia and Alzheimer's disease (ICD-10-AM Ed. 1 categories F00–F03, and G30—see NCCH 1998). Table excludes 115 cases with missing RCS, and all unlinked RAC hospital leave events.

8.3 Summary

The above results together show that, overall, the E linkage strategies result in linked data that largely reflect the N linkage match set in terms of the distributions across key variables. That is, while not exactly the same, the E linkage match sets' distributions look similar to those for the N linkage match set. In general, constrained methods using SLA group when matching perform better than those using straight postcode, and constrained matching (even without using 2-digit postcode matching) performs at least as well as or better than basic E matching.

In terms of practical utility, analyses by post-hospital destination and source of RAC admission indicate that, as expected from the sensitivities, the E linkage strategies underestimate the amount of movement between hospital and RAC, with permanent RAC admissions being particularly affected. Nevertheless, illustrative examples looking at patterns of use and characteristics of people moving between the two sectors show that analyses using links derived from the N and E strategies lead to very similar conclusions. Results indicate also that, irrespective of the linkage strategy, care needs to be taken when drawing conclusions as differences may not be statistically different due to small numbers in some groups.

9 Conclusions

The comparison between name-based and event-based linkage strategies in this project have led to a number of results:

- When identifying transition events using data linkage, detailed knowledge of both the service systems and the data collection practices within those systems is essential.
- Using detailed comparisons of N and E links, and expanding the event data used to identify links, it has been possible to improve the general performance of the E linkage strategy as first put forward in the initial 2001–02 feasibility study.
- E linkage strategies that adjust the linkage procedures according to available information (constrained E linkage) perform better than those that apply the same approach to all records (basic E linkage).
- Less restrictive regional matching (based on 2-digit postcode) can be used for areas with small populations, but not among those with denser populations.
- Links identified via E linkage are highly reliable, with over 97% of links being true matches when compared to N links. There is some small variation in the positive predictive value of the linkage strategies across movement type.
- E linkage tends to miss matches (8–14% among preferred strategies) more often than make false matches. The sensitivity of E linkage varies with both the strategy and movement type.
- As expected, the incidence of missed links results in the E linkage strategies underestimating the volume of flow from hospital to RAC. Among the strategies examined, this underestimation, which particularly affects people moving into permanent RAC, was lowest for the CSLA constrained E strategies.
- Within movement groups, the E linkage strategies result in linked data that largely reflect the N linkage match set in terms of the distributions across key variables. In general, constrained methods using SLA group when matching perform better than those using straight postcode, and constrained matching (even without using 2-digit postcode matching) performs at least as well as or better than basic E matching.
- In terms of practical utility, illustrative examples looking at patterns of use and characteristics of people moving between the two sectors show that analyses using links derived from the N and E strategies lead to very similar conclusions.
- Overall, the high positive predictive value of the E strategies—especially when movement is known to have occurred and postcode information is reliable—suggest that such an approach could be used in other areas. In particular, it is expected that a similar strategy could be used to derive whole-of-stay hospital records for cases when patients move within the hospital sector. In such cases, date of transfer and postcode information would be expected to be highly consistent on the before- and after- transfer records.
- Although this comparison study has been very labour intensive, a repeat comparison using more recent datasets or sources from another jurisdiction, forgoing the tuning step, would be an excellent verification of the generalisability of the use of the E-strategy.

- It should be noted, however, that E-linkages without recourse to a validation source need to be restricted to small populations moving across single interfaces within a narrow time window.

Appendix 1: Match rules for N linkage and constrained E linkage strategies

A1.1 Match rules for N linkage

The purpose of N linkage match rules is to find the best event match for in-scope events given established (person-based) links for all people using health services and residential aged care in Western Australia over the period 1980 to 2005. To achieve this, matching constraints were specified to identify the most appropriate event match from all possible event matches for a person. Different rules were used depending on the type of RAC event being considered for the match. For this project, matching to service event dates and selection of one–one matches from within the established many–many matches for individuals was undertaken using automated algorithms programmed in *SAS®* (without extensive clerical review).

Match rule steps

All hospital events within the allocated time period (July 2000 to June 2001) and all RAC events for the same period were retrieved from Western Australia's Hospital Morbidity Data System and RAC data custodians. Using HiL's established hospital–RAC person links, all corresponding event records (hospital and RAC) were joined using their associated person numbers. The resultant many-to-many links were then filtered to select linked sets of hospital records and RAC records that were within scope for the comparison with the E matches. In addition, within the set of all the in-scope hospital records for individuals those that were contiguous (that is, transfers) were joined together to form a concatenated string of events in order to have more accurate start and end dates for the whole hospital episode. Any contiguous periods of RAC leave were similarly joined (a rare occurrence). A further reduction step was then made to select the most appropriate one-to-one event match from within these combined strings of events for the comparison with the E strategy. The most appropriate hospital–RAC event link was chosen by measuring the closeness of hospital and RAC event dates in conjunction with the rules described below. Death information was added to the linkage dataset to aid event selection.

Specific match rules

1. Match hospital episode data and RAC event data using the established person match, leading to many-to-many matches (as described above).
2. Within RAC event type, select the best RAC event match using the following rules:

   RAC hospital leave:

   There are a number of possible ways that hospital and RAC event dates can align. In priority order for selecting the most appropriate match, these are:

A. Exact matching of hospital stay and RAC hospital leave dates (without transfers):
RAC leave start date = hospital admission date
RAC leave end date = hospital separation date

B. Exact matching of hospital stay and RAC hospital leave dates (with transfers):
RAC leave start date = hospital admission date
RAC leave end date = hospital separation date

C. Exact matching of hospital stay and RAC hospital leave end dates (with, or without transfers):
RAC leave end date = hospital separation date

D. RAC hospital leave encompasses the hospital stay:
RAC leave start date ≤ hospital admission date
RAC leave end date ≥ hospital separation date

E. RAC hospital leave is within the hospital stay:
RAC leave start date ≥ hospital admission date
RAC leave end date ≤ hospital separation date

F. RAC hospital leave starts before the hospital stay and ends before the hospital stay ends:
RAC leave start date ≤ hospital admission date
RAC leave end date ≤ hospital separation date

G. RAC hospital leave starts after the hospital stay and ends after the hospital stay ends:
RAC leave start date ≥ hospital admission date
RAC leave end date ≥ hospital separation date

H. Within the above groups, RAC discharges to hospital are excluded from linked data as the project is looking at movement from hospital to RAC. A discharge to hospital is identified as:
RAC discharge date < hospital separation date

If there is a choice between matches with the same priority ranking, the difference

Lag = Hospital separation date – RAC leave end date

is used to choose the preferred link, with the smaller *Lag* being chosen (so that negative differences are to be preferred over positive ones).

RAC social leave:

I. The RAC social leave event must encompass entirely the hospital episode (common start and/or end dates are allowed).

If more than one hospital stay matches to the social leave, the last hospital event is retained.

RAC admissions (permanent and respite):

J. A permanent RAC admission is said to match to a hospital episode if the RAC admission date is not more than 2 days before or 7 days after the hospital separation date (that is, a [–2 days, 7 days] acceptance interval). Note: the ‘–2’ allows for some error in dates, and the ‘7’ allows for pre-entry leave for permanent admissions.

K. A respite RAC admission is said to match to a hospital episode if the admission date is not more than 2 days before or 2 days after the hospital separation date (that is, a [–2 days, 2 days] acceptance interval). Note: the differences allow for some error in dates.

If there are several candidate matches, take the admission with the smallest date gap (*Lag* as above using the RAC entry date).

3. Some hospital episodes may match to more than one RAC event type (most commonly, hospital leave event and a permanent admission). The best match is selected based on the RAC event start and end dates (where the end date for an admission is the following discharge date). This means that hospital leave events are preferred over transfer admissions following hospitalisation.
4. As date of death data are available for N links, links to a hospital episode ending in the death of the patient are identified from date of death and date of hospital separation (rather than hospital mode of separation or RAC data) using:
   Date of death ≤ hospital separation date

   These are dropped as we are interested in movement from hospital to RAC.

   Note: using hospital mode of separation rather than date information to determine death in hospital results in 461 links being dropped compared with 464. Overall, seven links identified as deaths via dates had separation mode other than death, and four links with separation mode of death did not meet the above date criterion.

A1.2 Match rules for constrained E linkage

The purpose of constrained E linkage is to find the best match using all available event date information and event descriptors. To achieve this, matching constraints are specified separately for comparisons between different subsets of RAC and hospital events defined in terms of their purpose and/or admission and separation characteristics (refer to Appendix 2). Because two dates are available for RAC hospital leave (and social leave), match procedures for these events are the most complicated.

Constrained E matching is carried out in two stages. Initial matches are selected using 1:1 probability matching via the computer package *Websphere*®. Relatively broad match criteria are used to identify possible matches between RAC and hospital partitioned datasets. Each partitioned dataset pair is compared using a specific *Websphere*® procedure—for this project 12 such dataset pairs were used (see Table 5.1). Finer match rules are then applied (in *SAS*®) to select the final matches.

The rules applied to ensure that matches meet certain criteria are described below.

Stage 1 rules: *Websphere*® 1:1 probability matching

1. Sex must match.
2. Date of birth rules (done in *Websphere*®, checked in *SAS*®):
   – Allow differences in only one of day, month or year.
   – Year of birth differences must be less than 8 years.
3. Geographic matching: The geographic matching is based on postcode. This can either be done using straight postcodes, or using postcode-based SLA groups. Under the latter, postcodes are said to match if they include a common SLA using a postcode:SLA concordance. Note that for RAC admissions with their relevant ACAT in hospital, the RAC client postcode is compared both with the hospital client postcode and the hospital's own postcode.

4. Initial matching on event dates is carried out in *Websphere*® using relatively generous date intervals in the later passes (see Table 5.2). Note, if there are several possible matches to a particular record, under probability matching *Websphere*® chooses the closest match out of any possible matches (that is, the match with the highest weight; for example, that with the closest date when allowing date variation). The extent of duplicate matches (that is, more than one possible *Websphere*® match) is shown for the current project in Table A1.1. From this it can be seen that exact duplicates did not occur in the constrained CPC matching strategy, so that when one record in one dataset had more than one possible match in the other dataset there was a preferred match. Further, with one exception, exact duplicates only occurred in the constrained CSLA strategy in passes 6 and 7. However, it should be noted that in SLA group based datasets exact duplicates under match passes which do not include SLA matching result from the structure of the dataset—compare passes 6 and 7 for CSLA and CPC. In these cases, the choice of link did not affect the final linked dataset as the duplicates related to the same event. Exact duplicates may also occur if using basic E linkage (see BSESLA, with 192 exact duplicates compared with 13,989 links).

Stage 2 rules: *SAS*®

5. Dates must meet the criteria in Table 5.1.
6. Matches of RAC hospital leave are identified as discharges to hospital if
   RAC discharge date < *hospital separation date*
   These are dropped (separated out) as we are interested in movement from hospital to RAC.
7. Noting that date of death data are not available on either the RAC or NHMD datasets, matches to deaths in hospital are identified via hospital and RAC data as follows:
   (a) Links between a RAC episode of hospital leave are identified as linking to a death in hospital if the matching hospital episode was recorded as ending due to death of the patient.
   (b) For links to social leave, deaths in hospital are identified by:
   RAC entry date = *RAC discharge date*
   and
   RAC reason for discharge = *death.*
   (c) Links to RAC admissions are assumed not to relate to deaths in hospital.

   Links (a) and (b) are dropped as we are interested in movement from hospital to RAC.

   Aside: The small number of inconsistencies between hospital mode of separation and date of death data observed in the N links suggested that perhaps hospital mode of separation should not be used to exclude deaths in hospital in the E linkage strategy. Using hospital separation mode in conjunction with RAC date and discharge information (that is, both tests (a) and (b) above) to determine death in hospital resulted in 449 CSLA links being dropped compared with 438 if only RAC data were used (that is, applying just (b) above to all RAC events), so that 11 links identified as deaths via hospital separation mode (that is, test (a)) did not meet the RAC date and discharge criteria of test (b). Of these 11, four had hospital and death date data that were consistent with death in hospital but 'death' was not reported as the RAC reason for discharge. The remaining seven failed the RAC date test, with four having a day between the two dates; six of the seven had also linked under N linkage and the date of death information available from those links was consistent with a death in hospital. For links achieved under both N linkage and CSLA, only five out of 439 cases identified as deaths in hospital using both tests (a) and (b) above were not identified as such using death date data. These results indicate that both tests (a) and (b) should be used to determine death in hospital under the E linkage strategy.

8. The best match among duplicate matches resulting from matching a particular subset of RAC events to more than one subset of hospital events (for example, RAC admissions matching to hospital events discharged to usual residence and to those discharged to RAC) is selected using the following priority:
   i. Matches to RAC hospital leave (top priority). There may also be duplicates within RAC hospital leave. The priority ranking among these matches is:
      a. Hospital event with non-statistical admission, discharged to usual residence (top priority)
      b. Hospital event with non-statistical admission, discharged to death
      c. Hospital event with non-statistical admission, discharged to other (including to RAC)
      d. Hospital event with statistical admission, discharged to usual residence
      e. Hospital event with statistical admission, discharged to death
      f. Hospital event with statistical admission, discharged to other (including to RAC).
   ii. Matches to RAC admissions. There may also be duplicates within matches to RAC admissions. The priority ranking among these matches is:
      a. Hospital event reported as discharged to RAC (top priority)
         – match to hospital event using person postcode has priority over match using hospital postcode (used only if ACAT in hospital)
      b. Hospital event reported as discharged to usual residence
         – match to hospital event using person postcode has priority over match using hospital postcode (used only if ACAT in hospital)
      c. Hospital event reported as discharged to death (assumed invalid—see rule 7).
   iii. Matches to RAC social leave.

Table A1.1: Duplicates in AIHW *Websphere*® matching, by *Websphere*® procedure and pass (number before refining in *SAS*®)

Procedure/ pass[(a)(b)]	RAC records					Hospital records					[(b)]Links (before *SAS*® stage)
	Exact duplicates		Other duplicates		[(a)]Total records	Exact duplicates		Other duplicates		[(b)]Total records	
	1–5	[(c)]6–7	1–5	[(c)]6–7		1–5	[(c)]6–7	1–5	[(c)]6–7		
CSLA linkage											
CSLNST8H	—	—	63	—	13,756	—	—	—	—	5,802	901
CSLST8H	—	—	—	—	13,756	—	—	—	—	1,675	195
CSL0ADM	—	183	—	—	16,180	—	185	—	—	7,840	3,190
CSL9ADM	—	59	3	—	16,180	—	85	7	—	150,297	1,668
CSL9SOC	—	—	—	—	9,605	4	—	—	—	165,614	344
CSLH0ADM	—	72	2	—	6,936	—	436	1	—	10,064	2,240
CSLH9ADM	—	35	7	—	6,936	—	145	—	—	[(d)]205,756	857
CSLNST0H	—	—	299	—	13,756	—	9	6	—	5,007	1,827
CSLNST9H	—	25	256	—	13,756	—	37	284	—	137,731	8,317
CSLST0H	—	—	—	—	13,756	—	2	—	1	2,833	393
CSLST9H	—	—	1	3	13,756	—	—	—	7	12,566	899
All	. .	. .	. .	. .	**39,541**	. .	. .	. .	. .	**165,614**	. .
CPC linkage											
CPCNST8H	—	—	63	—	6,956	—	—	—	—	3,098	488
CPCST8H	—	—	—	—	6,956	—	—	—	—	917	98
CPC0ADM	—	—	—	—	8,157	—	—	—	—	3,969	1,712
CPC9ADM	—	—	1	—	8,157	—	—	4	—	79,250	948
CPC9SOC	—	—	—	—	4,720	—	—	4	—	87,234	187
CPCH0ADM	—	—	1	—	3,432	—	—	2	—	3,969	604
CPCH9ADM	—	—	3	—	3,432	—	1	1	—	79,250	406
CPCNST0H	—	—	157	—	6,956	—	—	7	—	2,514	890
CPCNST9H	—	—	136	—	6,956	—	—	171	—	72,310	4,246
CPCST0H	—	—	—	—	6,956	—	—	—	—	1,455	205
CPCST9H	—	—	—	—	6,956	—	—	—	—	6,940	486
All	. .	. .	. .	. .	**19,833**	. .	. .	. .	. .	**87,234**	. .

(continued)

Table A1.1 (continued): Duplicates in AIHW *Websphere*® matching, by *Websphere*® procedure and pass (number before refining in *SAS*®)

Procedure/ pass[(a)(b)]	RAC records: Exact duplicates 1–5	RAC records: Exact duplicates [(c)]6–7	RAC records: Other duplicates 1–5	RAC records: Other duplicates [(c)]6–7	RAC records: [(a)]Total records	Hospital records: Exact duplicates 1–5	Hospital records: Exact duplicates [(c)]6–7	Hospital records: Other duplicates 1–5	Hospital records: Other duplicates [(c)]6–7	Hospital records: [(b)]Total records	[(b)]Links (before *SAS*® stage)
BSESLA	192	. .	1	. .	39,541	8	. .	—	. .	165,614	13,989
Unlinked: SLA[(e)]	265				39,541	89				165,614	. .
Unlinked: postcode[(e)]	123				19,833	31				87,234	. .

(a) 'CSL' implies constrained SLA group matching, 'CPC' implies constrained postcode matching and 'BSE' implies the basic AIHW linkage strategy. See Table 5.1 for an explanation of the procedure dataset partition codes and Table 5.2 for a description of the passes.

(b) When using SLA group in the matching there are multiple records for an event, with each event being repeated for each SLA in the group, so that the only difference between the repeated records is the SLA data. Multiple matches between the same hospital and RAC events due to SLA group matching are identified and reduced back to a single match via *SAS*®. The repetition of records also leads to exact duplicates in passes that use postcode. On average there were 2.0 SLAs in an SLA group for RAC events, and 1.9 for hospital episodes.

(c) Matches in passes 6 and 7 are based on 3- and 2-digit postcode, and so repeated records will appear as exact duplicates. *Websphere*® chooses one of these as the match. In terms of final match outcomes, it is irrelevant which one gets chosen.

(d) SLA group based on hospital postcode rather than patient postcode. All other 'SLA' datasets use the SLA group of the client.

(e) Based on unpartitioned data using: date of birth, sex, SLA/postcode, hospital exit/RAC entry date.

Note: In the current application of *Websphere*®, a pair of records is said to match if they have a positive weight (that is, meet the specified *Websphere*® match rules). The table shows the number of duplicates identified in *Websphere*® passes, and whether or not these duplicates were identical in terms of match weights derived in *Websphere*®. For non-exact duplicates, *Websphere*® chooses the match with the highest match weight (that is, the 'nearest' match). For exact matches, the chosen match depends on file order. Note that the number of duplicates excludes the selected match.

Appendix 2: Illustrating event matching for constrained strategies

A2.1 Partitioning the datasets for constrained E matching

Event information may suggest that some matches are more likely to be correct than others (for example, a link *RAC admission–hospital discharge to RAC* has greater face validity than a link *RAC admission–hospital discharge to usual residence*). Thus, dataset partitioning based on event characteristics not only minimises coincident records (with respect to match data) within datasets by reducing the number of records being compared, but also allows link priorities to be set. Consequently, matching within partitioned datasets facilitates selection of the most likely match if duplicate links occur when the links from the partitioned datasets are combined. (Note, if exact duplicates within a partitioned subset match with a record in the other data subset then a choice has to be made between the two matches. In *Websphere*® this is determined by the record order in the datasets).

Table A2.1: Type of RAC events most likely to link to a hospital separation, by hospital mode of separation

Hospital mode of separation	Most likely type of RAC (re) entry (if any)
To other hospital (codes 1, 3)	Highly unlikely to have associated RAC event—excluded from linkage
Statistical discharge (code 5)	Can't have associated RAC event—excluded from linkage
To RAC, when this is not the usual residence (code 2)	Permanent or respite admission
To other health care accommodation (code 4)	Unlikely to have associated RAC event, but most likely associated with a permanent or respite admission
Left against medical advice, statistical discharge from leave, and unknown/not supplied (codes 6, 7 and 0)	No preferred RAC type
Death (code 8)	Hospital leave, and perhaps social leave
Other, including to RAC as usual residence (code 9)	Hospital leave, and perhaps social leave

The purpose of constrained event-based matching is to find the best match using all the available date information. To achieve this, matching procedures are specified separately for a range of data subset pairs derived by partitioning the hospital and RAC data. Four types of partitioning are used:

1. Hospital *Mode of separation*, indicating destination following hospitalisation: While the quality of *Mode of separation* data is not thought to be particularly high (see AIHW 2003), partitioning hospital data on this variable allows more likely sources of matches to be compared before less likely sources. Table A2.1 shows the more likely sources for links with RAC data for different modes of separation from hospital. Note that because of the possible confusion between reporting going to RAC and to another health care establishment, the relatively small numbers of separations via modes other than death and to usual residence (see Table 3.1), and the consequently small probability of

coincident matching data, hospital modes of separation other than death and to usual residence are often grouped together when undertaking matching.

2. Hospital *Mode of admission*, indicating hospital episodes starting with a transfer: For the E linkage strategy, *Mode of admission* primarily provides information on the accuracy of the start date of the hospital episode. As illustrated in Table A2.2, these start dates are useful when linking RAC hospital leave events (and, to a lesser degree, social leave events) with hospital separations, as the additional information allows more accurate linking, especially when all other linkage data are coincident. (In this report, a hospital admission is termed 'statistical' if the hospital patient is changing from one episode care type to another or transferring from one hospital to another. Other admissions are called 'non-statistical' admissions.)
3. RAC *Type of event*, distinguishing between RAC leave events and admissions. Initial analysis showed that many of the RAC events with coincident data for date of birth, sex, SLA and RAC in-date result from the same person having two RAC events on the same day—generally as the result of a return from hospital leave coinciding with a change in RAC facility (that is, they did not return to the facility from which they had hospital leave). Partitioning on type of RAC event allows these two events to be considered separately, and prioritisation can then be used to determine which link should be used if the two events link to a single hospital separation. In addition, there are some types of events that should not link to a hospital separation; in particular, a RAC admission immediately following a discharge from respite care should not be linked to a hospital discharge as a person cannot go on hospital leave while in residential respite care. If desired, these cases can be identified and then excluded from linked dataset.
4. RAC *Place of ACAT assessment* (categorised as in hospital, at home, in RAC, other): *Place of ACAT assessment* provides further information for linkage. In particular, if the assessment took place in hospital during the current hospital episode it may be more appropriate to compare the area of 'usual residence' as recorded in the RAC data (and which relates to a contact address for pre-2003 data) with the area of the hospital rather than the patient's usual residence as recorded in the hospital data.

Using the above variables, the hospital and RAC data are partitioned into a number of subsets to facilitate the event-based linkage. The partitioned set pairs used when matching are described below, and the partition code referred to in the main part of the report is given (see also Table 5.1). The priority of the matches is also indicated (in alphanumeric order). An overview of the linkage strategies to be used for the various partitioned pairs is given in Table A2.2.

Table A2.2: Possible links for hospital separations, and associated linkage strategy

Type of admission for hospital episode	Type of RAC entry			
	Respite admission	Permanent admission	Return from RAC hospital leave	Return to RAC social leave
With statistical admission (transfer from another hospital (code 1) and within hospital statistical admission (code 2))	Single date matching[(a)]	Single date matching[(a)]	End-date cover matching[(b)] Extended cover matching (discharge to hospital)[(c)]	Extended cover matching[(d)]
With non-statistical admission (other (code 3) and unknown (code 9) admission)	Single date matching[(a)]	Single date matching[(a)]	Period matching[(e)] Start-date cover matching (discharge to hospital)[(f)]	Extended cover matching[(c)]

(a) Linking on hospital separation date and residential aged care admission date.

(b) Linking on hospital separation date and RAC leave return date, and RAC leave period covers the hospital episode.

(c) Hospital episode partially covers the RAC leave period.

(d) RAC leave period covers the hospital episode.

(e) Linking on hospital separation start and end dates and RAC leave start and end dates.

(f) Linking on hospital start date and RAC leave start date, and the hospital episode covers the RAC leave period.

Note: Table assumes same-day and statistical hospital separations are excluded from the matching.

Source: Adapted from AIHW: Karmel 2004:19.

A2.2 Matching RAC hospital leave

There is a range of possible scenarios for event date overlap with hospital episodes for people on RAC hospital leave, and these are illustrated below. To assist in the matching, hospital episodes are partitioned by mode of admission and mode of separation.[7]

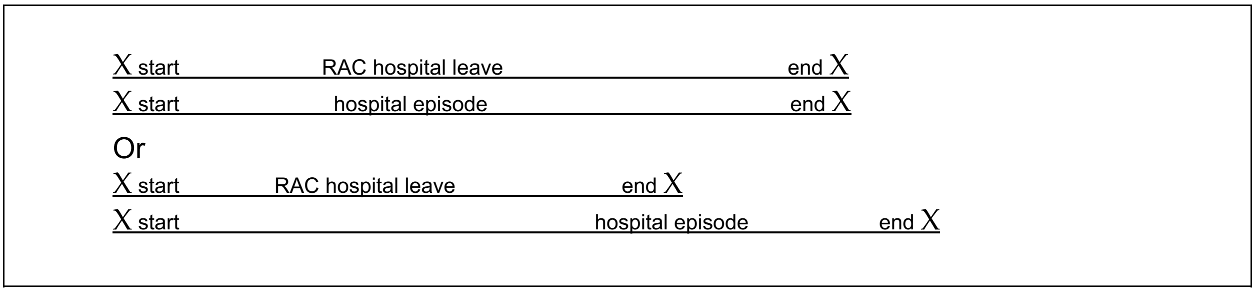
Matching to non-statistical hospital admissions (admission mode = 3, 9)

When linking RAC hospital leave to a hospital episode with a non-statistical admission, we expect the start of the RAC hospital leave to match the admission date of the hospital episode. However, if a person is discharged to hospital or dies in hospital, the RAC hospital leave end date may be earlier than the hospital episode separation date. Deaths in hospital and discharges to hospital are not retained when looking at movements from hospital to RAC. A person discharged to hospital while on RAC hospital leave may later leave hospital and go into the same or different RAC facility; in this case we retain the link to the relevant admission (achieved via matches 7 to 11).

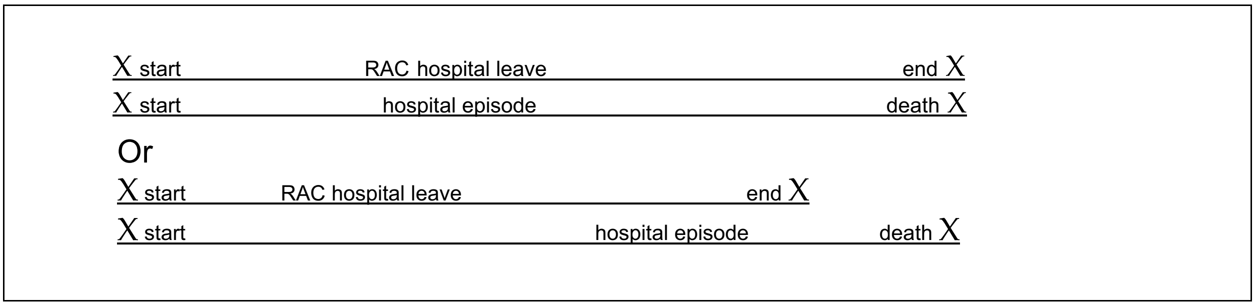
Matches 1, 2 and 3 all allow for the patient giving up or losing their RAC place while they are in hospital, that is, if the RAC resident was discharged from residential care while in hospital. Match 3 allows for misreporting of hospital mode of separation, that is, not reported as going to usual residence or death.

7 Codes refer to code sets used in 2000–01 NHMD.

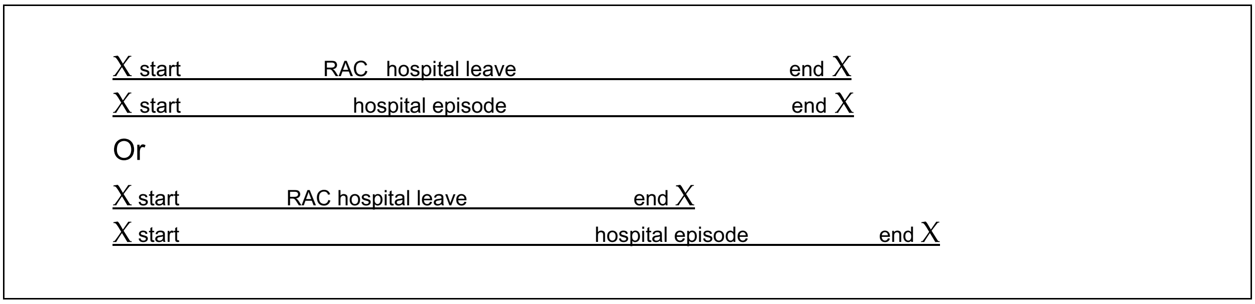
Match 1: Matching RAC hospital leave and non-statistical hospital episodes discharged to usual residence (separation mode = 9), Partition code NST9H, priority = H11



Match 2: Matching RAC hospital leave and non-statistical hospital episodes ending in death (separation mode = 8), Partition code NST8H, priority = H12



Match 3: Matching RAC hospital leave and other non-statistical hospital episodes (separation mode = 2, 4, 6, 7, 0), Partition code NST0H, priority = H13



Matching to statistical hospital admissions (admission mode = 1, 2)

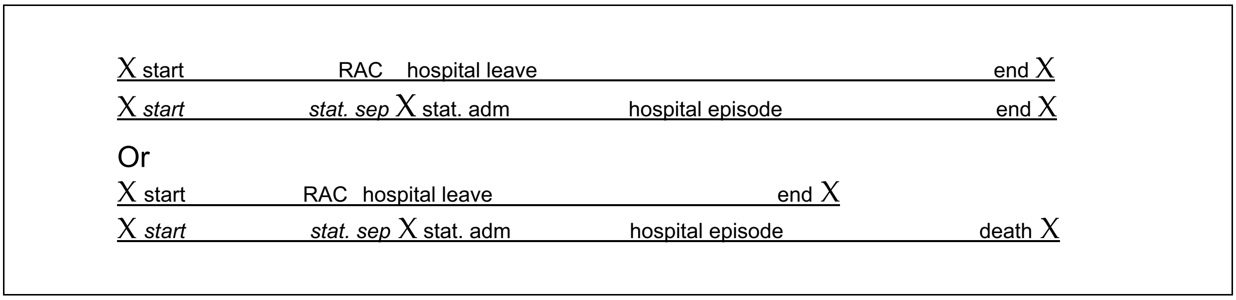
When linking RAC hospital leave to a hospital episode with a statistical admission, the start of the RAC hospital leave may be before the admission date for the hospital episode. In addition, if a person is discharged to hospital or dies in hospital, the RAC hospital leave end date may be earlier than the hospital episode separation date. As stated above, deaths in hospital and discharges to hospital are not retained when looking at movements from hospital to RAC. A person discharged to hospital while on RAC hospital leave may later leave hospital and go into the same or different RAC; in this case we retain the link to the relevant admission (achieved via matches 7 to 11).

Note that matches 4 to 6 allow for the patient giving up or losing their RAC place before the end of the hospital episode. Also, like match 3, match 5 allows for misreporting of hospital mode of separation; that is, not reported as going to usual residence or death. In addition, it should be noted that cases where the discharge to hospital occurs during the earlier contiguous hospital episode will not be identified in the current matching process, as such hospital episodes are excluded from the linkage dataset (see Section 3).

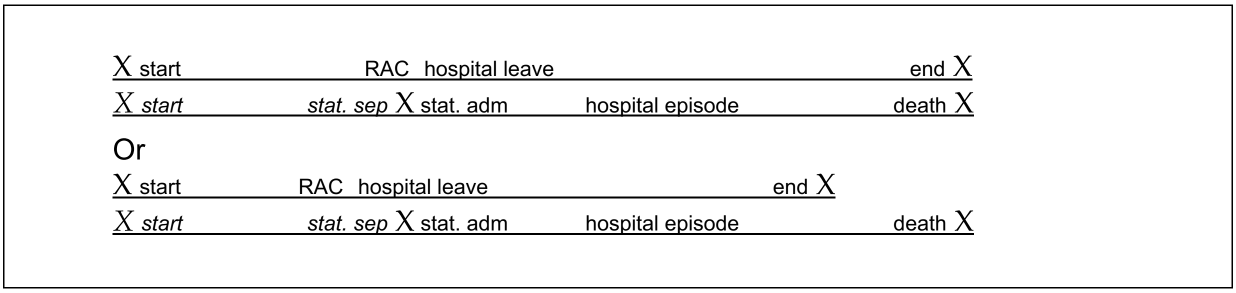
Match 4: Matching RAC hospital leave and statistical hospital episodes discharged to usual residence (separation mode = 9), Partition code ST9H, priority = H21

and

Match 5: Matching RAC hospital leave and statistical hospital episodes with other separation modes (separation mode = 2, 4, 6, 7, 0), Partition code ST0H, priority = H23



Match 6: Matching RAC hospital leave and statistical hospital episodes ending in death (separation mode = 8), Partition code ST8H, priority = H22



A2.3 Matching RAC admissions

For this matching, whether or not the hospital separation has a statistical or non-statistical admission is of no importance as only the hospital separation date is relevant for matching. However, partitioning on hospital mode of separation is again used to allow the most appropriate links to be identified. Also, partitioning on RAC place of ACAT assessment is used to aid region matching.

Note that while people moving from hospital to RAC usually leave hospital on the same day as they are admitted into permanent RAC admission, there may be occasions when this is not so. In particular, up to 7 days of social leave may be used as pre-entry leave immediately before a resident enters an aged care home: 'Pre-entry leave gives a prospective resident time to make arrangements to enter an aged care home or to transfer from one home to another home in a distant location. It enables the home to receive subsidy and keep the place vacant for a prospective resident for up to 7 days after he or she agrees to enter care…Pre-entry leave may be claimed for days on which the intending resident is in hospital' (DoHA 2005:195). Consequently, hospital separation dates are compared both with the RAC admission date and the date at the beginning of any related pre-entry leave; that is, in the following diagrams the RAC 'start' date may be either of these dates.

RAC admissions with ACAT assessment in hospital

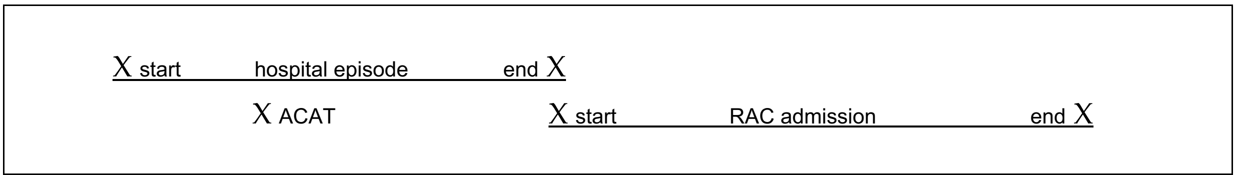
When the assessment is in hospital, the region reported for the RAC client may be that for the hospital (if this was given as the contact address for the ACAT). In this case, the reported RAC client region should be compared with both the hospital region and the usual residence region recorded by the hospital. However, in the hospital data hospital region will generally be known only for public hospitals. Note, that matches based on hospital region are given lower priority than those matching using person region.

Match 7 and, in particular, match 8 both allow for misreporting of hospital mode of separation, that is, not reported as going into RAC. Deaths in hospital are not included in this matching because of the possibility of introducing false matches due to the confined range of hospital regions and because there is only a very small chance of mode of hospital separation being coded incorrectly to death (see Appendix 1).

Match 7: Matching RAC admissions and ACAT assessment in hospital and hospital separations other than to usual residence (separation mode = 2, 4, 6, 7, 0), Partition code H0ADM, priority = R31

and

Match 8: Matching RAC admissions and ACAT assessment in hospital and hospital separations discharged to usual residence (separation mode = 9), Partition code H9ADM, priority = R32



Ignoring place of ACAT assessment

Even if assessment were in hospital, the contact address for RAC can still be the person's usual residence. However, having the ACAT assessment within the hospital period may help to distinguish between otherwise coincident matches.

Match 9, and in particular, matches 10 and 11 allow for misreporting of hospital mode of separation, that is, not reported as going into RAC. Deaths in hospital are included in this matching despite the unlikely event of mode of hospital separation being coded incorrectly to death because of the greater accuracy of the region data. (Note: when refining the constrained linkage strategies, links between RAC admissions and hospital discharges due to death are excluded due to their poor performance—see Table 6.3.)

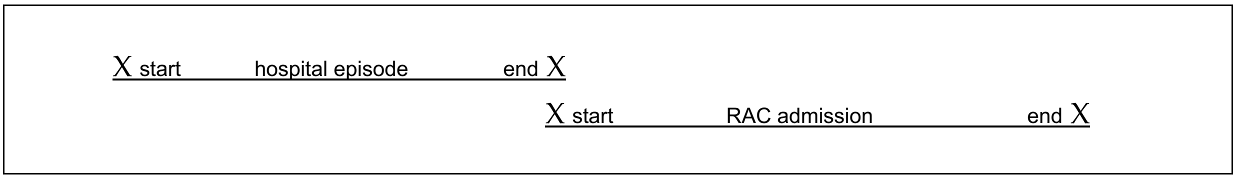
Match 9: Matching RAC admissions and hospital separations excluding discharge to usual residence and deaths (includes separation mode = 2, 4, 6, 7, 0), Partition code 0ADM, priority = R21

and

Match 10: Matching RAC admissions and hospital separations discharged to usual residence (separation mode = 9), Partition code 9ADM, priority = R22

and

Match 11: Matching RAC admissions and hospital separations ending in deaths (separation mode = 8), Partition code 8ADM, priority = R41

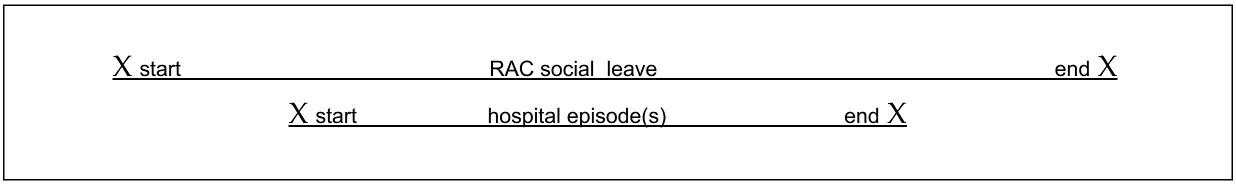


A2.4 Matching RAC social leave

RAC permanent residents can access both social leave and hospital leave. If a person needs to go to hospital while they are on social leave then they can change their leave type. It is to their advantage to do this as the amount of social leave that a person can take within any one financial year is limited. Consequently few valid matches to social leave are expected. When matching to social leave the type of admission to hospital is not considered as the period of RAC social leave should encompass the hospital episode. If a person does change their type of RAC leave when they enter hospital, a match between the hospital leave and hospital episode should be made when matching to hospital leave. Although unlikely, several hospital episodes may link to the one social leave event. The linkage strategy allows only one match to be made.

Note that the same matching procedure can be used for matching to any mode of hospital separation. Given the small number of matches expected for RAC social leave, matching with hospital episodes of all separation modes are carried out at the same time.

Match 12: Matching RAC social leave and hospital separations, Partition code 9SOC, priority = S1



Appendix 3: Preliminary CSLA analysis

Table A3.1: Summary of preliminary CSLA positive predictive values using N links as the reference standard, by *Websphere*® procedure and pass (% CSLA links)

	Pass							
	1	2	3	4	5	6	7	
Procedure	Exact match within SLA group	1-sided event date variation	YOB variation	Variation in month *or* day in DOB	1-sided event date variation, US date of birth	1-sided event date variation, YOB variation	1-sided event date variation, month *or* day DOB variation	All links
	PPV (per cent)[(a)]							N
Matching to deaths in hospital[(b)]								
CSL8ADM	. .	**—	**—	**—	**—	**—	**—	21
CSLNST8H	80.8	86.0	**100.0	*90.0	. .	. .	**—	403
CSLST8H	100.0	37.0	**100.0	. .	**100.0	. .	**—	95
Other matches (excluding to deaths in hospital and discharges to hospital)								
CSL0ADM	97.2	81.0	75.0	92.9	**20.0	. .	**—	1,267
CSL9ADM	96.3	72.7	32.7	8.9	4.5	—	1.1	906
CSL9SOC	95.5	10.0	—	3.0	. .	. .	. .	289
CSLH0ADM	97.7	88.9	**100.0	. .	**100.0	**100.0	. .	128
CSLH9ADM	94.1	69.2	. .	**14.3	*—	*—	—	109
CSLNST0H	99.4	100.0	*100.0	*100.0	**100.0	**100.0	**100.0	649
CSLNST9H	98.7	98.1	100.0	96.9	*75.0	*50.0	*66.7	3,690
CSLST0H	99.4	**100.0	**100.0	**100.0	. .	. .	. .	165
CSLST9H	99.0	**87.5	*100.0	*83.3	**—	**100.0	**—	436
Total (% match)	**98.2**	**90.7**	**61.0**	**47.2**	**16.1**	**11.9**	**7.0**	**90.3**
Total (all links)	**5,982**	**788**	**177**	**392**	**112**	**67**	**142**	**7,660**

(a) Links to hospital records: PPV rate includes links to same people, different RAC event.

(b) Links for RAC discharges to hospital have been excluded for N but not E links (affects 70 CSLNST8H and 17 CSLST8H matches).

* Based on 10–19 matches.

** Based on fewer than 10 matches.

Appendix 4: Analysis of CSLA missed and false links

A4.1 CSLA missed links

Several measures were looked at to identify reasons for missing N matches when using CSLA matching:

- Date of birth differences. Dates of birth on the RAC and hospital datasets are not sufficiently similar for CSLA matching if they differ by two or more elements, or if the years of birth differ by 8 or more years.
- Different sex on the two datasets (unacceptable in the AIHW strategy).
- Possible SLA group difference, identified by different postcodes on the two datasets.
- Poor hospital-to-RAC date match, measured by Lag2 = (RAC in date – hospital out date). Lag2 is relevant for all RAC events, and is unacceptable for CSLA links if it is negative or more than 2 days.
- Poor RAC-to-hospital date match, measured by Lag1 = (hospital in date – RAC out date). Lag1 is relevant for RAC leave events only, and is unacceptable for CSLA links if it is negative or more than 2 days.

Overall, there were 671 links made via N linkage and not by the refined CSLA strategy (excluding mixed matches, see Table 6.4). Over half of these related to hospital leave (53%), 46% were for RAC admissions and the remaining 1.5% were links between hospital events and RAC social leave (Table A4.1).

Table A4.1: Matches made by N linkage but not CSLA linkage,[(a)] by RAC event type and hospital separation mode

	Hospital separation mode						
RAC event type	To RAC	To other health care establishment	Left against medical advice/ statistical discharge from leave/unknown	Died	To usual residence /other	Total	
			Number				Per cent
Admission	101	60	18	—	129	308	45.9
Hospital leave	35	34	16	4	264	353	52.6
Social leave	2	—	—	—	8	10	1.5
Total	**138**	**94**	**34**	**4**	**401**	**671**	**100.0**

(a) Table excludes mixed CSLA–N links, and includes 2-digit postcode CSLA links.

There were 89 cases of missed matches which passed the individual match restrictions used in CSLA (that is, were 'CSLA-acceptable'), and so were matches that could possibly have been made under CSLA (in Table A4.2 and Table A4.3 combined). These 89 include the 42 'good' matches dropped when refining the CSLA strategy to exclude match passes with low

PPVs. All 89 had at least one element that could have led to a missed match under CSLA due either to exclusions or the *Websphere®* weighting algorithm:

- 4 were for cases with hospital separation mode of 'Death' (excluded from CSLA links, but not excluded from N links as the date of death was in fact after the hospital discharge date).
- 22 only had event date differences between the hospital and RAC data, with 21 of these relating to admissions with date differences of 2 days (missed because of the weights derived for *Websphere®*).
- 19 had differences for both date of birth and event dates.
- 44 only had date of birth differences. Of these, 10 had a different day, 4 had a different month, and 30 had a different year (20 with a year difference of 1–3 years, and 10 with a difference of 4–6 years).

Links to RAC admissions

Considering N-only matches to RAC admissions, poor region matching was the main reason for missing these matches under the CSLA strategy, with 70% (217) of the 308 missed matches involving records with different postcodes recorded in the hospital and RAC data (including 27 with poor event date matches as well; Table A4.2). Missed CSLA-possible matches accounted for 54 (18%) missed matches, with 42 of these resulting from excluding largely ineffective match passes (within match procedures CSL[H]9ADM). CSLA-unacceptable matches on either date of birth or event dates caused few missed links on their own (both under 25 cases). All of those with CSLA-unacceptable date matches related to matches where the RAC admission date was before the hospital discharge date (negative Lag2), with 14 out of 15 having a gap of 3 or more days (Table A4.2).

Table A4.2: Matches to RAC admissions made by N linkage and not CSLA linkage: indicators of poor variable matching leading to no match being made under CSLA

Reason for missing match	Poor date of birth: 1 element	Poor date of birth: >1 element	Lag2 <0	Other Lag2 difference	All
Missed possible match[(a)]	33	—	—	26	54
At least date of birth poor (different in more than one element; or >7 years between years of birth)	5	17	2	—	22
Sex different	—	—	—	—	0
Only possible SLA group difference	24	—	—	19	190
Only lag between hospital exit and RAC entry (Lag2) unacceptable[(b)]	2	—	15	—	15
Lag between hospital exit and RAC entry (Lag2) unacceptable and possible SLA group difference	2	—	27	—	27
All with possible SLA group differences	*26*	—	*27*	*19*	*217*
Total	**66**	**17**	**44**	**45**	**308**

(a) Possible missed matches are those with matching postcodes, and date of birth and event dates individually acceptable for CSLA linking.

(b) Lag2 (hospital to RAC) is relevant for all RAC episodes. Lag2 is unacceptable for CSLA links if (RAC in date – hospital out date) is negative or more than 2 days.

Links to RAC leave events

Unlike missed matches to RAC admissions, missed matches to RAC leave events were primarily the result of poorly matching event dates. Of the 363 N-only matches to RAC leave events (including 10 to social leave), 215 (59%) solely had date matches considered unacceptable in the CSLA strategy compared with just under 19% (68) with possible SLA-group differences; 12% had poor date of birth or sex matches (45). Almost one-quarter of the links with only poor event date matches had CSLA-unacceptable differences in both the start and end dates on the two datasets. The remainder were fairly evenly split between differences in the start date and differences in the end date. N matches where the hospital entry occurred before the RAC leave start date—that is, a negative Lag1 which is unacceptable in CSLA matching—were more commonly for hospital separations starting with a non-statistical admission than with a statistical admission (36 versus 5).

Looking in more detail at the 215 missed matches that had only CSLA-unacceptable event date matches (Table A4.4), more of the missed matches had a large positive gap than a negative gap between the dates recorded for exiting RAC (on leave) and entering hospital (Lag1; 96 versus 33). Two-thirds of the cases where the hospital entry date was before the RAC exit date involved gaps of less than 4 days, compared with half of the 'late' hospital entries involving more than a week's difference. On the other hand, similar numbers had the RAC return date preceding the hospital discharge date or unacceptably late after the hospital discharge date (Lag2 more than 2 days later). Of the latter, nearly two-thirds (40 out of 66) had a gap of more than a week.

Table A4.3: Matches to RAC leave events made only by N linkage: indicators of poor variable matching leading to no match being made under CSLA

	Poor date of birth		Lag2		Lag1 <0			
Reason for missing match	1 element	>1 element	<0	Other	[(a)]To stat. adm.	To non-stat. adm.	Other Lag1 difference	All
Missed possible match[(b)]	30	—	—	6	—	—	12	35
At least date of birth poor (different in more than one element; or >7 years between years of birth)	14	27	—	5	—	1	11	41
Sex different	—	—	—	—	—	—	2	4
Only lag between hospital exit and RAC entry (Lag2) unacceptable[(c)]	18	—	30	56	—	—	23	86
Only lag between hospital entry and RAC exit (Lag1) different[(d)]	11	—	—	11	2	22	55	79
Both Lag1 and Lag2 unacceptable	2	—	40	10	1	8	41	50
With unacceptable event dates only	*31*	—	*70*	*88*	*3*	*30*	*119*	*215*
Only possible SLA group difference	3	—	—	4	—	—	11	41
Lag between hospital exit and RAC entry (Lag2) unacceptable and possible SLA group difference	—	—	6	3	—	—	3	9
Lag between hospital entry and RAC exit (Lag1) unacceptable and possible SLA group difference	—	—	—	2	1	4	7	12
Both Lag1 and Lag2 unacceptable, and possible SLA group differences	—	—	4	2	1	1	4	6
All with possible SLA group differences	*3*	—	*10*	*11*	*2*	*5*	*25*	*68*
Total	**78**	**27**	**80**	**99**	**5**	**36**	**169**	**363**

(a) Among the missed matches there were 82 N-only matches between RAC leave events and hospital episodes starting with a statistical admission and 281 N-only matches starting with a non-statistical admission.

(b) Possible missed matches are those with matching postcodes, and date of birth and event dates individually acceptable for CSLA linking.

(c) Lag2 (hospital to RAC) is relevant for all RAC episodes. Lag2 unacceptable for CSLA links if (RAC in date – hospital out date) is negative or more than 2 days.

(d) Lag1 (RAC to hospital) is relevant for RAC leave episodes only. Lag1 unacceptable for CSLA links if (hospital in date – RAC out date) is negative or more than 2 days.

Table A4.4: Matches to RAC events made only by N linkage: variation in date matches, by RAC event type for events with CSLA-unacceptable date matches (unacceptable Lag1 and/or Lag2)(a)(b)

	Lag2 (RAC in date – hospital out date)		Lag1 (hospital in date – RAC out date)		
Days different	**Link to RAC admission**	**Link to RAC leave**	**Statistical admission**	**Non-statistical admission**	**Total**
< –3	7	16	—	10	10
–3	7	5	2	8	10
–2	—	13	—	2	2
–1	1	36	1	10	11
Total <0	*15*	*72*	*3*	*30*	*33*
0(c)	—	68	5	58	63
1	—	10	8	9	17
2	—	1	5	1	6
Total 0–2	—	*79*	*18*	*68*	*86*
3	—	8	6	5	11
4	—	4	5	6	11
5	—	6	6	5	11
6	—	5	1	4	5
7	—	3	—	5	5
>7	—	40	23	30	53
Total >2	—	*66*	*41*	*55*	*96*
Total	**15**	**215**	**62**	**153**	**215**

(a) Lag2 (hospital to RAC) is relevant for all RAC episodes. Lag2 unacceptable for AIHW links if (RAC in date – hospital out date) is negative or more than 2 days.

(b) Lag1 (RAC to hospital) is relevant for RAC leave episodes only. Lag1 unacceptable for AIHW links if (hospital in date – RAC out date) is negative or more than 2 days.

(c) Shading indicates CSLA-acceptable match—165 matches were CSLA-unacceptable on one date only.

Quality of HiL person links

Analysis of hospital events with mixed N and CSLA links (see Figure 6.2) indicated that in a small number of cases the N event linkage processes resulted in linking a hospital episode to an earlier, but close, RAC admission rather than the desired RAC hospital leave event (see note 3 to Table 4.1). Apart from the events identified among the mixed links, there were an additional 9 N-only links where N linkage matched a hospital event to a RAC admission just before the hospital event. Also, seven links to admissions were between hospital events and a RAC admission recorded as starting more than 3 days before the end of the hospital episode. For links to RAC leave events, three of the N-only links had non-overlapping hospital and RAC event dates, and 40 had more than a week's gap between the end of the hospital event and RAC leave dates. This latter is only conceptually valid for links to RAC social leave.

Overall there were 55 N-only problematic matches:

- links where the hospital episode did not overlap the matched RAC hospital leave
- links where the matched RAC admission started before the hospital episode

- links where the matched RAC admission began more than a week after the end of the hospital episode.

According to the RAC data, the people with these events had a total of 190 RAC events. On manual inspection, the N links were the favoured event matches in 49 of the 55 cases. Despite this, in a number of cases the chosen link matched very poorly on event dates, with the end of a RAC hospital leave event often being many days (even more than a month) after the end of the hospital episode. In at least one case, the data suggest that the person was in hospital and RAC at the same time.

A likely source of this error is the automated algorithm used to select one of the many possible linked events for individuals. Another possible cause is existence of errors in the HiL person links. The accuracy (false positives or mis-matches) of HiL person links has been assessed as 99.9% and 99.7% (Holman et al. 1999; Rosman et al. 2002). The completeness (missed matches or false negatives) has also been estimated as 99.9%. However, these estimates predate the extensive linkages to Australian Government Medicare and aged care clients data, which may have positively affected the completeness through wider population coverage, but negatively affected the accuracy due to increased volume of data processing. This hypothesis has not yet been tested. One factor that may have contributed to reduced accuracy and completeness of WADLS to RAC person links is the lack of specific event information (that is, event date and type) on the RAC records used to create links. Event information often assists the creation of person links (for example, within hospital records and between emergency and hospital records) when the other personal information is missing or inconsistent.

A4.2 CSLA false links

If it is assumed that N matches are highly likely to be correct and comprehensive, CSLA-only matches represent false matches made by the strategy. In the following analysis of the CSLA-only matches, the comparisons are generally based on the refined CSLA strategy retaining 2-digit postcode matching, and, as for N-only links, exclude any CSLA-only links that had a related link under N linkage.

Excluding mixed matches, there were 160 CSLA-only matches when using the refined CSLA strategy (Table A4.5). The majority of these were for admissions (100, or 63%), and one-third were for RAC hospital leave events (53); the remaining seven related to RAC social leave. Excluding 2-digit postcode matching from the strategy, the number of CSLA-only matches (excluding mixed matches) dropped to 136. All but one of the 24 links dropped related to RAC admissions, reflecting that 299 of 342 links made via 2-digit postcode were for admissions.

Around 15% of the CSLA-only matches to RAC admissions were for people that the hospital data recorded as going to another health facility; the remainder were fairly evenly split between those reported as going to a RAC facility and going to their usual residence. In general, there seems to be considerable confusion in the hospital data when reporting the discharge destination for people being admitted to RAC: among all CSLA links to RAC admissions, 1,701 had other health or RAC facility as their recorded post-hospital destination compared with 662 who were reported as going to their usual residence (which is what the RAC facility becomes if it is a permanent admission).

Table A4.5: CSLA-only links,[(a)] by RAC event type and hospital separation mode

	Hospital separation mode			
RAC event type	**To RAC**	**To other health care establishment**	**To usual residence/other**	**Total**
Accepting 2-digit postcode links				
Admissions	42	15	43	100
Hospital leave	3	—	50	53
Social leave	—	—	7	7
Total	**45**	**15**	**100**	**160**
Excluding 2-digit postcode links				
Admissions	35	12	30	77
Hospital leave	3	—	49	52
Social leave	—	—	7	7
Total	**38**	**12**	**86**	**136**

(a) Table excludes mixed CSLA–N links.

For CSLA-only links to RAC leave events, nearly all the matched hospital episodes reported that the person was returning to their usual residence, with just three out of 60 recorded as going into a RAC facility. Among all matches to RAC hospital leave, 4,235 were for people reported as discharged to their usual residence, compared with 843 discharged to another health facility or to a RAC facility. The latter may be valid if the person changes RAC facility on leaving the hospital.

Looking at the exactness of matching among the CSLA-only links suggests that many of the mismatches were caused by similar people living in a particular region—in terms of date of birth and sex (Table A4.6). Of the 100 RAC admissions linked only under CSLA, 94 had exactly matching dates of birth with their hospital record counterpart, and 86 had exactly matching event dates. Similar patterns were apparent for all levels of postcode matching. In addition, two-thirds of the links matched on at least the first 3 digits of the postcode, with more than half having exact postcode matches.

Table A4.6: Matches made by CSLA only: indicators of variable matching leading to likely false match, by RAC event type

Postcode matching	Exact date of birth	Different date of birth	Lag2 = 0	Lag2 ≠ 0	All	Lag1 = 0	Lag1 ≠ 0
Admission							
Same postcode	51	4	47	8	55	. .	. .
Same first 3 digits of postcode only	9	1	9	1	10	. .	. .
Same first 2 digits of postcode only	31	1	27	5	(a)32	. .	. .
Same on SLA group only	3	—	3	—	3	. .	. .
Total	*94*	*6*	*86*	*14*	*100*	. .	. .
RAC leave							
Same postcode	44	3	45	2	47	39	8
Same first 3 digits of postcode only	5	—	4	1	5	3	2
Same first 2 digits of postcode only	5	1	2	4	(b)6	3	3
Same on SLA group only	1	1	1	1	2	0	2
Total	*55*	*5*	*52*	*8*	*60*	*45*	*15*
All							
Same postcode	95	7	92	10	102	. .	. .
Same first 3 digits of postcode only	14	1	13	2	15	. .	. .
Same first 2 digits of postcode only	36	2	29	9	38	. .	. .
Same on SLA group only	4	1	4	1	5	. .	. .
Total	**149**	**11**	**138**	**22**	**160**	**. .**	**. .**

(a) Includes nine that also matched on SLA group.

(b) Includes five that also matched on SLA group.

Links to RAC leave also had a high percentage with exact matches on date of birth and hospital exit/RAC entry date (and to a lesser extent on RAC exit/hospital entry date). However, this pattern was not obvious in the very small numbers of links with poorly matching postcodes (that is, with only the first 2 digits matching, or matching on SLA group only).

The above comparisons do not readily suggest a way for reducing the number of false matches made under the CSLA strategy without losing a large number of true matches. Table A4.6 suggests that one of the most effective ways to reduce the number of false matches made under the CSLA strategy would be to reduce the size of the geographic region used in matching. For example, excluding 2-digit postcode matching would lead to dropping about 25% of the CSLA false matches, and excluding 3-digit postcode matching as well would increase this to one-third. However, narrowing the acceptable matching rules to lower the number of false matches could result in too many missed matches for little gain. For example, changing the strategy from refined CSLA matching (allowing 2-digit postcode matches) to exact matching on date of birth, sex, postcode and hospital exit/RAC entry date, would reduce the number of links from 7,595 to 5,559 (see Tables 6.3 and 6.10), while the PPV would increase only marginally—from 98.2% to 98.4%.

To confirm that many of the CSLA-only links were caused by similar people living in a particular region, the HiL-person links for this set of RAC records were re-investigated. Investigations by HiL of the client links implied by the CSLA-only links led to HiL identifying just two additional person links between hospital and RAC clients that had previously been missed.

Appendix 5: Analysis of population size of regions defined variously in terms of postcode

The following analysis shows the distribution of older people across postcode-based regions at the time of the 2001 population census. The population size of the region within which matching is being undertaken is an important consideration when undertaking event-based data linkage.

A5.1 Complete postcodes (4-digit)

In 2001, all postcodes had fewer than 7,500 men or women aged 65 years and over, with 97% having fewer than 2,500 (Tables A5.1 and A5.2). In addition, nearly 98% of people lived in postcodes with fewer than 5,000 older people of a particular sex (Tables A5.3 and A5.4).

Table A5.1: Distribution of postcodes, by size of population of men aged 65 years and over, by state/territory, 2001

Men aged 65+	State/territory								Total
	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	
0	0.5	1.1	0.5	3.5	0.6	—	3.8	—	1.1
1–2,500	97.2	98.2	96.3	96.2	99.4	99.1	96.2	100.0	97.5
2,501–5,000	2.3	0.8	3.0	0.3	—	0.9	—	—	1.4
5,001–7,500	—	—	0.2	—	—	—	—	—	—
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**598**	**653**	**405**	**312**	**311**	**106**	**26**	**25**	**2,436**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.2: Distribution of postcodes, by size of population of women aged 65 years and over, by state/territory, 2001

Women aged 65+	State/territory								Total
	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	
0	0.7	1.1	1.5	5.4	0.6	0.9	3.8	—	1.6
1–2,500	94.8	96.8	93.6	93.6	99.4	98.1	96.2	100.0	95.8
2,501–5,000	4.0	2.1	4.4	1.0	—	0.9	—	—	2.5
5,001–7,500	0.5	—	0.5	—	—	—	—	—	0.2
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**598**	**653**	**405**	**312**	**311**	**106**	**26**	**25**	**2,436**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.3: Distribution of men aged 65 years and over across postcodes, by postcode size, by state/territory, 2001

Men aged 65+	State/territory								Total
	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	
1–2,500	88.1	94.0	78.2	95.7	100.0	90.6	100.0	100.0	89.6
2,501–5,000	11.9	6.0	19.3	4.3	—	9.4	—	—	9.9
5,001–7,500	—	—	2.5	—	—	—	—	—	0.5
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**361,668**	**253,002**	**201,607**	**91,452**	**91,214**	**26,778**	**12,582**	**5,645**	**1,043,948**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.4: Distribution of women aged 65 years and over across postcodes, by postcode size, by state/territory, 2001

Women aged 65+	State/territory								Total
	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	
1–2,500	80.3	85.9	69.3	91.1	100.0	89.3	100.0	100.0	83.0
2,501–5,000	16.2	14.1	25.2	8.9	—	10.7	—	—	14.8
5,001–7,500	3.4	—	5.5	—	—	—	—	—	2.2
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**466,335**	**332,926**	**241,534**	**113,256**	**120,045**	**34,516**	**16,295**	**5,044**	**1,329,951**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

A5.2 First 3 digits of postcode

Areas defined by the first 3 digits of a postcode had fewer than 20,000 men and 30,000 women aged 65 years and over in 2001 (Tables A5.5 and A5.6). Just under 99% had fewer than 15,000 older women, and over 99% had under 15,000 men. Just over 98% of older men lived in 3-digit postcode areas with fewer than 15,000 men aged 65 years and over; 92% of older women lived in regions of this size, with nearly 2% living in areas with between 20,000 and 30,000 older women Tables A5.7 and A5.8).

Table A5.5: Distribution of 3-digit postcodes, by size of population of men aged 65 years and over, by state/territory, 2001

	State/territory								
Men aged 65+	**NSW**	**Vic**	**Qld**	**WA**	**SA**	**Tas**	**ACT**	**[(a)]NT**	**Total**
0	—	—	—	—	—	—	—	—	—
1–2,500	41.4	75.0	60.9	84.8	80.6	85.7	60.0	100.0	70.5
2,501–5,000	25.3	6.0	14.5	4.5	10.4	14.3	40.0	—	12.1
5,001–7,500	16.1	6.0	20.3	7.6	7.5	—	—	—	9.8
7,501–10,000	10.3	10.0	—	3.0	1.5	—	—	—	4.9
10,001–15,000	6.9	3.0	2.9	—	—	—	—	—	2.5
15,001–20,000	—	—	1.4	—	—	—	—	—	0.2
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**87**	**100**	**69**	**66**	**67**	**28**	**5**	**25**	**447**

(a) The Northern Territory only has 3 digits in its postcodes.

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.6: Distribution of 3-digit postcodes, by size of population of women aged 65 years and over, by state/territory, 2001

	State/territory								
Women aged 65+	**NSW**	**Vic**	**Qld**	**WA**	**SA**	**Tas**	**ACT**	**[(a)]NT**	**Total**
0	—	—	—	1.5	—	—	—	—	0.2
1–2,500	36.8	69.0	56.5	78.8	76.1	85.7	60.0	100.0	66.0
2,501–5,000	14.9	9.0	10.1	7.6	9.0	7.1	—	—	9.4
5,001–7,500	20.7	6.0	21.7	1.5	7.5	7.1	40.0	—	11.0
7,501–10,000	10.3	4.0	7.2	6.1	4.5	—	—	—	5.6
10,001–15,000	14.9	9.0	2.9	4.5	3.0	—	—	—	6.5
15,001–20,000	2.3	3.0	—	—	—	—	—	—	1.1
20,001–30,000	—	—	1.4	—	—	—	—	—	0.2
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**87**	**100**	**69**	**66**	**67**	**28**	**5**	**25**	**447**

(a) The Northern Territory only has 3 digits in its postcodes.

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.7: Distribution of men aged 65 years and over across postcodes, by 3-digit postcode size, by state/territory, 2001

	State/territory								
Men aged 65+	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Total
1–2,500	11.1	28.3	18.5	32.2	31.6	47.8	27.0	100.0	22.0
2,501–5,000	24.0	8.9	20.0	12.1	28.4	52.2	73.0	—	20.1
5,001–7,500	23.5	14.1	42.9	37.2	30.6	—	—	—	25.8
7,501–10,000	22.5	34.6	—	18.5	9.4	—	—	—	18.6
10,001–15,000	18.8	14.1	10.0	—	—	—	—	—	11.9
15,001–20,000	—	—	8.6	—	—	—	—	—	1.7
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**361,668**	**253,002**	**201,607**	**91,452**	**91,214**	**26,778**	**12,582**	**5,645**	**1,044,006**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.8: Distribution of women aged 65 years and over across postcodes, by 3-digit postcode size, by state/territory, 2001

	State/territory								
Women aged 65+	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Total
1–2,500	7.9	21.1	12.9	19.7	21.1	43.5	26.3	100.0	15.8
2,501–5,000	10.5	9.0	10.0	15.4	15.2	23.7	—	—	11.1
5,001–7,500	23.0	11.7	40.9	4.5	24.5	32.8	73.7	—	22.8
7,501–10,000	16.8	10.7	17.8	31.8	19.7	—	—	—	16.3
10,001–15,000	34.8	32.8	9.7	28.6	19.4	—	—	—	26.4
15,001–20,000	7.0	14.7	—	—	—	—	—	—	6.1
20,001–30,000	—	—	8.7	—	—	—	—	—	1.6
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**466,335**	**332,926**	**241,534**	**113,256**	**120,045**	**34,516**	**16,295**	**5,044**	**1,329,951**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

A5.3 First 2 digits of postcode

Using the first 2 digits of a postcode to define a region results in considerable aggregation of the population: in 2001 3% of 2-digit postcode areas had over 70,000 older men, and 9% had more than 70,000 older women (maximum of 87,250 men and 121,400 women, both in Victoria). The distribution varied considerably from state to state, with only New South Wales and Victoria having areas with populations greater than 50,000 people aged 65 years and over of either sex (Tables A5.9 and A5.10).

Looking at the distribution of the population across 2-digit postcodes, a number of states had quite a large proportion of their population living in the 2-digit postcodes with more than

30,000 people of one sex (Tables A5.11 and A5.12): New South Wales (93% of older women), Queensland (64%), Victoria (60%) and Western Australia (76%).

Table A5.9: Distribution of 2-digit postcodes, by size of population of men aged 65 years and over, by state/territory, 2001

	State/territory								
Men aged 65+	**NSW**	**Vic**	**Qld**	**WA**	**SA**	**Tas**	**ACT**	**[(a)]NT**	**Total**
1–2,500	10.0	—	10.0	22.2	33.3	33.3	50.0	100.0	30.3
2,501–5,000	—	—	10.0	44.4	33.3	16.7	—	—	13.6
5,001–7,500	—	10.0	—	—	—	33.3	—	—	4.5
7,501–10,000	—	10.0	—	—	—	16.7	—	—	3.0
10,001–15,000	20.0	40.0	10.0	11.1	11.1	—	50.0	—	15.2
15,001–20,000	0.0	10.0	20.0	—	11.1	—	—	—	6.1
20,001–30,000	20.0	10.0	20.0	11.1	—	—	—	—	9.1
30,001–40,000	—	—	30.0	11.1	—	—	—	—	6.1
40,001–50,000	20.0	—	—	—	11.1	—	—	—	4.5
50,001–60,000	10.0	—	—	—	—	—	—	—	1.5
60,001–70,000	10.0	10.0	—	—	—	—	—	—	3.0
70,001–high	10.0	10.0	—	—	—	—	—	—	3.0
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**10**	**10**	**10**	**9**	**9**	**6**	**2**	**10**	**66**

(a) The Northern Territory only has 3 digits in its postcodes.

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.10: Distribution of 2-digit postcodes, by size of population of women aged 65 years and over, by state/territory, 2001

	State/territory								
Women aged 65+	**NSW**	**Vic**	**Qld**	**WA**	**SA**	**Tas**	**ACT**	**[(a)]NT**	**Total**
1–2,500	10.0	—	10.0	22.2	33.3	33.3	0.0	100.0	28.8
2,501–5,000	—	—	10.0	33.3	22.2	16.7	50.0	—	12.1
5,001–7,500	—	—	—	11.1	11.1	—	0.0	—	3.0
7,501–10,000	—	10.0	—	—	0.0	33.3	0.0	—	4.5
10,001–15,000	—	20.0	10.0	11.1	11.1	16.7	50.0	—	10.6
15,001–20,000	20.0	30.0	10.0	—	—	—	—	—	9.1
20,001–30,000	—	20.0	20.0	—	11.1	—	—	—	7.6
30,001–40,000	20.0	—	30.0	11.1	—	—	—	—	9.1
40,001–50,000	—	—	10.0	11.1	—	—	—	—	3.0
50,001–60,000	20.0	—	—	—	—	—	—	—	3.0
60,001–70,000	—	—	—	—	—	—	—	—	—
70,001–high	30.0	20.0	—	—	11.1	—	—	—	9.1
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**10**	**10**	**10**	**9**	**9**	**6**	**2**	**10**	**66**

(a) The Northern Territory only has 3 digits in its postcodes.

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.11: Distribution of men aged 65 years and over across postcodes, by 2-digit postcode size, by state/territory, 2001

	State/territory								
Men aged 65+	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Total
1–2,500	—	—	0.1	0.8	3.8	1.2	15.8	100.0	1.2
2,501–5,000	—	—	2.0	16.8	12.7	10.7	—	—	3.2
5,001–7,500	—	2.9	—	—	—	50.8	—	—	2.0
7,501–10,000	—	3.4	—	—	—	37.3	—	—	1.8
10,001–15,000	7.5	19.7	6.8	11.8	11.1	—	84.2	—	11.7
15,001–20,000	—	6.9	17.5	—	18.4	—	—	—	6.7
20,001–30,000	13.9	8.6	27.2	31.5	0.0	—	—	—	14.9
30,001–40,000	—	—	46.5	39.1	0.0	—	—	—	12.4
40,001–50,000	24.5	—	—	—	54.1	—	—	—	13.2
50,001–60,000	14.8	—	—	—	—	—	—	—	5.1
60,001–70,000	19.3	23.9	—	—	—	—	—	—	12.5
70,001–high	20.0	34.5	—	—	—	—	—	—	15.3
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**361,668**	**253,002**	**201,607**	**91,452**	**91,214**	**26,778**	**12,582**	**5,645**	**1,043,948**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

Table A5.12: Distribution of women aged 65 years and over across postcodes, by 2-digit postcode size, by state/territory, 2001

	State/territory								
Women aged 65+	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Total
1–2,500	—	—	0.1	0.3	3.1	0.7	—	100.0	0.7
2,501–5,000	—	—	1.7	8.8	7.0	8.3	15.5	—	2.1
5,001–7,500	—	—	—	4.6	4.3	—	—	—	0.8
7,501–10,000	—	2.5	—	—	—	50.8	—	—	2.0
10,001–15,000	—	7.7	6.2	10.8	10.2	40.1	84.5	—	7.0
15,001–20,000	7.2	14.9	7.7	—	—	—	—	—	7.6
20,001–30,000	—	15.2	20.0	—	17.0	—	—	—	9.0
30,001–40,000	13.9	—	46.4	34.0	—	—	—	—	16.2
40,001–50,000	—	—	17.9	41.5	—	—	—	—	6.8
50,001–60,000	22.6	—	—	—	—	—	—	—	7.9
60,001–70,000	—	—	—	—	—	—	—	—	—
70,001–high	56.4	59.7	—	—	58.5	—	—	—	40.0
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**466,335**	**332,926**	**241,534**	**113,256**	**120,045**	**34,516**	**16,295**	**5,044**	**1,329,951**

Note: Three postcodes are in two states (ACT and NSW): 2618, 2619, 2620 with total populations aged 65+ years of 85, 32 and 3,165. The postcode population has not been split among its constituent states, and so these are included twice in the total.

Source: AIHW analysis of CDATA 2001 (ABS 2002).

A5.4 SLA group

The SLA group for a postcode is that set of SLAs that overlap the postcode (see Figure 5.1). In general, both postcodes and SLAs do not go across state boundaries (except for three postcodes that go across the Australian Capital Territory–New South Wales border). When comparing postcodes, two postcodes are said to match on SLA group if they have a common SLA in their respective SLA groups.

Looking at all postcode pairs that have overlapping SLA groups (excluding identical postcodes), in 2001 95% had fewer than 20,000 older men and 92% had fewer than 20,000 older women (Tables A5.13 and A5.14). Only New South Wales had any postcode pairs with more than 30,000 older men, and New South Wales, Victoria and Western Australia were the only states to have SLA groups with between 20,000 and 30,000 older women.

The distribution of postcode pairs by SLA group population size was between those for 3- and-2 digit postcodes: the majority of postcode pairs had combined SLA groups of less than 5,000 older men or women (63% for older men and 56% for older women), few had more than 40,000, but a substantial number had between 5,000 and 30,000 older people of one sex (36% for older men and 42% for older women).

Because postcode SLA groups are not mutually exclusive, it is not possible to derive the distribution of the population by size of SLA group.

Table A5.13: Distribution of matching postcode pairs, by size of SLA-group population of men aged 65 years and over, by state/territory of first postcode, 2001

	State/territory								
Men aged 65+	**NSW**	**Vic**	**Qld**	**WA**	**SA**	**Tas**	**ACT**	**NT**	**Total**
0–5,000	38.0	71.0	79.1	66.5	89.6	94.0	100.0	100.0	63.2
5,001–10,000	19.0	22.1	20.7	20.7	10.2	6.0	—	—	18.0
10,001–20,000	29.9	6.9	0.1	12.8	0.3	—	—	—	13.9
20,001–30,000	10.9	—	—	—	—	—	—	—	4.0
30,001–40,000	2.0	—	—	—	—	—	—	—	0.7
40,001–50,000	0.2	—	—	—	—	—	—	—	0.1
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**3,569**	**2,550**	**844**	**656**	**1,515**	**403**	**19**	**45**	**9,601**

Notes

1. For consistency with above tables, the table is based on the postcodes in CDATA 2001. The three postcodes that are in two states—2618, 2619, 2620 in ACT and NSW—are represented only once (chosen according to the alphabetic ordering of state abbreviations, i.e. in ACT).
2. SLA groups are for two different postcodes. Only distinct pairs are included in the table (i.e. postcode order is ignored).
3. Excludes postcode pairs without common SLAs.

Sources: AIHW analysis of CDATA 2001 (ABS 2002), using postcode–SLA concordances (ABS unpublished data).

Table A5.14: Distribution of matching postcode pairs, by size of SLA-group population of women aged 65 years and over, by state/territory of first postcode, 2001

Women aged 65+	State/territory								Total
	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	
0–5,000	32.8	60.4	75.5	60.5	83.0	78.7	89.5	100.0	56.1
5,001–10,000	16.7	25.0	23.8	19.7	14.3	21.3	10.5	—	19.5
10,001–20,000	29.6	13.8	0.7	18.1	2.6	—	—	—	16.4
20,001–30,000	14.3	0.8	—	1.7	—	—	—	—	5.6
30,001–40,000	4.3	—	—	—	—	—	—	—	1.6
40,001–50,000	2.0	—	—	—	—	—	—	—	0.7
50,001–60,000	0.2	—	—	—	—	—	—	—	0.1
60,001–70,000	0.1	—	—	—	—	—	—	—	—
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Number	**3,569**	**2,550**	**844**	**656**	**1,515**	**403**	**19**	**45**	**9,601**

Notes

1. For consistency with above tables, the table is based on the postcodes in CDATA 2001. The three postcodes that are in two states—2618, 2619, 2620 in ACT and NSW—are represented only once (chosen according to the alphabetic ordering of state abbreviations, i.e. in ACT).
2. SLA groups are for two different postcodes. Only distinct pairs are included in the table (i.e. postcode order is ignored).
3. Excludes postcode pairs without common SLAs.

Sources: AIHW analysis of CDATA 2001 (ABS 2002), using postcode–SLA concordances (ABS unpublished data).

A5.5 Conclusion

Noting that

- false match rates are expected to increase with increasing population size when matching using an event-based strategy, and
- theoretical analysis has indicated that false match rates may get unacceptably high within populations of more than 70,000 (AIHW: Karmel 2004)

the above analysis suggests the following for the E linkage strategy:

1. matching within 4-digit postcode will result in matching within **very small** population groups (generally within populations of under 5,000 older people of one sex)
2. matching within 3-digit postcode will generally result in matching within **acceptably small** population groups (within populations of under 20,000 older people of one sex for more than 98% of the older population)
3. matching within 2-digit postcode could result in matching within **unacceptably large** population groups for some states (in 2001 more than 40% of the older female population in New South Wales, South Australia and Victoria lived within populations of over 70,000 older women when grouped by 2-digit postcode)
4. matching via SLA groups for postcode pairs will generally result in matching within **acceptably small** regions, with the combined SLA group for a postcode pair (different postcodes) always being less than 30,000 older women for all states except New South Wales, where 7% of combined SLA groups had populations of between 30,000 and 70,000 older women.

Appendix 6: Additional linkage comparison tables

Table A6.1: Positive predictive value for CSLA$_s$, by RAC event type and hospital mode of admission

	Hospital admission mode		
RAC event type	**Statistical**	**Non-statistical**	**Total**
		Number of true matches by strategy	
Permanent admission	608	688	1,296
Respite admission	207	485	692
Hospital leave	602	4,360	4,962
Social leave	17	133	150
Total	**1,434**	**5,666**	**7,100**
		Total number of matches by strategy	
Permanent admission	647	713	1,360
Respite admission	213	491	704
Hospital leave	610	4,425	5,035
Social leave	19	135	154
Total	**1,489**	**5,764**	**7,253**
		PPV (per cent)	
Permanent admission	94.1	95.9	95.1
Respite admission	97.2	98.6	98.2
Hospital leave	98.5	98.7	98.7
Social leave	89.5	96.3	95.5
Total	**96.3**	**98.3**	**97.9**

Table A6.2: Positive predictive value for $CSLA_s$, by RAC event type and hospital mode of separation[(a)]

RAC event type	Hospital separation mode				Total
	To RAC	To other health care establishment	Left against medical advice/statistical discharge from leave/unknown	To usual residence/other	
			Number of true matches by strategy		
Permanent admission	851	176	4	265	1,296
Respite admission	134	253	8	297	692
Hospital leave	458	359	11	4,134	4,962
Social leave	19	9	1	121	150
Total	**1,462**	**797**	**24**	**4,817**	**7,100**
			Total number of matches by strategy		
Permanent admission	889	182	4	285	1,360
Respite admission	137	258	8	301	704
Hospital leave	458	361	11	4,205	5,035
Social leave	18	9	1	126	154
Total	**1,502**	**810**	**24**	**4,917**	**7,253**
			PPV (per cent)		
Permanent admission	96.2	96.7	100.0	90.5	95.1
Respite admission	97.8	97.7	100.0	98.7	98.2
Hospital leave	99.3	99.7	100.0	98.5	98.7
Social leave	100.0	100.0	100.0	94.4	95.5
Total	**97.3**	**98.4**	**100.0**	**98.0**	**97.9**

(a) $CSLA_s$ strategy excluded matching to hospital episodes recorded as ending in death. Consequently, this mode of separation does not appear in the table.

Table A6.3: Positive predictive value for $CSLA_s$, by RAC event type and year of birth (age in 2000)

	Year of birth					
RAC event type	**<1900 (>100 yrs)**	**1900–09 (91–100 yrs)**	**1910–19 (81–90 yrs)**	**1920–29 (71–80 yrs)**	**1935–39 (65–70 yrs)**	**Total**
			Number of true matches by strategy			
Permanent admission	—	230	625	371	70	1,296
Respite admission	1	69	325	244	53	692
Hospital leave	21	911	2,541	1,244	245	4,962
Social leave	—	21	80	37	12	150
Total	**22**	**1,231**	**3,571**	**1,896**	**380**	**7,100**
			Column per cent of true matches			
Permanent admission	—	18.9	17.9	20.3	19.3	18.8
Respite admission	4.5	5.5	9.1	12.8	13.9	9.7
Hospital leave	95.5	73.9	70.8	64.9	63.5	69.4
Social leave	—	1.7	2.2	2.0	3.3	2.1
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
			Total number of matches by strategy			
Permanent admission	—	237	652	396	75	1,360
Respite admission	1	69	330	250	54	704
Hospital leave	21	924	2,575	1,268	247	5,035
Social leave	—	21	80	40	13	154
Total	**22**	**1,251**	**3,637**	**1,954**	**389**	**7,253**
			PPV (per cent)			
Permanent admission	—	97.5	95.6	92.9	94.7	95.1
Respite admission	100.0	98.6	98.5	97.6	98.2	98.2
Hospital leave	100.0	98.6	98.8	98.4	98.8	98.7
Social leave	—	100.0	97.5	90.0	92.3	95.5
Total	**100.0**	**98.4**	**98.2**	**97.0**	**97.7**	**97.9**

Table A6.4: Sensitivity of strategy for CSLAs, by RAC event type and hospital mode of admission

RAC event type	Hospital admission mode		Total
	Statistical	Non-statistical	
		N matches	
Permanent admission	814	909	1,723
Respite admission	246	606	852
Hospital leave	695	4,675	5,370
Social leave	18	143	161
Total	**1,773**	**6,333**	**8,106**
		CSLAs sensitivity (per cent)	
Permanent admission	74.7	75.7	75.2
Respite admission	84.1	80.0	81.2
Hospital leave	86.6	93.3	92.4
Social leave	94.4	93.0	93.2
Total	**80.9**	**89.5**	**87.6**

Source: Table 7.4.

Table A6.5: Sensitivity of strategy for CSLAs, by RAC event type and year of birth

RAC event type	Year of birth					Total
	<1900 (>100 yrs)	1900–09 (91–100 yrs)	1910–19 (81–90 yrs)	1920–29 (71–80 yrs)	1935–39 (65–70 yrs)	
			N matches			
Permanent admission	—	298	842	488	95	1,723
Respite admission	3	83	400	299	67	852
Hospital leave	25	973	2,746	1,360	266	5,370
Social leave	—	22	86	41	12	161
Total	**28**	**1,376**	**4,074**	**2,188**	**440**	**8,106**
			CSLAs sensitivity (per cent)			
Permanent admission	—	77.2	74.2	76.0	73.7	75.2
Respite admission	33.3	83.1	81.3	81.6	79.1	81.2
Hospital leave	84.0	93.6	92.5	91.5	92.1	92.4
Social leave	—	95.5	93.0	90.2	100.0	93.2
Total	78.6	89.5	87.7	86.7	86.4	87.6

Source: Table 7.6.

Table A6.6: Independent variables included in logistic regressions for modelling N links missed by CSLA$_s$, N links, 2000–01

Variable	Value	Missed	[(a)]Linked	Total
		Row per cent		Number
Sex	Female	12.3	87.7	5,493
	Male	12.7	87.3	2,611
Age at RAC (re-) entry	60–64	6.5	93.5	31
	65–69	15.3	84.7	313
	70–79	13.3	86.7	1,946
	80–89	12.5	87.5	4,078
	90–99	10.8	89.2	1,692
	≥100	13.6	86.4	44
Marital status (on RAC)	De facto	8.3	91.7	24
	Divorced	11.2	88.8	401
	Married	12.6	87.4	1,714
	Never married	12.8	87.2	539
	Separated	14.4	85.6	132
	Unknown	9.8	90.2	153
	Widowed	12.4	87.6	5,141
Country of birth (on RAC)	Australia	12.4	87.6	5,083
	New Zealand/Oceania	10.7	89.3	75
	United Kingdom/Ireland	11.1	88.9	1,784
	Europe	14.8	85.2	722
	Asia	13.0	87.0	247
	Other/missing	16.1	83.9	193
RAC event type	Permanent admission	24.8	75.2	1,721
	Respite admission	18.8	81.2	852
	Hospital leave	7.6	92.4	5,370
	Social leave	6.8	93.2	161
Hospital separation mode	To RAC	17.5	82.5	1773
	To usual residence/other	9.5	90.5	5,319
	To other health care establishment	16.0	84.0	949
	Left against medical advice/statistical discharge on leave/unknown/death	61.9	38.1	63
Linked to unusual/unknown hospital separation mode for transfers (group E above)*	0 (no)	12.0	88.0	8,041
	1 (yes)	61.9	38.1	63
Hospital admission mode	Non-statistical	10.5	89.5	6,332
	Statistical	19.1	80.9	1,772

(continued)

Table A6.6 (continued): Independent variables included in logistic regressions for modelling N links missed by CSLAs, N links, 2000–01

Variable	Value	Missed	[(a)]Linked	Total
		Row per cent		Number
Location of ACAT assessment	Aged care facility	9.7	90.3	901
	At home	12.1	87.9	1,451
	Hospital	13.9	86.1	4,140
	Missing	8.5	91.5	1,282
	Other	18.2	81.8	330
RCS appraisal category on RAC (re-) entry	Missing	18.3	81.7	996
	S1	14.8	85.2	1,061
	S2	13.7	86.3	1,362
	S3	12.2	87.8	941
	S4	11.6	88.4	318
	S5	12.4	87.6	974
	S6	9.2	90.8	1,045
	S7	8.1	91.9	1,304
	S8	5.8	94.2	103
Principal diagnosis	Certain infectious & parasitic diseases (A00–B99)	10.9	89.1	101
	Neoplasms (C00–D48)	9.8	90.2	407
	Blood & blood-forming organs (D50–D89)	3.8	96.2	131
	Endocrine, nutritional, metabolic & immunity (E00–E90)	7.6	92.4	197
	Mental disorders (F00–F99)	21.2	78.8	472
	Nervous system & sense organs (G00–G99)	17.5	82.5	382
	Eye & adnexa (H00–H59)	6.6	93.4	122
	Ear & mastoid process (H60–H95)	0.0	100.0	12
	Cardiovascular disease (I00–I99)	10.1	89.9	1,085
	Respiratory system (J00–J99)	9.9	90.1	858
	Digestive system (K00–K93)	7.9	92.1	570
	Skin & subcutaneous tissue (L00–L99)	8.2	91.8	208
	Musculoskeletal system (M00–M99)	12.3	87.7	317
	Genitourinary system (N00–N99)	9.4	90.6	426
	Congenital anomalies (Q00–Q99)	0.0	100.0	1
	Symptoms, sign & ill-defined conditions (R00–R99)	11.1	88.9	539
	Injury & poisoning (S00–T98)	10.4	89.6	894
	Z code (Z00–Z99)	19.6	80.4	1,382
Dementia as principal diagnosis[(b)]	0 (no)	11.9	88.1	7,699
	1 (yes)	22.7	77.3	405

(continued)

Table A6.6 (continued): Independent variables included in logistic regressions for modelling N links missed by CSLA$_s$, N links, 2000–01

Variable	Value	Missed	[(a)]Linked	Total
		Row per cent		Number
Dementia as any diagnosis[(b)]	0 (no)	11.5	88.5	5,399
	1 (yes)	14.2	85.8	2,705
Any diagnoses include external causes[(b)]	0 (no)	12.3	87.7	5,795
	1 (yes)	12.6	87.4	2,309
Any diagnoses include 'Factors influencing health status etc.'[(b)]	0 (no)	11.2	88.8	7,233
	1 (yes)	22.3	77.7	871
Major diagnostic category (MDC)	Diseases and disorders of nervous system	17.1	82.9	1,123
	Diseases and disorders of eye	6.7	93.3	150
	Diseases and disorders of ear, nose, mouth & throat	4.6	95.4	87
	Diseases and disorders of respiratory system	10.2	89.8	903
	Diseases and disorders of circulatory system	9.4	90.6	1,015
	Diseases and disorders of digestive system	8.8	91.2	673
	Diseases and disorders of hepatobiliary system & pancreas	8.3	91.7	84
	Diseases and disorders of musculoskeletal system & connective tissue	11.6	88.4	854
	Diseases and disorders of skin, subcutaneous tissue & breast	9.0	91.0	402
	Endocrine, nutritional & metabolic diseases and disorders	6.0	94.0	166
	Diseases and disorders of kidney & urinary tract	9.7	90.3	475
	Diseases and disorders of male reproductive system	16.3	83.7	49
	Diseases and disorders of female reproductive system	7.9	92.1	38
	Diseases and disorders of blood & blood-forming organs & immunological disorders	3.8	96.2	132
	Neoplastic disorders	5.7	94.3	87
	Infectious & parasitic diseases	8.6	91.4	105
	Mental diseases and disorders	18.8	81.2	181
	Alcohol/drug use & induced organic mental disorders	36.4	63.6	11
	Injuries, poisonings etc.	8.0	92.0	174
	Factors influencing health status etc.	19.6	80.4	1,395

(continued)

Table A6.6 (continued): Independent variables included in logistic regressions for modelling N links missed by CSLAs, N links, 2000–01

Variable	Value	Missed	[(a)]Linked	Total
		Row per cent		Number
MDC is 'Factors influencing health status etc.'[(b)]	0 (no)	10.9	89.1	6,709
	1 (yes)	19.6	80.4	1,395
First 2 digits of person postcode (on RAC)	Perth	12.8	87.2	6,554
	South West	7.9	92.1	699
	Great Southern	9.9	90.1	303
	Goldfields	8.5	91.5	213
	Central coastal/Murchison	8.4	91.6	179
	North	24.4	75.6	131
	PO Box	77.8	22.2	9
	Not WA	56.3	43.8	16
First 2 digits of RAC outlet postcode	Perth	13.0	87.0	6,600
	South West	7.5	92.5	677
	Great Southern	11.1	88.9	315
	Goldfields	6.3	93.7	207
	Central Coastal/Murchison	7.4	92.6	176
	North	27.1	72.9	129
Locality of RAC facility	Inner regional	7.7	92.3	687
	Outer regional	9.0	91.0	700
	Remote	13.7	86.3	131
	Very remote	50.0	50.0	42
	Major city	13.0	87.0	6,544
RAC outlet in very remote region[(b)]	0 (no)	12.2	87.8	8,062
	1 (yes)	50.0	50.0	42
Poor quality person postcode data[(b)]	0 (no)	12.1	87.9	7,964
	1 (yes)	27.9	72.1	140
Person postcode in Goldfields region[(b)]	0 (no)	12.5	87.5	7,891
	1 (yes)	8.5	91.5	213
RAC facility postcode in Goldfields region[(b)]	0 (no)	12.6	87.4	7,897
	1 (yes)	6.3	93.7	207
Poor quality SLA data in hospital record	–1 (unable to be geo-coded)	9.9	90.1	364
	0 (geo-coded)	12.9	87.1	1,777
	1 (missing)	12.4	87.6	5,963

(continued)

Table A6.6 (continued): Independent variables included in logistic regressions for modelling N links missed by CSLAs, N links, 2000–01

Variable	Value	Missed	[(a)]Linked	Total
		Row per cent		Number
Hospital length of stay	1 day	7.7	92.3	621
	2 days	8.8	91.2	599
	3 days	8.4	91.6	607
	4–6 days	8.5	91.5	1,440
	7–9 days	8.7	91.3	984
	10–13 days	10.3	89.7	884
	14–20 days	12.8	87.2	950
	21–34 days	17.0	83.0	926
	≥35 days	25.1	74.9	1,095
Leave from hospital (number)	Missing	11.0	89.0	511
	0	11.8	88.2	7,450
	1	51.3	48.7	117
	≥2	39.3	60.7	28
Leave from hospital (days)	Missing	10.8	89.2	510
	0 days	11.8	88.2	7,451
	1 days	35.5	64.5	31
	2–6 days	49.4	50.6	87
	≥7 days	63.0	37.0	27
Total	. .	**12.4**	**87.6**	**8,104**

(a) Sensitivity.

(b) Derived binomial variable.

Note: Table excludes two cases due to missing data in locality of RAC facility.

Table A6.7: Results from logistic regressions for modelling BSESLA missed link status among N links, within RAC event type, 2000–01

Variable	Model for RAC permanent admissions	Model for RAC respite admissions	Model for RAC leave events
Intercept	-	***	-
Sex	-	-	-
Age at RAC (re-) entry	-	-	*
Marital status (on RAC)	**	-	-
Country of birth (on RAC)	-	-	-
Hospital separation mode	-	-	-
Unusual/unknown hospital separation mode*	*	-	***
Hospital care type	-	-	-
Hospital admission mode	*	-	-
Hospital length of stay (categorised)	-	-	*
Number of episodes of leave from hospital (categorised)	-	-	-
Days of leave from hospital (categorised)	***	-	**
Principal diagnosis	-	-	-
Dementia as principal diagnosis[(a)]	-	-	-
Dementia as any diagnosis[(a)]	-	-	-
Any diagnoses include external causes[(a)]	*	-	-
Any diagnoses include 'Factors influencing health status and other contacts with health services'[(a)]	-	-	-
Major diagnostic category (MDC)	-	-	-
MDC is 'Factors influencing health status and other contacts with health services'[(a)]	-	-	**
RAC event type	. .	. .	-
Location of ACAT assessment		-	-
RCS appraisal category on RAC (re-) entry		-	-
RAC outlet postcode region	***	-	**
RAC outlet in very remote region[(a)]	-	-	**
RAC facility postcode in Goldfields region[(a)]	-	-	-
Locality (DoHA) of RAC facility	-	-	**
Person postcode region (in RAC data)	(b)	(b)	-
Poor quality person postcode data[(a)]	(b)	(b)	-
Person postcode in Goldfields region[(a)]	(b)	(b)	-
Quality of SLA data in hospital record	*	-	-

(a) Derived binomial variable.

(b) Variable excluded from model due to complete separation of data points. Note that person postcode is highly associated with RAC facility postcode for linked records.

- Not statistically significant at 0.05 level.

* Significant at 0.01–<0.05 level.

** Significant at 0.0001–<0.01 level.

*** Significant at <0.0001 level.

Table A6.8: Relative standard error of binomial proportions, by proportion and sample size

Proportion (%)	Number of cases					
	100	200	500	750	1,000	3,000
	Relative standard error (%)					
2.5	62.4	44.2	27.9	22.8	19.7	11.4
5.0	43.6	30.8	19.5	15.9	13.8	8.0
10.0	30.0	21.2	13.4	11.0	9.5	5.5
15.0	23.8	16.8	10.6	8.7	7.5	4.3
20.0	20.0	14.1	8.9	7.3	6.3	3.7
25.0	17.3	12.2	7.7	6.3	5.5	3.2
30.0	15.3	10.8	6.8	5.6	4.8	2.8
35.0	13.6	9.6	6.1	5.0	4.3	2.5
40.0	12.2	8.7	5.5	4.5	3.9	2.2
45.0	11.1	7.8	4.9	4.0	3.5	2.0
50.0	10.0	7.1	4.5	3.7	3.2	1.8
55.0	9.0	6.4	4.0	3.3	2.9	1.7
60.0	8.2	5.8	3.7	3.0	2.6	1.5
65.0	7.3	5.2	3.3	2.7	2.3	1.3
70.0	6.5	4.6	2.9	2.4	2.1	1.2
75.0	5.8	4.1	2.6	2.1	1.8	1.1
80.0	5.0	3.5	2.2	1.8	1.6	0.9
85.0	4.2	3.0	1.9	1.5	1.3	0.8
90.0	3.3	2.4	1.5	1.2	1.1	0.6
95.0	2.3	1.6	1.0	0.8	0.7	0.4
97.5	1.6	1.1	0.7	0.6	0.5	0.3

Note: Standard errors (SEs) can be used to estimate the likely range of estimates of proportions. Using a normal approximation, the 95% confidence interval is given by (p +/– 1.96*SE), where SE ~ $(n(1–p)/p)^{1/2}$ for a proportion p and a total of n cases. The relative standard error RSE is (100*SE/p)%.

References

ABS (Australian Bureau of Statistics) 2002 . Census of Population and Housing: CDATA 2001—Full GIS, Australia, 2001. Canberra: ABS.

AIHW (Australian Institute of Health and Welfare) 2002. Residential aged care in Australia 2000–01: a statistical overview. Canberra: AIHW.

AIHW 2003. Interface between hospital and residential aged care: feasibility study on linking hospital morbidity and residential aged care data. Canberra: AIHW.

AIHW 2006. Data linkage and protecting privacy: a protocol for linking between two or more data sets held within the Australian Institute of Health and Welfare. Canberra: AIHW. Viewed 28 July 2006, <http://www.aihw.gov.au/dataonline/privacy_of_data.cfm>.

AIHW: Karmel R 2004. Linking hospital morbidity and residential aged care data: examining matching due to chance. Canberra: AIHW.

Brook EL, Rosman DL, Holman CDJ & Trutwein B 2005. Summary report: research outputs project, WA Data Linkage Unit (1995–2003). Perth: Department of Health, Western Australia.

DoHA (Department of Health and Ageing) 2005. Residential care manual. 22 September 2004. Canberra: DoHA. Viewed 26 October 2006, <http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/1CC3ACD213466762CA256F19000FC9A5/$File/rcmfull.pdf>.

Hoel PG 1971. Introduction to mathematical statistics. John Wiley & Sons

Holman CDJ, Bass AJ, Rouse I & Hobbs MST 1999. Population-based linkage of health records in Western Australia: development of a health services research linked database. Australian and New Zealand Journal of Public Health 23:453–9.

Kelman CW, Bass AJ & Holman CDJ 2002. Research use of linked health data: a best practice protocol. Australian and New Zealand Journal of Public Health 26:251–5.

Medical University of South Carolina 2001. Diagnostic tests. Viewed 2 June 2006, <http://www.musc.edu/dc/icrebm/diagnostictests.html>.

NCCH (National Centre for Classification in Health) 1998. The international statistical classification of diseases and related health problems, 10th revision, Australian modification (ICD-10-AM). Sydney: University of Sydney.

Rosman DL 1996. The feasibility of linking hospital and police road crash information without names. Accident Analysis and Prevention 28(2):271–4.

Rosman DL, Garfield CP, Fuller SA, Stoney AJ & Owen T 2002. Measuring data and link quality in a dynamic multi-set linkage system. Paper presented at Data Linkage Symposium, Sydney, 21–22 March.

List of tables

List of figures