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*®)

	RAC records					Hospital records					
	Exact duplicates		Other duplicates			Exact duplicates		Other duplicates			
Procedure/ pass[(a)(b)]	**1–5**	**[(c)]6–7**	**1–5**	**[(c)]6–7**	**[(a)]Total records**	**1–5**	**[(c)]6–7**	**1–5**	**[(c)]6–7**	**[(b)]Total records**	**[(b)]Links (before *SAS*® stage)**
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)

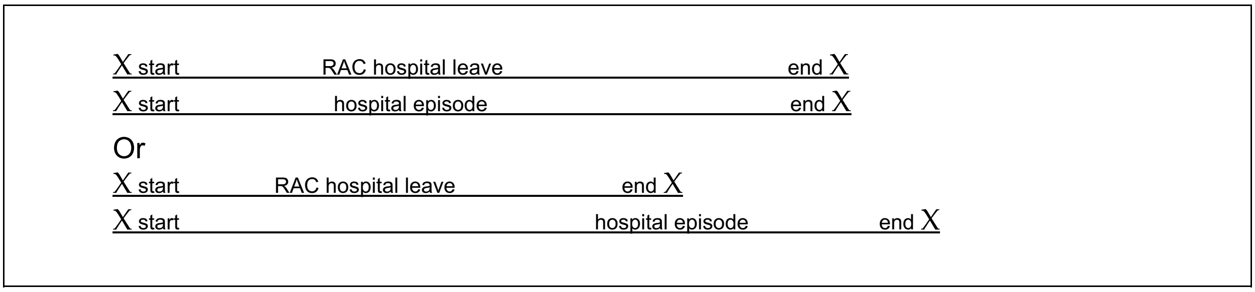
When linking RAC hospital leave to a hospital episode with a non-statistical admission, we expect the start of the RAC hospital leave to match the admission date of the hospital episode. However, if a person is discharged to hospital or dies in hospital, the RAC hospital leave end date may be earlier 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.

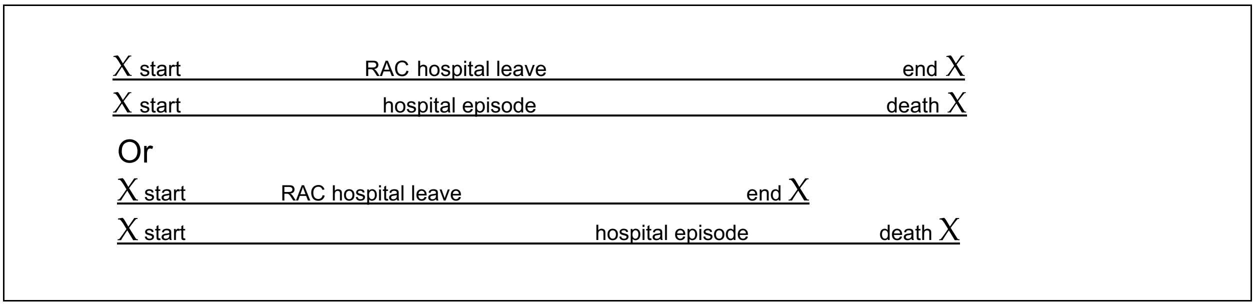
---

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

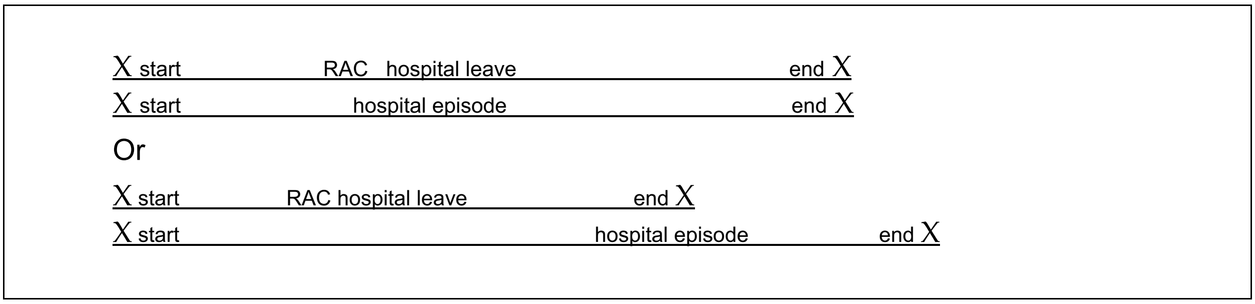
Match 1: Matching RAC hospital leave and non-statistical hospital episodes discharged to usual residence (separation mode = 9), Partition code NST9H, priority = H11



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)

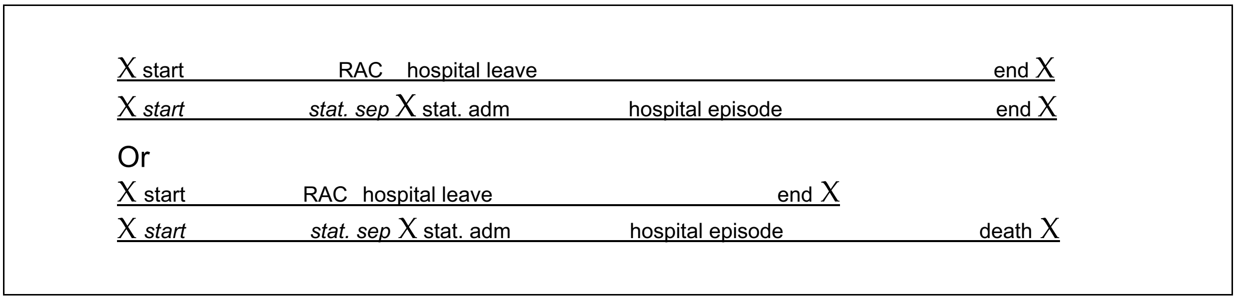
When linking RAC hospital leave to a hospital episode with a statistical admission, the start of the RAC hospital leave may be before the admission date for the hospital episode. In addition, if a person is discharged to hospital or dies in hospital, the RAC hospital leave end date may be earlier than 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).

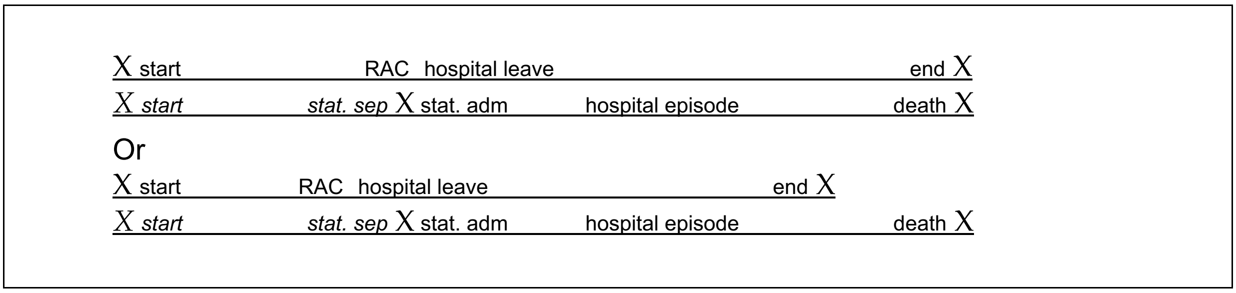
Match 4: Matching RAC hospital leave and statistical hospital episodes discharged to usual residence (separation mode = 9), Partition code ST9H, priority = H21

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

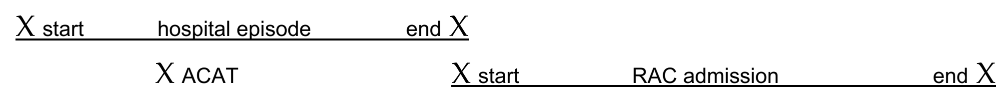
When the assessment is in hospital, the region reported for the RAC client may be that for the hospital (if this was given as the contact address for the ACAT). In this case, the reported RAC client region should be compared with both the hospital region and the usual residence region recorded by the 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.)

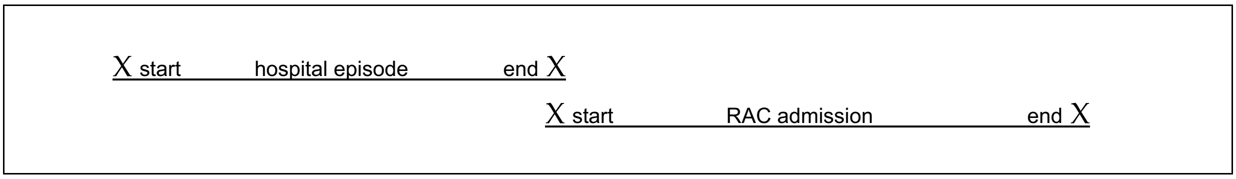
Match 9: Matching RAC admissions and hospital separations excluding discharge to usual residence and deaths (includes separation mode = 2, 4, 6, 7, 0), Partition code 0ADM, priority = R21

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

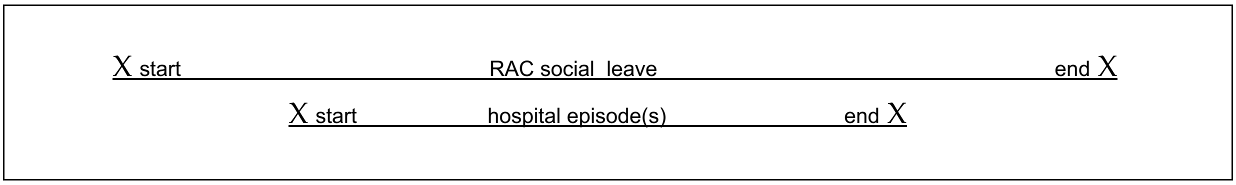


A2.4 Matching RAC social leave

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

Reason for missing match	Poor date of birth: 1 element	Poor date of birth: >1 element	Lag2: <0	Lag2: Other	Lag1 <0: [(a)]To stat. adm.	Lag1 <0: 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

Men aged 65+	State/territory								Total
	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	
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

Women aged 65+	State/territory								Total
	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	
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)]

	Hospital separation mode				
RAC event type	To RAC	To other health care establishment	Left against medical advice/statistical discharge from leave/unknown	To usual residence/other	Total
			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)

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
	Year of birth					
	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 $CSLA_s$, 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 $CSLA_s$, 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 CSLAs, 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 CSLA$_s$, 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