incorrect matching. This, however, led to the exclusion of some valid new year birthdays.

- Same day hospital admission and separation: people admitted and discharged on the same day are unlikely to be discharged to a residential aged care facility unless they are going from a residential aged care facility to a hospital for a day procedure. In this case the person is unlikely to be recorded as an admission by the residential aged care facility. Therefore to avoid spurious matches between aged care admissions and hospital separations into the community, records with the same admission and separation dates were excluded.

For 1999–00, after taking into account the above exclusions, in New South Wales and the Australian Capital Territory 328,220 hospital separations for those aged 65 and over were extracted for analysis (Table A35).

The residential aged care extract

The extract from the residential aged care data included demographic data, details of place and time of assessment prior to entry, type of entry record, demographic data and resident characteristics related to the Resident Classification Scale (that is, dependency). In addition, postcode of usual residence prior to admission was included. A variable was also created to identify multiple entries to residential aged care for the same person during the study period. In order to maximise the capture of movements from hospital to residential aged care, the residential aged care data extract included not only permanent admissions, but also respite care admissions and residents returning from hospital leave.[2] A full list of variables is presented in Appendix 5.

Admissions relating to people born on the 1st of January were excluded from the residential aged care data extract to mirror the exclusion applied to hospital separations as these admissions could never be matched to the hospital extract data.

For 1999–00 in New South Wales and the Australian Capital Territory, the number of residential aged care admissions extracted from the database for analysis was 32,870. These entries were for people aged 65 or more and included permanent admissions, respite admissions and permanent residents returning from hospital leave (Table 1).

4 Testing the feasibility of the linkage strategy

In the discussion below the combination of variables being used to identify individual records is called the linkage key; for example a linkage key could consist of date of birth, sex, postcode of usual residence and exact date of separation / admission. A unique linkage key is one where there is only one instance of the specific combination of variables on the database; for example, using a linkage

[2] Preliminary analyses had revealed that the numbers of patients recorded as discharged from hospital to nursing homes was approximately double the number of permanent residential aged care admissions for the same time period.

key consisting of date of birth, sex, postcode of usual residence and exact date of separation / admission, if there is only one case on the database where a man was born on 14 March 1935, with usual residence postcode 2617, and a hospital separation date / residential aged care admission date of 17 July 1999 then this is a unique linkage key. A non-unique key (termed a 'duplicate') is one where more than one record on the database contains a specific combination of the variables in the linkage key. As the proportion of non-unique linkage keys increases, then the validity of the linkage process necessarily decreases.

Analyses were undertaken to establish the proportion of records with a unique linkage key in the residential aged care database, the hospital morbidity database, and the linked database, on the basis of various combinations of the proposed linkage variables. In the analyses it has been assumed that each record in an extract relates to a different separation (in the case of hospitals) or admission (in the case of residential aged care).

4.1 Unique linkage keys in the residential aged care data

The residential aged care database contains unique numerical identifiers for individuals which allow the analysis of these data at both the record and resident level. Residents have more than one record (i.e. more than one entry) if they entered a residential aged care service more than once during the 12 month study period, for example a permanent admission followed by a return from hospital leave, or a respite admission followed by a permanent admission. Analysis of the proportion of unique linkage keys was undertaken at the record level since the current project is primarily concerned with the movement of people between sectors and not their care histories. In addition, the hospital morbidity data does not support analysis by person.

The results using four different linkage keys are presented in Table 1. In the first half of the table, admission dates were matched only if they were exactly the same; in the second half they were matched if they occurred within a 3 day period.[3]

Using date of birth, sex and exact date of admission for the linkage key, 99.0% of residential aged care records (32,540) had unique keys. Adding postcode of usual residence prior to admission to the linkage key increased this proportion to 99.9%. The table also shows the breakdown of these results within the three categories of residents entering residential care—permanent entries, respite entries and return from hospital leave entries.

As expected, allowing matching of admission dates if they were within 3 days of each other reduced the accuracy of the linkage keys compared with allowing only same date of admission. However, the level of duplicates was still quite small for this

[3] Because the proportion of admissions with unique keys was very high, even without including a geographic indicator, only postcode was used in this analysis. If SLA were to be substituted for postcode, the proportion with unique linkage keys would be between those for the linkage keys with and without postcode.

data set. Using date of birth, sex and admission dates within 3 days as the linkage key, the proportion of unique keys was 97.0%. Again, adding postcode increased the accuracy, with the proportion of unique keys increasing to 99.5%.

These results show that a linkage key based on date of birth, sex and admission date is very good at identifying individual admissions to residential aged care. As expected, adding a geographical element increased the already high accuracy of the linkage key. However, if exact admission date is used (and assuming that it is accurately reported), a geographic dimension is not required for identifying cases within the NSW / ACT residential aged care data set.

Table 1: Duplicates in the residential aged care extract using different linkage keys, by type of resident, NSW/ACT, 1999–00 (admissions, number and per cent)

Linkage key	Permanent	Respite	Leave	All	Permanent	Respite	Leave	All
Exact date		**Number**				**Per cent**		
Linkage key = Date of birth, sex and admission date								
Unique	14,851	14,243	3,634	32,540	99.5	99.5	100.0	99.0
2	68	74	—	330	0.5	0.5	—	1.0
All	**14,919**	**14,317**	**3,634**	**32,870**	**100.0**	**100.0**	**100.0**	**100.0**
Linkage key = Date of birth, sex, admission date and postcode								
Unique	14,917	14,309	3,634	32,832	100.0	99.9	100.0	99.9
2	2	8	—	38	0.0	0.1	—	0.1
All	**14,919**	**14,317**	**3,634**	**32,870**	**100.0**	**100.0**	**100.0**	**100.0**
Within 3 days								
Linkage key = Date of birth, sex and admission date								
Unique	14,746	14,146	3,617	31,881	98.8	98.8	99.5	97.0
2	172	170	17	973	1.2	1.2	0.5	3.0
3	1	1	—	16	0.0	0.0	—	0.1
All	**14,919**	**14,317**	**3,634**	**32,870**	**100.0**	**100.0**	**100.0**	**100.0**
Linkage key = Date of birth, sex, admission date and postcode								
Unique	14,917	14,308	3,622	32,692	100.0	99.9	99.7	99.5
2	2	9	12	176	0.0	0.1	0.3	0.5
3	—	—	—	2	—	—	—	0.0
All	**14,919**	**14,317**	**3,634**	**32,870**	**100.0**	**100.0**	**100.0**	**100.0**

Note: Duplicates among 'All' records include cases with the same linkage key but a different type of resident. Consequently the number of unique linkage keys among 'All' records is smaller than the sum of duplicates in the three resident types.

4.2 Unique linkage keys in the hospital morbidity data

Hospital separation data are episode based, and (unlike the residential aged care database) it was not possible to identify multiple hospital separations for individuals within the study period. The analysis of the proportion of unique records was therefore undertaken at the record (i.e. separation) level.

The proportion of unique linkage keys in the hospital morbidity data based on date of birth, sex and date of separation within 3 days was relatively low at 80.8% (Table 2). This result was quite different from that reported for the residential aged care data, with the difference being the result of the substantial difference in the size of the two data sets (328,220 hospital separations versus 32,870 residential aged care admissions). When the date of separation criterion was tightened so that an exact (i.e. same day) match was required, the proportion of unique linkage keys increased substantially to 94.4% (or 309,910 records).

Adding a geographic linkage variable

Adding postcode of usual residence to the linkage requirements increased substantially the proportion of unique linkage keys. With regard to date of birth, sex and date of separation within 3 days, adding postcode of usual residence increased the proportion of unique keys from 80.8% to 99.1% (Table 2). Using the tighter requirement of an exact day match for date of separation, the proportion of unique records increased from 94.4% to 99.8% with the addition of the geographic variable.

In the last third of Table 2, SLA was substituted for postcode. SLA is in most cases a broader geographic area than postcode and, therefore, when linking across data sets, reduces the likelihood of missed matches due to slight errors in recording postcode (e.g. recording 2614 instead of 2615). On the other hand, it is a less stringent test for matching than postcode since it is more likely to lead to the same linkage key for different separations. Consequently, it is more likely to produce false matches than matching by postcode. Furthermore, in the hospital morbidity data the SLA of a patient's usual residence is derived from either postcode or SLA information, depending on what is available.[4] This may lead to either missed matches or false matches if a patient is assigned to the wrong SLA for their usual residence.

The results when SLA is included reflect these expectations. With regard to date of birth, sex and date of separation within 3 days, the proportion of unique linkage keys was 97.8%—higher than that without a geographical variable, but slightly lower than that for postcode. Using the tighter requirement of an exact day match for date of separation, the proportion of unique records was 99.6%, again an improvement on the 94.4% achieved with no geographical indicator, but only marginally less than that achieved with postcode (99.8%).

[4] In the analysis for this section SLA has been derived from postcode or SLA data provided for the hospital morbidity data, depending on which information was available and where the SLA data may relate to out-of-date SLA boundaries. Postcodes and SLAs from previous versions are assigned to current SLAs with probability equal to the proportion of the postcode's, or old SLA's, population that live within a particular current SLA.

Table 2: Duplicates in the hospital morbidity extract using different linkage keys, NSW/ACT, 1999–00 (separations)

	Date of birth and sex only		Date of birth, sex and postcode		Date of birth, sex and SLA	
Number of duplicates	Exact day	Within 3 days	Exact day	Within 3 days	Exact date	Within 3 days
			Number			
Unique	309,910	265,053	327,688	325,116	327,021	320,954
2	17,602	55,333	532	3,051	1,196	7,108
3	696	7,093		53	3	156
4	12	666	—	—	—	2
5	—	68	—	—	—	—
6	—	7	—	—	—	-
All	**328,220**	**328,220**	**328,220**	**328,220**	**328,220**	**328,220**
			Per cent			
Unique	94.4	80.8	99.8	99.1	99.6	97.8
2	5.4	16.9	0.2	0.9	0.4	2.2
3	0.2	2.2	—	0.0	0.0	0.0
4	0.0	0.2	—	—	—	0.0
5	—	0.0	—	—	—	—
6	—	0.0	—	—	—	—
All	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Note: SLA is derived from postcode and SLA data provided in the hospital morbidity data—see footnote 4.

The effect of age

The proportion of linkage keys that are unique increases with age, although this effect is dependent on the linkage mechanism used. The age-related trend is of potential relevance given that most residential aged care admissions are at advanced ages: in 1999–00, 85% of persons aged 65 and over admitted to residential aged care (permanent and respite) were aged over 75 (Table A1).

For any linkage key, the potential for duplicate keys decreases with the size of the population being examined. Thus the proportion of patients of the same sex with the same date of birth decreases with age because the number of patients in a particular age group decreases with age (see Table 3 and Table A8). Consequently, using date of birth, sex and exact date of separation, the proportion of unique linkage keys increased from between 93% and 95% in the age groups between 65 to 84 years, to 96.1% in the 85 to 89 age group, and to 99.4% in the 95 and over age group (Table 3).

When postcode was added to the matching requirement, the effect essentially disappeared, with the proportion of unique keys being 99.8% or more for all ages. This is because the groups of keys for comparison were so small that duplicate keys were highly unlikely. Similarly when SLA was substituted for postcode, there was no age-related trend, with the proportion of linkage keys that were unique varying between 99.6% and 99.8%.

Table 3: The effect of age on the per cent of unique linkage keys in the hospital morbidity data, using different linkage keys, NSW/ACT, 1999–00 (separations)

Linkage key	65–69	70–74	75–79	80–84	85–89	90–94	95+	All
Exact date				**Number with unique linkage keys**				
Date of birth and sex	57,088	70,224	74,150	56,250	36,107	13,310	2,781	309,910
Date of birth, sex and postcode	60,238	74,806	79,442	59,378	37,511	13,520	2,793	327,688
Date of birth sex and SLA	60,160	74,682	79,230	59,229	37,425	13,508	2,787	327,021
Within 3 days								
Date of birth and sex	48,949	58,541	61,349	48,475	32,385	12,622	2,732	265,053
Date of birth, sex and postcode	59,705	74,176	78,858	58,931	37,235	13,431	2,780	325,116
Date of birth sex and SLA	58,942	73,220	77,768	58,184	36,786	13,300	2,754	320,954
All records	**60,344**	**74,916**	**79,584**	**59,480**	**37,567**	**13,532**	**2,797**	**328,220**
Exact date				**Per cent with unique linkage keys**				
Date of birth and sex only	94.6	93.7	93.2	94.6	96.1	98.4	99.4	94.4
Date of birth, sex and postcode	99.8	99.9	99.8	99.8	99.9	99.9	99.9	99.8
Date of birth, sex and SLA	99.7	99.7	99.6	99.6	99.6	99.8	99.6	99.6
Within 3 days								
Date of birth and sex only	81.1	78.1	77.1	81.5	86.2	93.3	97.7	80.8
Date of birth, sex and postcode	98.9	99.0	99.1	99.1	99.1	99.3	99.4	99.1
Date of birth, sex and SLA	97.7	97.7	97.7	97.8	97.9	98.3	98.5	97.8

Notes

1. SLA is derived from postcode and SLA data provided in the hospital morbidity data—see footnote 4.
2. See Table A2 for complete data.

The above results indicate that a linkage key based only on date of birth, sex and separation date within 3 days is not sufficient for identifying individual separations from hospital for NSW / ACT. Adding a geographical element increases noticeably the accuracy of the linkage key for both linkages based on exact day separation / admission date matches and those allowing up to 3 days difference: duplicates accounted for less than 3% of records for all keys tested which incorporated a geographical dimension. However, increasing the size of the reference area from postcode to SLA increases only very slightly the incidence of duplicates. If exact separation date is used in conjunction with a geographic dimension the resulting linkage key is very effective in distinguishing between separations.

4.3 Unique keys in the linked database

Examining the proportion of unique records in the linked database is a more complex task than that involved in identifying the proportion within the one data set. The review process must take place in the linked data set in two directions. In other words it is necessary to ask what proportion of hospital morbidity records have been linked to more than one residential aged care record, and then what proportion of

residential aged care records have been linked to more than one hospital morbidity record. By combining the two, the number of non-unique links can then be examined. The results of these analyses are presented in Table 4. It is interesting to note that for linkage keys incorporating a geographic indicator, the number of records in the linked database obtained using the 3 day match was only about 7% more than the number using the exact separation / admission date match.

Proportion of hospital records linked to more than one residential aged care record

On the basis of date of birth, sex, and a date of admission to residential aged care *within 3 days* of hospital separation, 17,749 linked records were obtained (Table 4).[5] Of these, there were 469 instances (2.6%) where a hospital record was linked to more than one residential aged care record. When postcode of usual residence was added to the linkage requirements, the number of linked records dropped to 4,301 and the number of instances where hospital records were linked to more than one residential aged care record fell to 22 (0.2%). When SLA group was substituted for postcode,[6] the number of linked records more than doubled, to 10,570, and the number of instances where hospital records were linked to more than one residential aged care record was 56 (0.5%).

Note that including a geographic indicator in the linkage key means that cases may not be matched if a hospital patient was living permanently in residential aged care immediately prior to hospital admission. In this case the hospital may record the postcode of the residential aged care service as the patient's usual residence, while the residential aged care service will have recorded usual residence based on residence prior to admission into the service. It is quite likely for the residential aged care service to have a different postcode from that recorded as the resident's usual residence prior to admission. However, if people move into a residential aged care service near to their place of residence, then the residential aged care service may well be in the same SLA group as the person's previous residence. In 1999–00 for NSW / ACT, 11% (or 3,634) of admissions to residential aged care were for residents returning from hospital leave (Table 1). This difference in recorded usual residence, and the likelihood of error in recording postcode, accounts for the much higher level of matching when using SLA group rather than postcode in the linkage key.

Linking on the basis of date of birth, sex, and the *same* date of separation / admission resulted in 13,459 linked records. Among these, there were 192 instances (1.4%)

---

[5] If a record in one extract links to more than one record in the other extract then the linked data set contains records for each link. For example, one residential aged care admission linking to two hospital separations results in two linked records, and vice versa.

[6] SLA group is based on postcode. For a particular postcode, the corresponding SLA group includes all SLAs which have some residents in that postcode. Because of this, SLA groups may overlap. Using SLA group overcomes the problems of changing SLA boundaries over time and the inaccuracies associated with the derivation of the SLA of usual residence in the hospital morbidity data (see footnote 4).

where a hospital record was linked to more than one residential aged care record. When postcode was added to the linkage requirements, the number of linked records dropped to 4,051, and the number of instances where hospital records were linked to more than one residential aged care record fell to 8 (0.2%). When SLA group was substituted for postcode, as expected the number of linked records was higher than when using postcode, at 9,922; the number of instances where hospital records were linked to more than one residential aged care record was still small at 14 (0.1%).

Proportion of residential aged care records linked to more than one hospital record

Because there were many more hospital separations included in this study than residential aged care admissions—almost 330,000 compared with 33,000—there are much more likely to be multiple hospital records linking to single residential aged care records than the other way round. Consequently, among the 17,749 linked records resulting from linking on the basis of date of birth, sex, and a date of admission to residential aged care *within 3 days* of hospital separation there were 4,550 instances (25.6%) where a single residential aged care record linked to more than one hospital record (Table 4). Including a geographic dimension to the linkage dramatically reduced the number of multiple links. In particular, using SLA group of usual residence, single residential aged care records were linked to more than one hospital record in only 3.1% of cases (that is for 323 out of 10,570 linked records).

As above, allowing only exact separation / admission date matches reduced the incidence of multiple links to single aged care records. On the basis of date of birth, sex, and the *same* date of separation / admission, 9.2% of links related to single residential aged care records linking to more than one hospital record. Adding SLA group to the linkage key reduced this level of duplication to just 0.8% (78 out of 9,922 records). As would be expected, using postcode rather than SLA group resulted in even fewer multiple links.

From the above it can be seen that most duplicate links are caused by one residential aged care record linking to several hospital records. The number of records in the linked data set that result from a record in one data set linking to more than one in the second data set was quite high if the linkage was not restricted according to geographic area. Even when linking was limited to exact separation / admission date matches, 10.6% of linked records related to multiple matches if the linkage key did not incorporate a geographic indicator (Table 4). Requiring a common SLA group of usual residence reduced the level of duplicate keys in the linked data set to 3.6% using the 3 day separation / admission criterion and to under 1% allowing exact only separation / admission date matches.

Table 4: Duplicates in the linked data using different linkage keys, NSW/ACT, 1999–00

	Exact date match		Within 3 days	
Number of records linking	**Number**	**Per cent**	**Number**	**Per cent**
Hospital records linked to one residential aged care record				
Linkage key = Date of birth, sex and date of separation/admission				
1	12,224	90.8	13,199	74.4
More than 1	1,235	9.2	4,550	25.6
Total	*13,459*	*100.0*	*17,749*	*100.0*
Linkage key = Date of birth, sex and date of separation/admission and postcode				
1	4,019	99.2	4,223	98.2
More than 1	32	0.8	78	1.8
Total	*4,051*	*100.0*	*4,301*	*100.0*
Linkage key = Date of birth, sex and date of separation/admission and SLA group				
1	9,844	99.2	10,247	96.9
More than 1	78	0.8	323	3.1
Total	*9,922*	*100.0*	*10,570*	*100.0*
Residential aged care records linked to one hospital record				
Linkage key = Date of birth, sex and date of separation/admission				
1	13,267	98.6	17,280	97.7
More than 1	192	1.4	469	2.6
Total	*13,459*	*100.0*	*17,749*	*100.0*
Linkage key = Date of birth, sex and date of separation/admission and postcode				
1	4,043	99.8	4,279	99.5
More than 1	8	0.2	22	0.2
Total	*4,051*	*100.0*	*4,301*	*100.0*
Linkage key = Date of birth, sex and date of separation/admission and SLA group				
1	9,908	99.9	10,514	99.5
More than 1	14	0.1	56	0.5
Total	*9,922*	*100.0*	*10,570*	*100.0*
All duplicates				
Linkage key = Date of birth, sex and date of separation/admission				
Unique links	12,032	89.4	12,730	72.1
Non-unique links	1,427	10.6	5,019	27.9
Total	**13,459**	**100.0**	**17,749**	**100.0**
Linkage key = Date of birth, sex and date of separation/admission and postcode				
Unique links	4,011	99.0	4,201	98.0
Non-unique links	40	1.0	100	2.0
Total	**4,051**	**100.0**	**4,301**	**100.0**
Linkage key = Date of birth, sex and date of separation/admission and SLA group				
Unique links	9,830	99.1	10,191	96.4
Non-unique links	92	0.9	379	3.6
Total	**9,922**	**100.0**	**10,570**	**100.0**

Note: If a record in one extract links to more than one record in the other extract then the linked database contains records for each link. For example, one residential aged care admission linking to two hospital separations results in two linked records, and vice versa.

The effect of age

As for the morbidity database, in the linked database the proportion of unique records tended to increase with age. On the basis of date of birth, sex and an admission within 3 days of separation, the percentage of records relating to unique linkage keys increased from 64.9% in the 70–74 age group, to 69.0% in the 80–84 age group, and to 93.9% in the 95 and over age group (Table 5). When an exact match on separation / admission date was used, the proportion of records with unique linkage keys increased from around 86% in the age groups under 80 years, to 88.5% in the 80–84 age group, and to 97.0% in the 95 and over age group.

The above age effect virtually disappeared when postcode of usual residence was added to the linkage key because of the very high proportion of unique records in this matched group. However, when SLA group was used rather than postcode a small effect was still observed, with the trend being more noticeable if the 3 day match was used in the linkage key rather than exact date.

4.4 Summary

Overall, for the linkage keys tested, the relatively small number of duplicate linkage keys in both the two data extracts and in the linked data set when a geographic dimension was included, suggests that the proposed linkage strategy can provide a sample of linked client records which could successfully be used to examine the association between resident characteristics, diagnostic and episode variables, and length of stay in the hospital sector. However, restrictions applied when extracting the data sets, uncertainty in a proportion of the linkages due to duplicate linkage keys and inaccuracies in some of the linkage key variables, especially in the geographic variables, mean that the resulting linked data set should not be used for estimating the volume of client flows at this stage.