Linking hospital morbidity and residential aged care data

Examining matching due to chance

Rosemary Karmel

October 2004

Australian Institute of Health and Welfare
Canberra

AIHW cat. no. AGE 40

© Australian Institute of Health and Welfare 2004

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced without prior written permission from the Australian Institute of Health and Welfare. Requests and enquiries concerning reproduction and rights should be directed to the Head, Media and Publishing, Australian Institute of Health and Welfare, GPO Box 570, Canberra ACT 2601.

A complete list of the Institute's publications is available from the Publications Unit, Australian Institute of Health and Welfare, GPO Box 570, Canberra ACT 2601, or via the Institute's website (http://www.aihw.gov.au).

ISBN 1 74024 413 3

Suggested citation

Karmel R 2004. Linking hospital morbidity and residential aged care data: examining matching due to chance. AIHW Cat. No. AGE 40. Canberra: AIHW.

Australian Institute of Health and Welfare

Board Chair
Hon. Peter Collins, AM, QC

Director
Dr Richard Madden

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

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

Phone: (02) 6244 1000
Email: agedcare@aihw.gov.au

Published by Australian Institute of Health and Welfare

Printed by Pirion

Contents

List of tables

List of figures

Acknowledgments

This report was prepared and written by Rosemary Karmel. However, it would not have been possible without the guidance and advice on statistical methods provided by Dr Terry Neeman, consultant statistician to the Australian Institute of Health and Welfare. Discussions with Greg Griffiths also provided useful input into the theoretical aspects of this paper. In addition, Peter Braun of the Ageing and Aged Care Unit assisted with the extraction of data on residential aged care.

A number of Institute staff, including Diane Gibson, Ken Tallis, Ann Peut, Cathy Hales and Jenny Hargreaves, provided comments on drafts of the report. Their contributions are also acknowledged.

Executive summary

Background

The interface between acute hospital care and residential aged care has long been recognised as an important issue in aged care services research. Investigations into the feasibility of linking hospital morbidity and residential aged care data to examine the interface between the two sectors using linkage keys which did not include name, part of name or person identifier were conducted in 2001 and 2002. The study was carried out by the Australian Institute of Health and Welfare under the auspices of the Australian Health Ministers' Advisory Council (AHMAC) Care of Older Australians Working Group, and a final report was presented to this Working Group in June 2002. Findings from the feasibility study suggested that the set of linked client records resulting from the anonymous linkage strategy could provide a valuable source of information on the client characteristics and service use patterns associated with movements between the two sectors.

In the feasibility study, matching was based on date of birth, sex, region of usual residence, and hospital separation and residential aged care entry dates. In the absence of validation against a gold standard linkage, doubts concerning the efficacy of the linkage strategy were raised because of the lack of either name or a common person identifier on the two data sets. This paper employs statistical theory to establish why and under what circumstances the no-name strategy developed in the feasibility study is useful. It investigates the effectiveness of the strategy in a range of linkage situations, and using these results refines the strategy for use in future work.

Introduction

The aim of the linkage strategy is to match hospital separations to entries into residential aged care for people who are admitted into residential aged care following a hospital episode, or who are already permanent aged care residents. Put simply, the strategy links admissions into residential aged care and any hospital stays by permanent aged care residents with a hospital separation by matching on birth date, sex, region of usual residence, and hospital separation/residential aged care admission or re-entry dates. Note that for permanent residents of aged care, the aged care service is considered to be their place of usual residence. When linking records four outcomes are possible: a true link, no link, a false link (false positive) and a missed link (false negative).

Intuitively, among people who usually live in a particular small region we would expect there to be few cases in which a person leaving hospital on a particular day has the same date of birth as someone else entering residential aged care on that day.

In the feasibility study, the regions used in the matching were based on postcodes and grouped Statistical Local Areas (SLAs). Daily rates of hospital separations and RACS admissions and leave returns for people in these regions are sufficiently low that we would expect to get very few matches between the two sets of events just because two people happen to have the same birth date. That is, we would expect only a small number of false matches. The actual statistical probability of such chance matches (that is, of false positives) and the effect that they have on the overall utility of the linked data are the prime focus of this paper. These issues are examined by applying standard statistical theory to the linkage process.

Experience from the feasibility study suggests that when using the AIHW linkage strategy there are likely to be more missed matches than false matches. The issue of missed matches cannot be examined using statistical theory, but can only be investigated by comparing results from the final matching strategy with those from a rigorous (gold standard) name-based matching system. A project which compares links achieved through the AIHW linkage strategy with those obtained using a name-based strategy is currently being negotiated with the Western Australian Department of Health.

Method

In this paper, the reliability of matches is examined by considering the occurrence of a match between a hospital separation and a residential aged care entry purely by chance due to the distribution of birth dates. Two measures are used:

- the probability of there being any matches by chance due to the distribution of birth dates; and
- the rate of such chance matches among achieved matches.

Estimates of these measures are based on analysis carried out for the initial feasibility study, and on broad event rates and distributions for hospital separations and residential aged care admissions and re-entries separately. No additional linkage was carried out for this study.

Data

The calculation of the two measures of reliability is based on data easily available to the author at the time, and so estimates for different aspects of the linkage may have slightly different bases. In addition, a number of simplifying assumptions were made when deriving the reliability measures. In general, the data used when estimating the probability of making any chance matches and the false match rate lead to an upward bias in the estimates of these two measures. Furthermore, the assumptions used when deriving the measures are also likely to lead to over-estimation. Consequently, the results presented in this paper are conservative in that in practice both the probability of any chance matches and the false match rate in the final linked data will be lower than the estimates presented in this report suggest.

Analysis

In 2001–02, for people aged 65 and over there were 1,000,000 hospital episodes involving at least an overnight stay (including those ending in death, change of type of care, or in transfer to another hospital). A proportion of the hospital episodes involved people associated with a residential aged care service (RACS), either as an admission directly following the hospital episode, as a permanent resident returning from a stay in hospital (hospital leave), or as a permanent resident returning to (or going on) social leave from residential aged care. In 2001–02, there were 91,000 permanent and respite admissions into residential aged care (86,000 for people aged 65 and over), 70,000 episodes of hospital leave and 54,000 episodes of social leave. The people to whom these events related lived throughout Australia—across 2,618 postcodes.

In practice, hospital separations on a particular day can be compared with all entry events in residential aged care at a particular time; that is, with residential aged care admissions, returns from hospital leave and periods of social leave. For a particular hospital separation, it is known whether the episode's admission involved a transfer between hospitals or a change in care type within a hospital (that is, if it had a statistical admission). The total length of the period in hospital is only known for separations with a non-statistical admission.

The dates available to identify links vary between the different RACS events, and are listed below. Dates may be matched exactly, or some variation may be allowed.

- All hospital separations can be compared with respite and permanent admissions, matching the hospital separation date with the admission date (termed 'single-date matching').
- As an alternative to using single-date matching, hospital episodes can be compared with RACS hospital leave periods encompassing the hospital episode. Hospital episodes that start with a non-statistical admission can be compared with RACS hospital leave periods matching the hospital episode (termed 'period matching').
- Hospital episodes that start with a statistical admission (including hospital transfers) can be compared with RACS hospital leave periods ending on (or close to) the same date and encompassing the hospital episode (termed 'end-date cover matching').
- All hospital separations can be compared with RACS social leave encompassing the hospital episode (termed 'extended cover matching').

Matching within SLA groups and using exact single-date matching when linking hospital separations to residential aged care admissions and returns from hospital leave, and allowing hospital separations to match to periods of social leave covering the hospital episode—a linkage strategy similar to that used in the feasibility study—the average regional daily probability of any chance matches is estimated to be 0.2%, and the national maximum false match rate is estimated to be 0.4%.

Effect of different linkage strategies

A range of linkage strategies can be used for each of the types of RACS entry. The choices made for the different types affect the overall reliability of the linkage strategy.

Allowing 3-date matching to admissions: When matching to RACS admissions, either exact date or approximate date matching can be used. The latter allows for some inaccuracies in the dates by tolerating a gap between the hospital separation date and the residential aged care admission date. Accepting 3-date matching to admissions rather than only exact-date matching increases both the overall probability of getting at least one chance match and the maximum total false match rate by around two-thirds.

Period matching to hospital leave: Insisting on exact period matching, rather than single-date (end-date) matching, to RACS hospital leave reduces both the daily probability of making any chance matches and the maximum false match rate by up to one-quarter.

Requiring exact period matching when linking hospital episodes to hospital leave reduces the number of potential matches and therefore the rate of chance matches because of the stricter linking requirement. However, in some cases it is too strong a requirement and so will lead to missed matches. In particular, among hospital episodes with a statistical admission no allowance is made for the length of stay of the preceding hospital episode. For these hospital episodes single-date (end-date) or cover matching rather than period matching to hospital leave may be more appropriate.

Cover matching to social leave: Excluding matching to social leave periods, as opposed to using extended cover matching, reduces the daily probability of making at least one chance match and the maximum false match rate by around one-third. When considered on its own, the false match rate within links to RACS social leave may be quite high, depending on the linkage rate achieved for these RACS entries. If after linking, match rates to periods of social leave are found to be very low then including such leave in the candidate matches will only add to the number of false matches without substantially contributing to the number of valid matches.

Effect of birth date distribution and population size

The size of the population within which matching is taking place and the spread of birth dates for that population affect the number of chance matches that will be made during the linkage process.

Distribution of birth dates: The spread of birth dates for a population affects the chance match rate when using the AIHW linkage strategy: the smaller the spread, the greater probability of chance matches. For example, if birth dates are spread roughly normally over 30 years, the probability of any chance matches and the maximum false match rate are both likely to be about one-third lower than those obtained for the same size population with birth dates spread uniformly over 15 years.

Population size: The size of the population within which matching is taking place is a major factor affecting both the probability of making any chance matches and the overall false match rate. The false match rate increases linearly with population size, while the probability of making any chance matches increases exponentially.

Using exact single-date matching when linking to residential aged care admissions and returns from hospital leave (as in the feasibility study) in conjunction with extended cover matching for social leave results in a probability of any chance matches as high as 5.0% for matching regions with 35,000 people (single sex), compared with 0.02% for a population of 2,000. The corresponding false match rates are 4% and 0.2%. When matching within SLA groups using this linkage strategy, due to the predominance of small regions the false match rate in the national linked data will be less than 0.4%. Excluding social leave from the matching process and insisting on exact period matching to hospital leave reduces the maximum false match rate by nearly two-thirds (to 1.5% for a region of 35,000), giving a national false match rate of under 0.2%.

Discussion

The purpose of the anonymous linkage strategy developed by the Australian Institute of Health and Welfare was to find a pragmatic way of linking currently available data on hospital separations and residential aged care entries to provide a linked data set that could be used to investigate the hospital/aged care interface. A key factor determining whether or not the AIHW strategy can be used is the number of candidate matches being considered: if this is too high the probability of chance matches becomes too great and the false match rate increases accordingly. In the current application of the strategy the number of candidate matches is primarily driven by the size of the population within which matches are to be made. By limiting the size of the population group for linkage, the number of chance matches can be kept to such a level that results from many types of analyses would not be affected by their presence.

Because the false match rate increases with population size, using a single linkage strategy across all sizes results in bias in the linked data: larger regions will be over-represented in the data compared with smaller regions. Since changing the linkage strategy can also affect the chance match rate, the linkage strategy could be adjusted according to the size of the population to achieve a near-uniform false match rate for all regions. However, such an approach will introduce other biases by reducing the achieved match rate in the regions with the stricter linkage rules. One possible way of reducing the effect of such biases on analyses is to use a uniform linkage strategy for all regions irrespective of size to obtain a uniform true match rate, and then weighting the resulting links according to the regional false match rates to adjust for any over-representation. In this paper it has not been possible to examine the trade-off between false and missed matches—an issue faced by many linkage strategies.

The relationship between these two may be investigated in a proposed study that compares various linkage strategies with a full name-based linkage.

Overall, estimates based on the theoretical analysis presented in this report indicate that the Institute's linkage strategy results in acceptably low numbers of false matches, and so can be used to derive a data set useful for investigating the hospital/aged care interface. Furthermore, although the linkage strategy has been developed for linking hospital separations to residential aged care entries, it could be used in any situation where similar data are available (that is, date of birth, sex, region and event dates).

1 Background

The interface between acute hospital care and residential aged care has long been recognised as an important issue in aged care services research. Investigations into the feasibility of linking hospital morbidity and residential aged care data to examine the interface between the two sectors using linkage keys which did not include name, part of name or person identifier were conducted by the Australian Institute of Health and Welfare (AIHW) in 2001 and 2002. The study was undertaken under the auspices of the Australian Health Ministers' Advisory Council (AHMAC) Care of Older Australians Working Group, and a final report was presented to this Working Group in June 2002 (AIHW 2003a).

Findings from that study suggested that the set of linked client records resulting from an anonymous linkage strategy could indeed provide a valuable source of information on the client characteristics and service use patterns associated with movements between the two sectors. Analyses such as the relationship between hospital length of stay, or primary diagnosis, and subsequent dependency levels or length of stay in residential care, could therefore be successfully undertaken. However, as the matching rate was less than expected, the linked data set was not considered suitable for the analysis of aggregate flows between sectors, or for comparisons of rates of flow over time. In addition, in the absence of validation against a gold standard linkage, doubts concerning the efficacy of the linkage strategy were raised because of the lack of either name or a common person identifier on the two data sets.

When deriving the final linked database, the linkage strategy for linking a hospital separation to a residential aged care service (RACS) entry will be a refinement of that used in the feasibility study. In that study, matching was based on date of birth, sex, region of usual residence, and hospital separation date matching to a residential aged care entry date. The same linkage strategy was used for all types of residential aged care entries. As different information on event dates is available depending on the type of residential aged care entry, adjusting the linkage strategy for the different types of entries may lead to more accurate matching. The following analysis employs standard statistical theory to investigate why and under what circumstances the no-name strategy developed in the feasibility study is useful. This paper investigates the effectiveness of the no-name strategy in a range of linkage situations, and discusses how the results can be used to refine the strategy for use in future work. The issue is examined using two measures of the reliability of matches between hospital separations and candidate residential aged care entries.

2 Introduction

The aim of the linkage strategy is to match hospital separations by people who then go into residential aged care, or who are already permanent aged care residents, with the relevant aged care entries.

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

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

b going to live temporarily with family and friends before returning to a residential aged care service;

c going into a residential aged care service;

d returning to the residential aged care service where they usually live;

e transferring to another hospital;

f transferring to residential health care services;

g changing episode type within the hospital; or

h dying.

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

A being admitted into residential aged care for permanent care;

B being admitted into residential aged care for respite care;

C transferring between residential aged care services (for either permanent or respite care);

D returning to residential aged care following an episode in hospital (termed 'hospital leave'); or

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

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

In terms of the events described above, the linkage strategy tries to match hospital separations of type *c* to the relevant RACS entries of types *A* and *B*; separations of type *d* to entries of type *D*; separations of type *b* to the relevant entries of type *E,* and separations of type *h* to deaths of aged care residents while in hospital (type *F*). Admissions into residential aged care and any hospital stays by residential aged care residents recorded as hospital leave have an exact date suitable for linkage—the admission date or the return from leave date—to a hospital separation. In addition, periods of RACS hospital leave should cover the hospital length of stay. In cases where hospital episodes are in the middle of social leave from residential aged care, the situation is more complicated as there is not an exact date suitable for linkage, but comparisons of the period of hospital episodes with the period of social leave can still be useful.

2.1 False matches

When linking records four outcomes are possible: a true link, no link, a false link (false positive) and a missed link (false negative). Intuitively, among people who usually live in a particular small region we would expect there to be few cases in which a person leaving hospital on a particular day has the same date of birth as someone else entering residential aged care on that day. In the linkage feasibility study, the regions used in the matching were based on postcodes and grouped Statistical Local Areas (SLAs). Daily rates of hospital separations and RACS admissions and leave returns for people in a region of 70,000 people aged 65+—around the size of the largest SLA group in 2001 (see Table A12)[1]—are sufficiently low that we would expect to get very few matches between the two sets of events just because two people happen to have the same birth date (Table A1). The actual probability of such false, or chance, matches, and the effect that they have on the overall utility of the linked database, are the prime focus of this paper.

2.2 Missed matches

Experience from the feasibility study suggests that using the AIHW linkage strategy there are likely to be more missed matches than false matches. The issue of missed matches cannot be addressed using a theoretical approach, as in this paper, but can only be investigated by comparing results from the final matching strategy with those from a rigorous (gold standard) name-based matching system. A project which compares links achieved through the AIHW linkage strategy with those obtained using a name-based strategy is currently being negotiated with the Western Australian Department of Health. Such a comparison would identify both false matches (false positives)and missed matches (false negatives).

2.3 Data quality

The analysis in this paper assumes that all information concerning hospital episodes and residential aged care entries is correct. However, false matches and missed matches may arise because of differences in the data recorded for the various episodes; that is, for date of birth, sex, geographic region and event date.

Of particular concern are data on region of usual residence which may differ for a number of reasons. It may particularly be a problem for people who are not staying

[1] In 2002, the largest SLA had 29,100 people aged 65+ (both sexes together) and there were only 37 SLAs with more than 10,000 people aged 65+. In addition, 582 SLAs (42%) out of 1,384 had fewer than 500 people aged 65+; the median size of SLAs was 676 people aged 65+.

in their usual residence just prior to hospitalisation; for example, when people usually living in residential aged care are staying with relatives (on social leave). In this situation, for the hospital episode the address of the relatives may be recorded as the address of usual residence rather than that of patients themselves. If these addresses are in different regions, then match errors may occur.

Another source of differences in recorded region relates to the way the region of a residential aged care service is recorded on the Aged and Community Care Management Information System (ACCMIS) database—the source of data on residential aged care. The region of usual residence for a permanent residential aged care resident is the region in which their residential aged care service is located. However, in some cases a residential aged care service may consist of several outlets. As ACCMIS can record only one region for a residential aged care service, if the outlets are located in different regions the location of usual residence for a permanent aged care resident could be recorded differently in the hospital data (which would probably record the outlet's region) and the residential aged care data (which may record the region related to the administrative centre of the service).

Differences in recorded region may also occur for people being admitted into residential aged care from hospital. On ACCMIS the information on usual residence refers to where the client could be contacted following an Aged Care Assessment Team (ACAT) assessment. Under a range of circumstances this could be the home of relatives or friends, or a hospital, rather than the client's usual residence (as would be recorded on the hospital database).

The prevalence of erroneous and conflicting data for related hospital and residential aged care episodes is unknown, but it would be a cause of both missed and false matches. The number of false matches in the discussion below relates only to those matches occurring due to coincident birth dates and does not include those caused by differences in data.

The extent of both missed and false matches in a linked data set affects the uses to which the data can be put. While the number of missed matches and false matches may be important in certain contexts (for example, when looking at patient flows), in many analyses—such as those investigating patterns of use—biases in the linked data are more critical. As part of the comparative study referred to previously, analysis of the two linked data sets will be undertaken to identify any differences in the data resulting from the two linkage strategies. Such differences could arise from the linkage strategies themselves, or be the result of poor linkage data, especially with respect to date of birth and recorded region of usual residence.

3 Method

In this paper the reliability of matches is examined by considering the occurrence of a match between a hospital separation and a residential aged care entry purely by chance due to the distribution of birth dates. Two measures are used:

- the probability of there being any matches by chance due to the distribution of birth dates; and
- the rate of such chance matches among achieved matches.

Estimates for these measures are based on analysis carried out for the initial feasibility study (AIHW 2003a), and on broad event rates and distributions for hospital separations and residential aged care admissions and re-entries separately. No additional unit-level linkage was undertaken.

3.1 Probability of any chance matches

As the probability of any chance matches decreases, our confidence in any matches that have been identified increases. Thus, the probability of any chance matches (or false positives) between hospital separations and entries into residential aged care can be used as an indicator of the confidence we can have in resulting matches.

The probability of a group of hospital separations (on a particular day) having different dates of birth from a particular number of possibly linking RACS admissions/returns is readily derived using two assumptions: that the match candidate residential aged care entries all have different birth dates, and that these birth dates are distributed uniformly across 15 years. Under these assumptions

P = Probability of m hospital separations having different birth dates from all k RACS admissions
= [Probability of a single hospital separation having a
different birth date from all RACS admissions]$^{(\text{no. hospital separations})}$
= [(total number of allowed birth dates)/(total number of possible birth dates)]$^{(\text{no. hospital separations})}$
= $[(5{,}475 - k)/5{,}475]^m$ where m is the number of hospital separations on a day, and
k is the number of associated candidate residential aged care entries

The probability Q of there being any, that is 1 or more, matches by chance is then Q = (1-P).

Using this formula, it is estimated that the probability of there being no chance matches is over 99% when 20 or fewer hospital separations are compared with 2 or fewer residential aged care admissions or returns from leave, as might reasonably happen for a region of 15,000 men or women aged 65 and over (see Table A3).

3.2 False match rate

Whether or not chance matches make up a sizeable proportion of all matches, and therefore affect the utility of the linked data, can be seen by comparing the number of chance matches with the total number of achieved matches; that is, by estimating a false match rate per hospital separation date (termed the 'daily false match rate').

The daily expected number of chance matches C due to coincident birth dates when comparing m hospital separations (for a particular day) with k candidate residential aged care admissions/returns is simply:

C = Estimated number of chance matches

[Probability of a single hospital separation having the same birth date as a RACS admission/return] x [number of comparisons]

= 1/(Total number of possible birth dates)

x (Number of possibly matching residential aged care entries)
x (Number of hospital separations)

= $k \times m/5{,}475$

Since final matching has not yet been undertaken, the number of matches achieved between separations and residential aged care entries is not known. In the absence of such information, the false match rate F of a particular linkage strategy due to chance matches can be estimated assuming a certain match rate α between hospital separations and residential aged care entries; that is,

F = daily false match rate = C/(Expected number of matches to m hospital separations)

= $C/(\alpha \times m)$

Rather than assume a single final match rate, a range for the false match rate is estimated from a corresponding range of match rates. A range for the false match rate using a particular linkage strategy can then be estimated using an ideal high match rate α_H (to give a lower limit) and the match rate observed in the feasibility study α_L (to give an upper limit):

$$C/(\alpha_H \times m) \leq F \leq C/(\alpha_L \times m)$$

or equivalently:

$$k/(5{,}475 \times \alpha_H) \leq F \leq k/(5{,}475 \times \alpha_L)$$

Under the linkage strategy used in the feasibility study, for NSW/ACT around 2.5% of hospital separations matched to a residential aged care admission when using exact date matching, and 0.5% matched to a return to residential aged care following hospital leave (AIHW 2003a:Tables 2, 5, 6, A3).[2] On the other hand, it estimated that between 45% and 50% of residential aged care admissions are from hospitals (Gibson 2002:134), and theoretically all periods of RACS hospital leave should have a corresponding hospital episode. For NSW/ACT in 1999–00, 50% of residential aged care admissions and 100% of hospital leave returns equated to 4.5% and 1.1% of hospital separations, respectively (AIHW 2003a:Tables 1 and 2).[3] Therefore, for the linkage strategy used in the feasibility study $\alpha_H = 0.056$ and $\alpha_L = 0.03$, giving a range for the false match rate of

$$k/(5{,}475 \times 0.056) \leq F \leq k/(5{,}475 \times 0.03)$$

Continuing the example of 20 hospital separations and 2 residential aged care admissions for a population of 35,000 people, the expected number of chance matches between these two groups is much less than 1 (0.007). This means that if 1,000 such comparisons were made—for example, over 1,000 days of hospital separations—we would expect to see 7 matches purely due to the distribution of birth dates. That these chance matches would have only a small effect on many analyses can be seen by looking at the false match rate: when comparing 20 hospital separations with 2 residential aged care entries, and using the linkage strategy trialled in the feasibility study, the false match rate is estimated to be between 0.6% and 1.2% (that is, 100*0.007/(0.056*20) ≤ F ≤ 100*0.007/(0.030*20)). For regions with more than 35,000 people the effect would be greater; for smaller regions it would be less. Given that few SLA groups have more than 15,000 people aged 65 and over (single sex) (Table A12), these initial results suggest that chance matches would have a minor effect on the utility of the linked data resulting from the AIHW linkage strategy as used in the feasibility study.

---

2 These figures do not include hospital separations due to death, as these were excluded from the feasibility study. In addition, the same linkage strategy was used in the feasibility study to match hospital separations with both admissions and returns to residential aged care.

3 The number of periods of hospital leave in the study were low compared with the national experience. If it is assumed that 12.6% of non-same-day hospital separations for people aged 65 and over are due to transfers and statistical discharges (as found for NSW/ACT for 1999–00), and that 95.6% of periods of leave from residential aged care are for people aged 65 and over (as for permanent residents 2001–02), then 50% of admissions (43,000) equate to 4.5% of hospital separations for people aged 65 and over excluding statistical discharges and transfers (Table A11), and 100% of hospital leave returns (66,900) equate to 6.2% of such hospital separations (Tables A1 and A11). Using these numbers, the lower limit for the false match rate becomes $C/0.107 \times m$; that is, it roughly halves.

3.3 Assumptions and data

Assumptions

Several key assumptions are made when deriving the formulae for the two reliability measures: that birth dates are spread across 15 years, that birth dates have a uniform distribution across the years, and that match candidate residential aged care entries for a particular hospital separation all have different birth dates. Using just a 15-year spread for birth dates assumes a greater concentration of birth dates than actually happens as dates of birth for those aged over 65 spread across at least 35 years. Consequently, this assumption will result in over-estimation of the probability of there being any matching birth dates purely by chance, and also of the number of chance matches. On the other hand, assuming a uniform distribution of ages leads to either under- or over-estimation of the measures, depending on the part of the age distribution to which a patient/resident belongs. Using just a 15-year range of birth dates therefore ameliorates the effect of the assumption of a uniform birth date distribution. The effects of these distributional assumptions are examined numerically in Section 4.2.

The last of the assumptions concerning birth dates (that is, that match candidate residential aged care entries for a particular hospital separation all have different birth dates) is used when estimating the probability of no chance matches. This assumption was made to simplify the derivation of the probability and will have only a marginal effect on the results. That the effect will be very limited can be seen from the very small probability that people of the same sex admitted into, or returning to, residential aged care over a few days with the same region of usual residence have identical birth dates. Using the conservative assumption that birth dates are distributed uniformly across 15 years (that is, across 5,475 dates), the probability of k residential aged care entries all having different birth dates is $(5474!/\{(5{,}475 - k)! \times 5{,}475^{(k-1)}\})$. This probability is at least 98% when 15 or fewer entries into residential aged care are being compared with hospital separations; it is at least 99.5% when comparing with eight or fewer RACS entries, and at least 99.9% when comparing with fewer than five entries. The analysis in Chapter 4 indicates that in almost all cases there will be fewer than eight residential aged care admissions and returns from leave available for matching to hospital episodes ending on a particular day. In addition, birth dates have a much greater spread than 15 years. Consequently, the assumption that match candidate residential aged care entries for a particular hospital separation all have different birth dates will have little effect on estimates of the probability of there being no chance matches. As this assumption was not used when deriving the expected number of false matches, it will have no effect on the estimated false match rate.

Data

In this paper the calculation of the two measures of reliability is based on data easily available to the author at the time. The following information should be noted with regard to the data used in the analysis.

Hospital data

- Data on hospital episodes are for 2000–01 and are from the National Hospital Morbidity Database (NHMD) maintained by AIHW.
- Data on hospital episodes include only those episodes for people aged 65 and over lasting longer than a day (that is, that involve at least an overnight stay).
- In general, hospital separation rates exclude same-day separations as same-day hospital episodes are unlikely to lead to admission into residential aged care, and same-day hospital episodes by permanent aged care residents are unlikely to be recorded in the aged care data.
- The NHMD records episodes of hospital care: if the type of care changes, then a hospital episode is closed (a statistical discharge) and a new episode is started (a statistical admission). Also, new episodes are recorded when people transfer between hospitals. Because such separations are not associated with people leaving hospital, in the AIHW linkage strategy all hospital separations ending with a statistical discharge or a hospital transfer are precluded from linking. All other hospital separations—termed 'non-statistical'—are available for linkage, irrespective of their type of hospital admission. In this report, the term 'statistical separation' includes all separations that were either statistical discharges or transfers to another hospital.
- 'Statistical admission' is used to refer to all hospital episodes that began as the result of a change in hospital care type or as a transfer from another hospital.[4] Hospital episodes available for linkage can result from either a statistical or non-statistical admission.
- In general, hospital separation rates include deaths and statistical separations. While deaths could relate to people living permanently in residential aged care, as stated above statistical separations should not match to people entering or returning to residential aged care. The inclusion of statistical discharges and transfers in the separation rates will therefore lead to over-estimation of the daily rate of separations suitable for linking, and therefore to an over-estimation of both the probability of making at least 1 chance match and the false match rate. For NSW/ACT 1999–00, among episodes involving at least 1 night 8.5% of separations were for hospital transfers, 4.1% were for statistical discharges and 4.9% were for people who died in hospital.
- Distribution of hospital episodes by length of stay, mode of admission and mode of separation are based on ACT/NSW 1999–00 data. Slightly different results

[4] Mode of admission is available on the NHMD from 1999–00.

would be obtained if national data, or data from a different region, were used. Any differences are unlikely to affect the conclusions of the analysis.

Residential aged care data

- Data on residential aged care are for 2001–02 and are from the Aged and Community Care Management Information System (ACCMIS) maintained by the Australian Department of Health and Ageing.
- Residential aged care admission rates are for people aged 65 and over.
- Data on residential aged care leave rates and distribution of length of leave are for all permanent residents. In 2001–02, 96% of permanent residents were aged 65 and over (AIHW 2003b:29). Including younger residents in the leave rates will lead to over-estimation of the daily rate of separations suitable for linking, and therefore to over-estimation of both the probability of making at least 1 chance match and the false match rate.

General

- Hospital separation and residential aged care entry rates relate to all of Australia.
- Match rates between hospital separations and residential aged care entries are from the feasibility study (AIHW 2003a), and so are based on data for NSW/ACT 1999–00.
- When using distributions across length of stay, a maximum stay of 30 days is assumed. In 2001–02, for permanent aged care residents 7.8% and 1.3% of hospital and social leave periods, respectively, were at least 30 days long; in NSW/ACT in 1999–00, 13.2% and 2.2% of statistical and non-statistical admission hospital episodes, respectively, were for at least 30 days. This collapsing of the data will lead to an overstatement of probability of making any chance matches as grouping across length of stay assumes matching is possible between events that would not be considered candidate matches using exact period length.
- Due to the complexities involved, variations in rates across regions and within population groups have not been incorporated into the estimates. Consequently, the same hospital separation and residential aged care admission and re-entry rates are assumed to apply nationally.

From the above it can be seen that, in general, the data used when estimating the probability of making at least 1 chance match and the false match rate lead to an overstatement of these two measures. Furthermore, the assumptions used when deriving the measures are also likely to lead to over-estimation. Consequently, the results presented in this paper will be conservative in that the probability of any chance matches and the false, or chance, match rate in the final linked data will be lower than estimates presented in this report would suggest.

4 Analysis

In practice, people can enter residential aged care services in a number of ways: as admissions, returning from hospital leave or returning from social leave. In 2001–02, for people aged 65 and over there were 1,000,000 hospital episodes involving at least an overnight stay (including those ending in death, change of type of care, or in transfer to another hospital; Table A11). Also in that year, there were 91,000 permanent and respite admissions into residential aged care (86,000 for people aged 65 and over; AIHW 2003b), 70,000 episodes of hospital leave and 54,000 episodes of social leave. A proportion of the hospital episodes involved people associated with residential aged care, either as an admission directly following the hospital episode, or as a permanent resident returning from a stay in hospital (hospital leave), or as a permanent resident returning to (or going on) social leave from residential aged care. The people to whom these events related lived throughout Australia—across 2,618 postcodes. The effect of chance matches on the resulting linked data for these three entry types is discussed separately below. The overall effect of various linkage strategies is then examined.

4.1 Matching to different types of residential aged care entries

Linking to RACS admissions

Accounting for around two-fifths of entries into residential aged care, admissions are the greatest supply of possible links between hospital separations and entries into residential aged care. As a consequence, admissions are likely to be the greatest source of chance matching. When matching to RACS admissions, the region of usual residence used is that of the person's region of usual residence prior to the hospital episode. Single-date matching is used; that is, a hospital separation date is compared with a RACS admission date. In this context, either exact date or approximate date matching to the hospital separation date can be used.

Because the number of comparisons being made increases with the size of the regions within which matching is taking place and with the number of admission dates being considered as possible matches, the probability of any chance matches and the expected number of such matches both increase with population size and the number of dates being considered for matching. In the feasibility study, both exact separation/admission date matching and 3-date matching to admissions were investigated, where 3-date matching allows a hospital episode to match to residential aged care admissions on 3 dates (including the same date, exact date + 1 day, and exact date + 2 days). Such a gap in dates allows for inaccuracies in recording movement dates, and for patients who return home to prepare for the move into residential care, for example, before being admitted. In the feasibility study, allowing for 1 or 2 days delay between leaving hospital and entering residential aged care increased the number of achieved matches by 7% (AIHW 2003a:Tables 6, A3), suggesting that such delays, or differences in reporting dates, may be occurring.

Among populations of up to 10,000 (single sex), the probability of making any chance matches for hospital separations on a particular day is quite small: less than 1% allowing 3-admission-date matches and assuming that the expected number of hospital separations and residential aged care admissions are actually observed on a particular day (Table 1, and examples in Table A4). The effect of chance matches on analysis is also likely to be small: for populations of less than 10,000—with fewer than 12 hospital separations and 1 residential aged care admission expected daily—the daily expected number of chance matches is 0.002, and the false match rate for admissions is estimated to be less than 2.0% using 3-date matching, compared with 0.7% using exact date matching.[5]

For large populations, allowing 3-date matching could result in an unacceptable number of chance matches. However, using exact date matching the level of chance matches is still at an acceptable level for many types of analyses for populations as large as 35,000 (single sex). For example, for a population of 35,000, where on average we expect 42.4 hospital separations and 3.3 residential aged care admissions a day, there is less than a 3% probability of at least 1 match between hospital separations for a particular day and residential aged care admissions for that day due just to the random distribution of birth dates; the associated false match rate is also under 3% (Table 1). The corresponding figure when using 3-date matching is 7% for both measures.

Linking to returns to RACS following hospital leave

People living permanently in residential aged care may need to go to hospital for treatment. After admissions, hospital leave episodes are the second largest reason for entry in residential aged care, accounting for one-third of entries into RACS annually. The number of chance matches is therefore expected to be fewer among returns from hospital leave than among admissions into residential aged care. On the other hand, given the nature of the leave, the proportion of such entries matching to a hospital separation is expected to be high.

As for RACS admissions, a single-date strategy could be used when linking to RACS hospital leave. However, the linkage can be made more precise for these entries by incorporating the length of leave into the linkage rules. The region of usual residence used when undertaking linkage for people returning to their residential care after a

---

[5] The range for the false match rate using exact day matching for admissions is given by $[(km/5{,}475)/(0.045*m) \leq F \leq (km/5{,}475)/(0.025*m)]$ (see page 6). Assuming a 7%increase in the match rate when using 3-date matching changes the range to $[(3km/5{,}475)/(0.045*m*1.07) \leq F \leq (3km/5{,}475)/(0.025*m*1.07)]$.

period of hospital leave is the region in which the residential aged care service is located, rather than their usual residence prior to admission to that service.

The procedures used to record hospital episodes in the NHMD affect the way in which the length of the hospital episode can be included in the linkage process. Due to the way that hospital episodes are recorded in the data set (see Section 3.2), not all recorded hospital separations result in the patient leaving the hospital system: some separations end in transfer to another hospital or in change of care type (see types *e* and *g* earlier). These statistical separations mean that some continuous stays in hospital are associated with several hospital separations. As a consequence, the length of hospital leave from residential care may be longer than the hospital separation associated with return to residential care.

Table 1: Measures of match reliability, by RACS entry type (per cent)

	Population 10,000 (single sex)		Population 35,000 (single sex)	
Linkage strategy	**Probability of any chance matches (Q)**	**Range of false match rate (F)**	**Probability of any chance matches (Q)**	**Range of false match rate (F)**
RACS entry type				
Admissions				
Exact date matching	0.21	0.39–0.69	2.51	1.35–2.43
3-date matching	0.63	1.08–1.95	7.36	3.78–6.80
Hospital leave				
Period matching–for hospital episodes of 1 day, allowing for 1 day in emergency	0.007	0.30–0.66	0.09	1.03–2.26
End-date cover matching–for hospital episodes of 1 day	0.03	1.28–2.81	0.39	4.47–9.83
End-date cover matching–for hospital episodes of 2 days	0.02	1.10–2.41	0.20	3.85–8.47
Exact date matching[a]	0.17	1.28–2.81	2.06	4.47–9.83
Social leave				
Cover matching–for hospital episodes of 1 day	0.48	?[a]	5.75	?[a]
Cover matching–for hospital episodes of 2 days	0.04	?[a]	0.50	?[a]

(a) False match rate cannot be estimated as matching to social leave was not considered in the feasibility study.

Note: Calculations use average person separation/admission/leave rates (see Tables A1, A5, A7 and A8). Hospital separation rates exclude same-day episodes.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD; see also Tables A9 and A14.

Linking hospital separations with a non-statistical admission

In the case of a hospital separation with a non-statistical hospital admission (that is, not a statistical admission and not a transfer from another hospital), the total length of the hospital stay is known. Consequently, the matching strategy can include both the start and end dates of the hospital stay; that is, 'period matching' could be employed. In addition, the period matching could be adjusted to allow for 1 additional day (at the beginning of the period) to account for time spent in the emergency department away from residential aged care but not yet admitted into hospital. Similarly, when permanent residential aged care residents die in hospital some leeway in the period end date may be appropriate.

Because of its greater specificity, a linkage strategy using period matching would be expected to result in fewer chance matches than matching using only hospital separation date and RACS hospital leave return date. In addition, the expected daily number of hospital separations decreases with the length of stay in hospital (Table A5), so that chance matches are less likely as the length of hospital stay increases. Even for populations as large as 35,000 (single sex), the expected number of leave returns equal to, or 1 day longer than, a particular hospital stay is small at less than 1 for hospital separations of all lengths ending on a given day (Table A5). Consequently, the probability of making any matches by chance is small, and the expected number of chance matches is similarly low. Examples showing the reliability of period matching to RACS hospital leave are given below. Note that the estimates of the two reliability measures provide upper limits because they assume that all hospital separations have a non-statistical admission and therefore overstate the number of comparisons being made when trying to link hospital and residential aged care events. For NSW/ACT in 1999–00, 89% of non-same-day separations began with a non-statistical admission.

As an extreme example, consider linking overnight hospital stays (the most common length of stay) for an area with a population of 35,000 (a very large SLA group) to RACS hospital leave ending on the same day and beginning on or the day before the hospital admission date to allow for time in an emergency department. Daily averages for this scenario are up to 7.9 hospital separations for overnight stays with non-statistical admissions and 0.6 RACS leave returns following 1 or 2 days in hospital (Tables A5 and A8). Matching between these numbers of events results in a very low probability of any chance matches—just 0.1% (Table 1). However, because of the smaller number of achieved matches expected through linking to hospital leave returns compared with RACS admissions, the false match rate for hospital leave episodes is estimated to be between 1.0% and 2.3%—similar to that for same-date matching to admissions.[6] For longer hospital periods, matches will have greater reliability because of the smaller numbers involved in the matching.

Delays which might occur between a death in hospital and the end of the associated RACS leave period being recorded by the residential aged care service could be accounted for by using 3-date matching for the leave return date. As such an approach leads to more RACS candidate links for a hospital separation, it increases both the probability of making any false matches and the false match rate. However,

---

[6] The range for the false match rate is $[(km/5{,}475)/(0.011*m) \leq F \leq (km/5{,}475)/(0.005*m)]$ (see page 6). This range is likely to be biased upward due to the exclusion of cases where people have died in hospital and also to the relatively low RACS hospital leave rate in the source data used to derive the expected number of achieved matches.

the effect would not be as great as when using 3-date matching for admissions because the use of the start date in the linkage means that, rather than increasing three-fold, the number of candidate matches would increase by the number of RACS hospital leave episodes of sufficient length to cover the hospital episode as recorded by the hospital and the additional 'notification' days. In addition, the effect could be further reduced by allowing for additional days only when the hospital separation is identified as a death, thereby reducing the number of hospital separations being compared with candidate RACS hospital leave returns—in NSW/ACT in 1999–00, 6% of hospital episodes ended in death, excluding statistical separations and same-day episodes.

In the example above, an average of 1.5 hospital leave returns would cover an overnight stay in hospital plus an additional 'emergency' day at the beginning and 2 additional 'notification' days at the end (up from 0.6). The probability of any false matches among resulting links would be 0.2% and the maximum false match rate would rise to 5.2% (again assuming a 7% increase in identified links when allowing 3-date matching). If only 1 of the average 7.9 hospital separations ended in death, then the probability of making a false match falls to 0.03%. The resulting false match rate cannot be estimated until the final linkage is undertaken because the linkage rate for hospital deaths is unknown as these episodes were excluded from the feasibility study.

Linking hospital separations with a statistical admission

If a hospital separation begins with a statistical admission, the total length of the stay in hospital is not known. In this situation, all episodes of RACS hospital leave that end on the same date as the hospital separation and lasting at least as long as the hospital episode could be considered candidates for matching (termed 'end-date cover matching'). While such a link criterion is likely to lead to more chance matches than exact period matching, there is still an element of start and end-date matching in this approach, so that the probability of any chance matches would be expected to be lower than when using single-date (end-date) matching.

Table A5 shows that comparing length of hospital stay to RACS hospital leave covering the hospital episode results in more candidate matches than exact period matching. For a population of 35,000, using cover matching between (on average) up to 7.9 overnight hospital episodes and 2.7 RACS hospital leave returns covering those episodes, the probability of making any matches purely by chance is less than 0.5% and the false match rate would be between 4.5% and 9.8% (Table 1). For a population of 10,000, on average we expect 2.3 hospital separations a day that involve an overnight stay and 0.8 candidate hospital leave returns. For this comparison there is less than a 0.03% probability of a chance match, and the false match rate is less than 3%. Note again that these estimates provide upper limits for the two reliability measures because they assume that all hospital separations have a statistical admission and therefore overstate the number of comparisons being made.

If the type of admission for hospital episodes is ignored, and exact end-date matching is used between hospital separations and returns from RACS hospital leave instead of period or end-date cover matching, acceptable levels of chance matches are still expected for populations under 10,000 (single sex). For a population of 10,000 there is less than a 0.5% probability of a chance match between the 12.1 hospital separations and 0.8 returns from RACS hospital leave expected daily, and the corresponding false match rate would be under 3%. For larger populations, the chance matches would start to affect the utility of the linked data. For example, for a population of 35,000 the probability of making any matches purely by chance using end-date matching (between, on average, 42.4 hospital separations and 2.7 returns from RACS hospital leave) is 2%; however, despite the low probability of chance matching, the false match rate is estimated to be between 4.5% and 9.8% because of the relatively low number of expected matches between hospital separations and RACS hospital leave.

Linking to RACS social leave

People living permanently in residential aged care may go on social leave, and while on social leave they may have a period in hospital. There is no funding incentive for that episode to be recorded as hospital leave (sandwiched between two episodes of social leave) unless the resident has used up all of their claimable social leave entitlements of 52 days for the financial year.[7] It is therefore possible that a hospital separation should be linked to a permanent aged care resident on social leave (Figure 1).

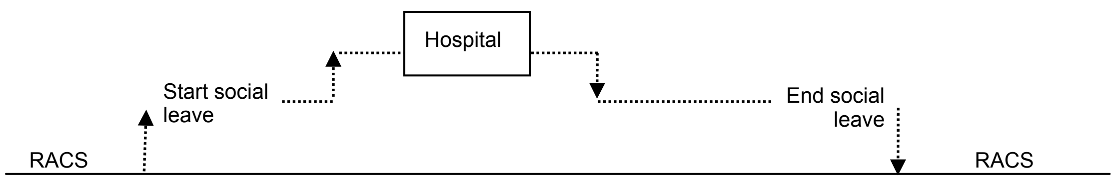


Figure 1: Hospital episode during social leave from residential aged care

Social leave happens less frequently than hospital leave. Because there are no exact dates for linking a hospital separation to a period of social leave, linking to RACS social leave was not considered in the feasibility study. However, such linkage is examined here in order to determine whether it should be attempted in the final linkage.

As for RACS admissions and hospital leave, a single-date strategy could be again used when linking to RACS social leave. However, for a link between a hospital separation and a period of RACS social leave to be valid, the period of social leave

[7] There are two types of social leave: claimable and non-claimable. A resident's usual subsidies are payable during claimable leave, but not during non-claimable leave.

must at least cover the period of the hospital episode.[8] Therefore, the linkage strategy could be made more precise by using ‘extended cover matching’; that is, by ensuring that the hospital episode lies within the period of RACS social leave. Again, the region of usual residence used when undertaking linkage to periods of RACS social leave is the region in which the residential aged care service is located, rather than the person’s usual residence prior to admission to that service.

Analysis in the feasibility study indicated that people going from hospital to the community tend to have shorter stays than those going to residential aged care (medians of 5 and 15 days, respectively, in NSW/ACT 1999–00; Table A6). Also, in 2001–02 the median length of residential aged care hospital leave was 6 days and that of claimable social leave was 2 days (90th percentiles were 26 and 8 days, respectively). The generally short nature of social leave suggests that the social leave/hospital stay period comparisons may provide quite a strong match criterion since a hospital episode is only likely to occur within relatively long periods of social leave. In addition, by flagging hospital episodes with a statistical admission, those hospital separations with known and unknown total length of stay can be identified, and, as for hospital leave, this information can inform the matching process.

Although the daily average number of social leave episodes covering hospital episodes of fewer than 5 days is more than the daily average number of admissions into residential aged care (Table A5), the relatively small numbers of hospital separations of various lengths mean that the probability of chance matches is quite small (under 0.5% for populations of up to 35,000 for all lengths of hospital episodes except overnight stays; Table 1). Because social leave was not considered for matching in the feasibility study, any estimate of limits for the false match rate are necessarily based on guessed match rates. However, it is highly likely that fewer hospital separations will link to periods of social leave from residential aged care than to periods of hospital leave. If, for example, the number of hospital separations linking to periods of social leave is one-fifth the number linking to episodes of RACS hospital leave, then for short hospital episodes (under 5 days) among large populations the false match rate for matches to social leave could be 70% or more.[9]

The above discussion shows that when there are low match rates the trade-off between achieving additional valid matches and increasing the number of false links due to chance matching should be considered. Since people going on social leave are most likely more robust than those going on hospital leave, the match rate between hospital separations and periods of social leave is expected to be less than one-fifth of

---

[8] If a period of claimable leave is directly followed by a period of non-claimable social leave, then the matching period should be the concatenation of the two. In practice, change of type of social leave does not happen very often: analysis of 2001–02 data showed that non-claimable social leave accounted for only 2.3% of all social leave, and these 1,200 episodes of social non-claimable leave were spread across only around 200 residents. Over a 5-year period from October 1997, there were 40 cases of non-claimable social leave directly preceding claimable social leave, and around 640 cases where non-claimable social leave followed straight on from claimable social leave.

[9] Adjusting the formula in footnote 6, the range for the false match rate is derived as $[(km/5{,}475)/(0.011*0.2*\mathrm{m}) \leq \mathrm{F} \leq (km/5{,}475)/(0.005*0.2*m)]$.

the hospital leave match rate. It is therefore important to compare the number of achieved matches with the number of expected chance matches before a final decision is made concerning the inclusion of social leave in the linkage process. If after linking, match rates to periods of social leave are found to be very low then including such leave in the candidate matches will add to the number of false matches without substantially contributing to the number of valid matches.

4.2 Looking at all possible links

The above discussion considered residential aged care admissions and leave returns separately, and it was seen that, when using the single-date matching employed in the feasibility study, the probability of any chance matches due to the distribution of birth dates is quite low for small and medium regions (less than 10,000 of one sex) for individual types of entries into residential aged care, even when using 3-date matching. The resulting false match rates within RACS entry type are generally also acceptably small, with the possible exception of matches to social leave. Restricting the matching to same-date matching to admissions and period-based matching for RACS leave results in acceptable false match rates for larger regions, again with the possible exception of social leave.

In practice we will be comparing hospital separations on a particular day with all different types of entries into residential aged care at a particular time; that is, with residential aged care admissions, returns from hospital leave and periods of social leave. The effect of chance matching on the reliability of matches when these three types of RACS entries are considered together is examined below.

Table 2 shows the residential aged care entries that could be considered for linking with particular hospital separations. In summary, among hospital separations (excluding same-day and statistical separations) the following comparisons can be made to identify links:

- All hospital separations can be compared with respite and permanent admissions.
- All hospital separations can be compared with RACS social leave encompassing the hospital episode.
- As an alternative to using single-date matching, hospital episodes can be compared with RACS hospital leave periods encompassing the hospital episode. Hospital episodes that start with a non-statistical admission can be compared with RACS hospital leave periods matching the hospital episode.
- Hospital episodes that start with a statistical admission (including hospital transfers) can be compared with RACS hospital leave periods ending on (or close to) the same date and encompassing the hospital episode.

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

Type of hospital episode	Respite admission	Permanent admission	Return from RACS hospital leave	Return to RACS social leave	Return to the community
	Type of entry associated with residential aged care				
With statistical admission	Single date matching[(a)]	Single date matching[(a)]	End-date cover matching[(b)]	Extended cover matching[(c)]	Linking not possible
With non-statistical admission	Single date matching[(a)]	Single date matching[(a)]	Period matching[(d)]	Extended cover matching[(c)]	Linking not possible

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

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

(c) RACS leave period covers the hospital episode.

(d) Linking on hospital separation start and end dates and RACS leave start and end dates.

Note: Table excludes same-day and statistical hospital separations.

The expected numbers of entries into/back to residential aged care that could be considered for matching to hospital episodes for areas of 10,000 and 35,000 people (single sex) are illustrated in Table A5. From Tables A5 and A7 it can be seen that, for a population of 10,000, over 100 days we would expect to see 80 (0.08 x 100) hospital separations of length 5 days that started with either a statistical or non-statistical admission. In comparison, over the 100 days we would expect 225 (100 x (0.90 + 0.47 + 0.88)) residential aged care entries (including 90 admissions) that could be candidates for matching to some of the separations with statistical/transfer admissions, depending on the day of the hospital separation. Furthermore, 188 of these residential aged care entries (100 x (0.90 + 0.10 + 0.88)) could be candidates for matching to a number of the separations with non-statistical admissions, again depending on the day of the hospital separation. For a population of 35,000 people, these expected numbers are commensurately larger.

Considering the above, the question then arises of finding, among all these possibilities, matches purely through chance just because two people have the same birth date. To answer this question, a similar approach can be taken as that used when considering particular types of residential aged care entries separately. That is, the two reliability measures, the probability of any chance matches and the false match rate, are calculated—this time taking into account all of the different comparisons being made.

Probability of any chance matches

The probability of making any chance matches between hospital separations and all candidate RACS entry events is derived in Appendix A1.1 using the same assumptions as before: that birth dates are spread across 15 years, that birth dates have a uniform distribution across time, and that match candidate residential aged care entries for a particular hospital separation all have different birth dates. Since the resulting equation involves only the expected daily number of hospital separations of different lengths and the expected daily number of candidate RACS matches, it can be used to explore theoretically the effect of different linkage strategies on the overall reliability of identified links.

False match rate

As discussed earlier, two numbers are required to estimate the false match rate: the expected number of chance matches and the total number of matches made. These are derived simply by summing across the expected numbers for the comparisons with the different RACS entries (see Appendix A1.2). The false match rate is then the total expected number of chance matches divided by the total number of matches made.

As for the probability of making any chance matches, the false match rate can be derived for a number of linkage strategies. However, because of limitations in the data available on achieved match rates (see Appendix A1.2), the upper limits of the false match range derived for the following discussion are biased upwards. Therefore, actual false match rates resulting from the final linkage strategy will be lower than those presented here.

Note that, while the probability of making any chance matches increases when aggregating across days (or regions), the expected false match rate remains the same whether separations for 1 day or 365 are being considered.

Comparing linkage strategies

The effect of chance matching when linkage between hospital separations and all candidate RACS entries (that is, RACS admissions and periods of hospital and social leave) are considered together is examined below. The effect of different linkage strategies for the various types of RACS entries on the overall reliability of the linkage strategy is examined for a number of likely strategies. The base matching strategy against which others are compared uses exact single-date matching of hospital separations to residential aged care admissions and returns from hospital leave, and allows hospital separations to match to periods of social leave covering the hospital episode. In addition, the same linkage strategy is used for matching to hospital episodes irrespective of their mode of admission (that is, statistical or non-statistical admission), and birth dates are assumed to be spread uniformly over a period of 15 years.

Using the base linkage strategy, for populations of up to 10,000 (single sex), the probability of getting any chance matches between hospital separations on a particular day and all candidate residential aged care entries is small—less than 1%. The corresponding maximum false match rate for a population of 10,000 people is 1.7% (Table 3). For larger populations, the probability of getting any chance matches can be quite substantial: for 35,000 people it is 7.2%, resulting in an upper limit for the false match rate of 5.9%.

Allowing 3-date matching to admissions

When linking hospital episodes to admissions into residential aged care, the separation date of the hospital episode is compared with the admission date of the RACS admission. Allowing 3-date matching to admissions rather than same-date matching increases both the overall probability of getting at least 1 chance match and the maximum total false match rate by around two-thirds (Table 3).

For small populations, such rises will affect the utility of the linked data set only marginally, with the chance match probability remaining under 1%, and the corresponding false match rate under 3%, for populations under 10,000. However, for populations greater than 25,000 the probability of getting at least 1 chance match is over 6% when 3-date admission matching is allowed, and the upper limit for the false match rate is at least 7%, rising to over 9% for populations of 35,000.

Table 3: Measures of match reliability looking at all possible matches

	Population 10,000 (single sex)		Population 35,000 (single sex)	
Linkage strategy	Probability of any chance matches (Q)	Range of false match rate (F)	Probability of any chance matches (Q)	Range of false match rate (F)
Base linkage strategy[(a)]	0.61	0.90–1.68	7.20	3.14–5.87
+ 3-date admission matching	1.02	1.52–2.67	11.86	5.31–9.36
+ hospital leave exact period matching	0.45	0.66–1.24	5.38	2.32–4.34
– linking to social leave	0.38	0.56–1.05	4.55	1.96–3.66
+ hospital leave exact period matching – linking to social leave	0.22	0.33–0.61	2.68	1.14–2.14
+ birth dates spread uniformly over 30 years	0.30	0.45–0.84	3.66	1.57–2.93
+ birth dates spread across 5-year age groups as per RACS admissions over 30 years (see Table A11)	0.40	0.59–1.10	4.79	2.07–3.86

(a) Base linkage strategy and assumptions:

- 15 years of birth dates, uniform distribution
- ignore admission mode of hospital separations
- single-date (exact end-date) matching only of separations to admissions and hospital leave returns
- allow separations to match to social leave covering the hospital episode.

Note: Calculations use average person separation/admission/leave rates (see Tables A1, A5, A7 and A8). Hospital separation rates exclude same-day episodes. Reliability measures take into account the distribution of hospital episodes by length of stay.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD; see also Tables A9 and A14.

Period matching to hospital leave

Insisting on period matching to RACS hospital leave, rather than using single-date end-date matching, reduces both the daily probability of making at least 1 chance match and the maximum false match rate by up to one-quarter (Table 3). For populations of 10,000, insisting on period matching reduces the estimated upper limit of the false match rate from 1.7% to 1.2%, while for populations of 35,000 the upper limit decreases from 5.9% to 4.3%.

Requiring exact period matching when linking hospital episodes to hospital leave reduces the number of potential matches, and therefore the rate of chance matches and the false match rate, because of the stricter linking rule. However, in some cases it is too strong a requirement and so will lead to missed matches. In particular, among hospital episodes with a statistical admission no allowance is made for the length of stay of the preceding hospital episode. Consequently, because the length of stay of the previous adjoining episode is unknown, comparisons with leave from residential aged care of the wrong length could happen. For people aged 65 and over, around 25% of hospital episodes ending in a transfer to another hospital are same-day episodes, with a further 50% being for 2 to 10 days (NSW/ACT 1999–00, Table A10). For statistical discharges, less than 1% are same-day episodes, and over half are at least 10 days long. For these hospital episodes single-date matching rather than period matching to hospital leave may be more appropriate.

Cover matching to social leave

Reducing the daily probability of making at least 1 chance match and the maximum false match rate by around one-third, excluding matching to social leave periods has a greater impact on the reliability of the linked data than insisting on period matching to RACS hospital leave (Table 3). For example, for larger populations of 35,000, excluding social leave reduces the probability of any chance matches to below 5% (down from 7.2%) and the upper limit of the false match rate to 3.7%, down from 5.9%.

The above effect is slightly overstated as achieved matches to social leave periods have not been included in the denominator (total number of achieved matches) because the relevant matching rate is unknown, but expected to be very low. As discussed earlier, the false match rate within links to RACS social leave may be quite high, depending on the linkage rate achieved for these RACS entries. If, after linking, match rates to periods of social leave are found to be very low then including such leave in the candidate matches will add to the number of false matches without substantially contributing to the number of valid matches.

Effects of birth date distribution and population size

The above estimates were derived assuming that birth dates for people leaving hospital and for people entering residential aged care span 15 years , and that birth dates for people leaving hospital or entering residential aged care are spread uniformly. These assumptions were made to simplify the calculations, and make the problem more tractable. As a consequence of limiting birth dates to 15 years, birth dates are assumed to be more concentrated than they are in practice, so that, for a population of a given size, people are assumed more likely to have coincident birth dates than will actually be the case. The effect of this assumption is therefore similar to assuming that smaller populations have the same chance of false matches as larger populations. The impact on the reliability measures of the birth date assumptions and of population size is examined below.

Distribution of birth dates

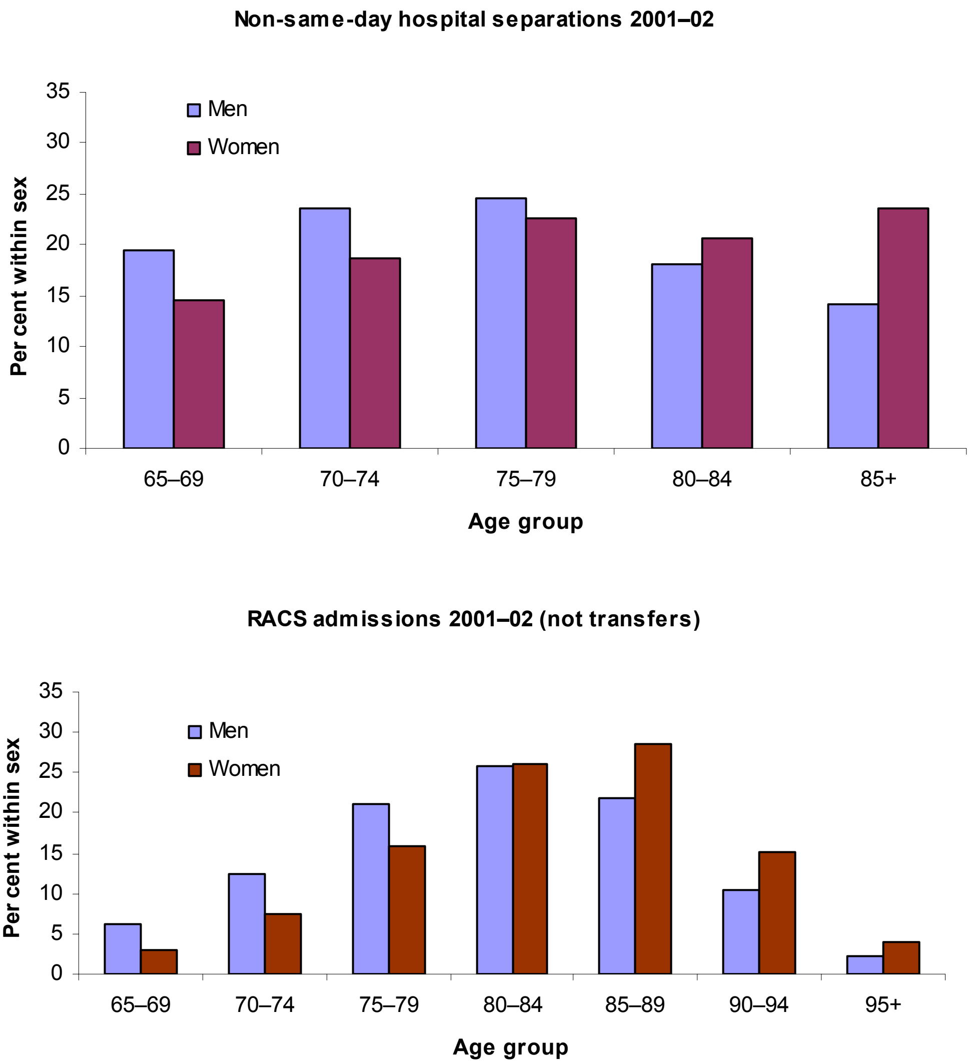
The estimates in the preceding discussion overstate the probability of chance matches because birth dates for people leaving hospital and for people entering residential aged care span more than the 15 years assumed in the calculations. As Figure 2 shows, birth dates for people aged 65 and over are spread over considerably more than 15 years. If it is assumed instead that birth dates are spread uniformly over 30 years, the probability of making at least 1 chance match roughly halves—to 0.3% per separation day for a population of 10,000, and to 3.7% for a population of 35,000 (Table 3). Similarly, the false match rate halves, with the upper limit remaining under 3% for all populations up to 35,000.

Assuming a uniform distribution of birth dates is also a broad simplification: with regard to 5-year age groups, the distribution of birth dates appears to be closer to normal than uniform, especially for residential aged care admissions (Figure 2). The probability of making at least 1 chance match is therefore likely to be somewhere between the two estimates based on 15 and 30 years of uniformly distributed birth dates.

With regard to 5-year age groups, the largest number of hospital separations (almost 25%) are in the 75–79 age group for both men and women; residential aged care admissions peak at ages 80–84 for men (26%) and 85–89 for women (29%). These distributions of birth dates mean that the probability of making any chance matches will be lower among people aged under 75 and over 90 than among those in the 75 to 90 age groups.

To get a feel for the effect of such clustering on the estimated reliability, consider the case when 29% of hospital separations and residential aged care entries are for a particular 5-year age group. For a large area of 35,000 people (single sex), this would mean that just over 10,000 people would be in the one age group. The daily probability of making at least 1 chance match within this group is estimated at 1.9%, with the false match rate being up to 5.1%, assuming a uniform distribution of birth dates across the 5 years (Tables A9 and A14). For all other age groups, these figures would be lower because of their smaller size.

The combined effect of the greater spread of birth dates and their non-uniform distribution can be illustrated by an example. If the birth date distributions for both hospital separations and residential aged care entries were the same as that for women admitted to residential age care in 2001–02 (see Table A11), overall for a population of 35,000 people the probability of getting any chance matches would be 4.8% per separation day, compared with probabilities of 3.7% and 7.2% assuming birth dates are spread uniformly over 30 and 15 years, respectively. The corresponding maximum false match rate would be 3.9%, compared with 2.1% and 5.9% assuming 30-year and 15-year uniform distributions.



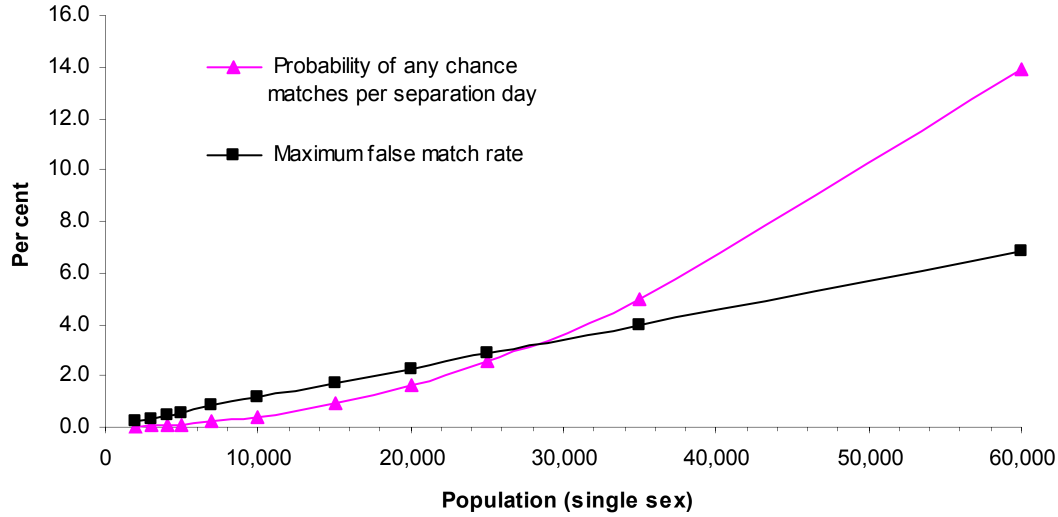
Source: Table A11.

Figure 2: Age distributions for hospital separations and residential aged care admissions, 2001–02

Overall, the actual probability of any chance matches and the maximum false match rate are both likely to be about one-third lower than the estimates obtained using the assumption contained in the base strategy—that birth dates are spread uniformly over 15 years.

Population size

As can be seen from the preceding discussion, the size of the population within which matching is taking place is a major factor affecting both the probability of making any chance matches and the overall false match rate. The effect of population size on the two measures of reliability is shown clearly in Figure 3. These estimates were derived assuming exact end-date matching only of hospital separations to admissions and RACS hospital leave returns, in conjunction with cover matching to social leave (as for the base linkage strategy). In addition, birth dates are assumed to be spread uniformly over 22 years (a compromise distribution which gives results similar to that using the distribution of birth dates seen for female RACS admissions).



Notes

1. Assumes single date (end date) matching only of separations to admissions and hospital leave returns, and allows extended cover matching to social leave.
2. Assumes birth dates are spread uniformly over 22 years (compromise distribution).
3. See notes to Table A15.

Source: Table A15.

Figure 3: The effect of population size on chance matches

From Figure 3 it can be seen that using end-date matching to admissions and RACS hospital leave results in both the probability of making any chance matches and the maximum false match rate being below 3% for populations of up to 25,000. Considering the trade-offs identified above, this would still be the case for populations of less than 20,000 if 3-date admission matching were allowed, and matches to hospital leave were restricted to period matching when hospital separations had non-statistical admissions. For populations greater than 20,000, allowing 3-date admission matching would result in too many chance matches. Excluding social leave from the matching process and restricting matches to hospital leave to period or cover matching, as appropriate, would result in both the probability of making any chance matches and the maximum false match rate being below 2% even for large populations: using this restricted strategy, the maximum false match rate is 0.8% for a population of 20,000, rising to 1.5% for a population of 35,000.

To minimise the effect of population size, the regions within which matching is taking place should be kept as small as possible. In the feasibility study, matching within postcodes and within SLA group were tested. The SLA group for a postcode is that set of SLAs that overlap the postcode, and postcodes with a common SLA in their respective groups are considered to match. It is proposed that such regional matching be done again as this allows for some variation in reporting the area of usual residence (see Section 2.2). In 2001, in 99% of cases the SLA group for a postcode had fewer than 15,000 men and 20,000 women aged 65 and over (Table A12). Because two-thirds of SLA groups have fewer than 2,000 people (single sex) aged 65 and over, under the linkage strategy illustrated in Figure 3, the average regional probability of any chance matches is estimated to be 0.2%, and the national false match rate is estimated to be under 0.4% (Table A15).

5 Discussion

The purpose of the anonymous linkage strategy developed by the Australian Institute of Health and Welfare was to find a pragmatic way of linking currently available data on hospital separations and residential aged care entries to provide a linked data set that could be used to investigate the hospital/aged care interface. Estimates based on the theoretical analysis presented in this report indicate that the Institute's linkage strategy results in acceptably low numbers of false matches, and so can be used to derive such a data set in a range of circumstances. As seen above, using a linkage strategy similar to that used in the feasibility study, but allowing matching to RACS social leave, the average regional daily probability of any chance matches is estimated to be 0.2%, and the national false match rate is estimated to be under 0.4%. Replicating the strategy used in the feasibility study—that is, excluding matching to RACS social leave—reduces the upper limit for the false match rate to 0.25%.

A key factor determining whether or not the AIHW strategy can be used is the number of candidate matches being considered: if this is too high, the probability of chance matches becomes too great and the false match rate increases accordingly. In the current application of the strategy, the number of candidate matches is primarily driven by the size of the population within which matches are to be made. By limiting the size of the population group for linkage, the number of chance matches can be kept to a level such that results from many types of analyses would not be affected by their presence.

As the false match rate increases with population size, using a single linkage strategy across all sizes results in a bias in the linked data: larger regions will be over-represented in the data compared with smaller regions. Since changing the linkage strategy can also affect the chance match rate, the linkage strategy could be adjusted according to the size of the population to achieve a near-uniform false match rate for all regions. However, such an approach will introduce other biases. This is because, although tightening the linkage strategy will reduce the false match rate, it is also likely to reduce the overall achieved match rate. Consequently, tailoring the linkage strategy to population size to reduce the false match rate could lead to under-representation of larger regions in the linked data. One possible way of reducing the effect of such biases on analyses is to use a uniform linkage strategy for all regions irrespective of size to obtain a uniform valid match rate, and then weighting the resulting links according to the regional false match rates to adjust for any over-representation.

In general, tightening linkage rules decreases the likelihood of making false matches. However, it also increases the number of missed matches. This issue is common to most linkage processes, and is not restricted to the AIHW linkage strategy. It has not been possible to examine the trade-off between missed and false matches, but it may be possible to investigate this issue in a proposed study to compare the AIHW linkage strategy with a rigorous name-based linkage.

5.1 Linkage rules for regional analysis

Australia-wide, the largest region within which matching will take place when using SLA groups is about 60,000 people (two SLA groups, allowing for overlap). If regional analysis is being carried out, for example within a smaller state, the largest matching region could be considerably smaller. Consequently, different linkage strategies could be used for different analyses to take advantage of the likely increase in valid matches when using less specific linkage strategies while keeping the false match rate under desired limits. The linkage strategy that can be adopted therefore depends on the size of the largest region within which matching is being undertaken. Findings in this report suggest that the following linkage rules could be used for linking for regional analyses, with the resulting false match rate remaining below 2% of matches for all but the largest regions (Figure 4).

All matching regions <10,000 single sex (small)

When the matching regions included in the area of interest have fewer than 10,000 people, a relatively loose linkage strategy can be employed without leading to unacceptable numbers of false matches. For example, linkage employing the following rules for linking to the three types of RACS events will result in a false match rate of under 2% for a matching region of 10,000 people:

- 3-date matching to RACS admissions
- Single-date (exact end-date) matching to RACS hospital leave
- Cover matching to RACS social leave.

Excluding matching to social leave, and allowing 3-date single-date matching to RACS hospital leave (to allow for death notification days), may also result in a false match rate of less than 2%, although analysis of the final linkage would be needed to confirm this.

Includes matching regions 10,000–25,000 single sex (medium)

When medium to large matching regions are included in the area of interest (10,000–25,000 people), the linkage strategy needs to be tightened up if we want to keep the false match rate below 2%, for example, for all matching regions. The following linkage strategy achieves this:

- Exact date matching to RACS admissions
- Period matching of hospital separations to RACS hospital leave
- Extended cover matching to social leave.

Using end-date cover matching to RACS hospital leave for the estimated 11% of hospital separations that begin with a statistical admission—rather than insisting on period matching for all hospital separations—could also be considered. Such an approach would increase the false match rate slightly, but would result in fewer missed matches.

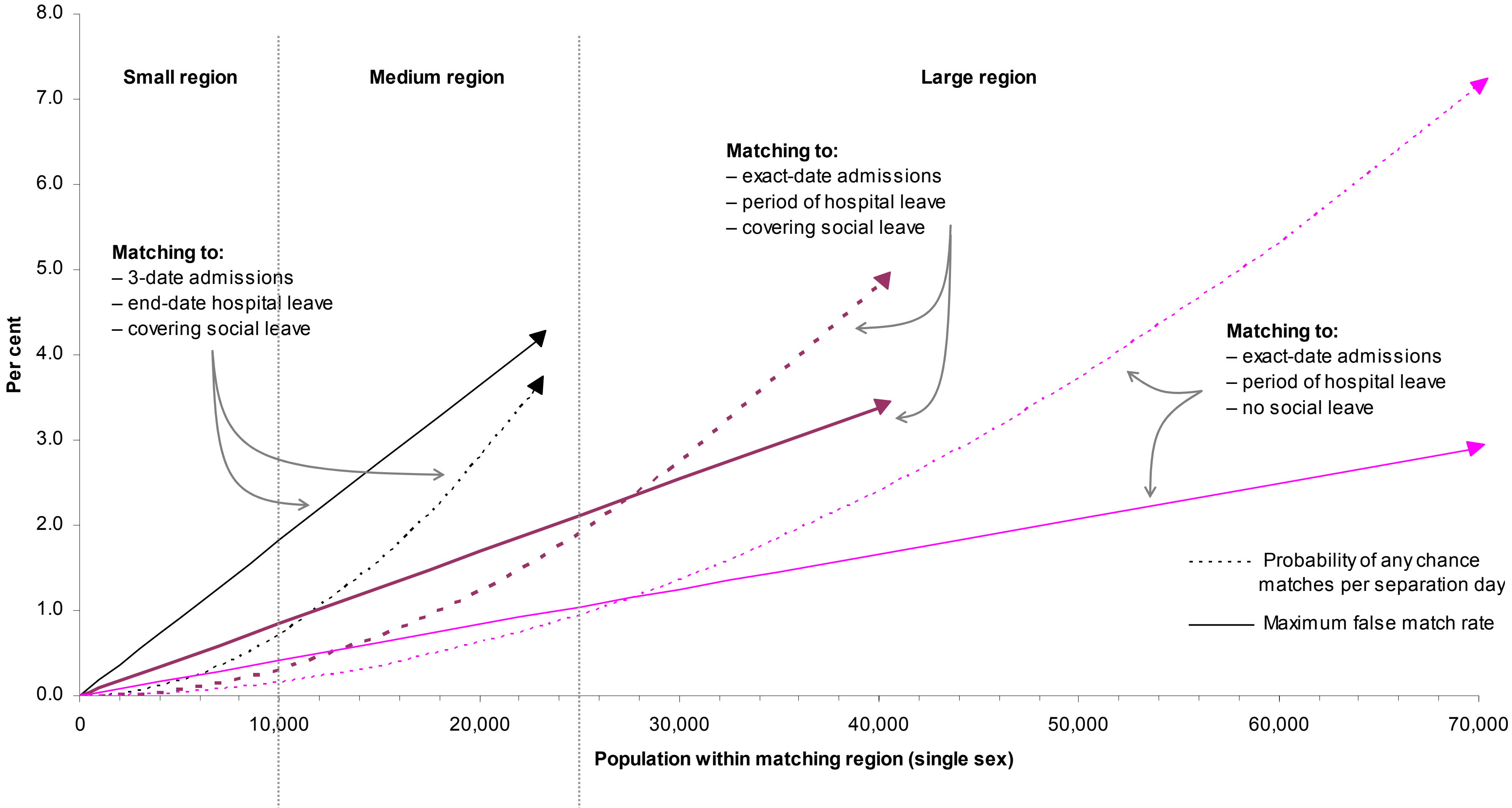
Includes matching regions >25,000 single sex (large)

For an area of interest which includes large matching regions (>25,000 people), the linkage strategy needs to be quite restrictive if we want to keep the false match rate below 2% for as many matching regions as possible. Note that since the information available for linking is limited, it is not always possible to keep the false match rate below some predetermined desirable level for all matching regions. The following linkage strategy has an estimated upper limit for the false match rate of 2% or less for all matching regions with fewer than 50,000 people:

- Exact date matching to RACS admissions
- Period matching of hospital separations to RACS hospital leave
- No matching to social leave.

As for medium size matching regions, using end-date cover matching to RACS hospital leave could be considered for hospital separations that begin with a statistical admission.

Note that the false match rate for linking to social leave needs to be investigated further once the linkage has been carried out to determine whether it should be included in the linkage for any of the population groups.



Note: Period matching to hospital leave ignores the admission type of hospital episodes.

Source: Table A16.

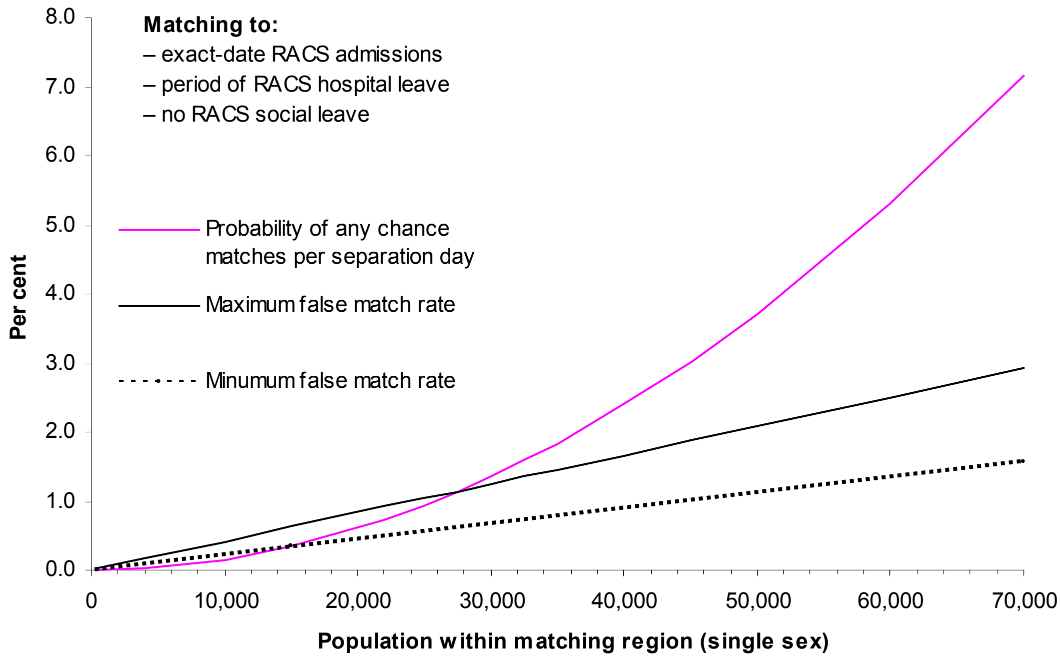
Figure 4: Chance matching in possible linkage strategies for regional analyses

5.2 Linkage rules for national analysis

Since national analyses necessarily include the largest matching regions, the linkage strategy given above for areas including large matching regions should be used across all regions if we want to limit the false match rate without introducing size-based biases in the valid matches. That is, the linkage strategy could include:

- Exact date matching to RACS admissions
- Period matching of hospital separations to RACS hospital leave
- No matching to social leave.

This strategy results in a false match rate of 2% or less for all matching regions with fewer than 50,000 men or women, and under 3% for populations of up to 70,000 (Figure 5). Furthermore, the lower limit for the false match rates suggests that the achieved false match rate could in fact be less than 2% for all matching regions under 70,000 people. Such a linkage strategy is more precise than the base linkage strategy discussed at the end of the last chapter, and will result in a false match rate of under 0.2%.



Note: Period matching to hospital leave ignores the admission type of hospital episodes.

Source: Table A16, and AIHW analysis.

Figure 5: Chance matching in a possible linkage strategy for national analyses

If regional false match rates larger than 2% are considered acceptable, looser match rules could be employed. For example, Figure 4 indicates that using the linkage strategy suggested for areas including medium size matching regions will result in false error rates of under 5% for nearly all matching regions nationally.

Although the linkage strategy has been developed for linking hospital separations to residential aged care entries, it could be used in any situation where similar data are available (that is, date of birth, sex, region and event dates). In particular, within the National Hospital Morbidity Database it could be used for linking adjacent hospital episodes for people to develop a data set reflecting all-of-hospital-stay rather than individual hospital episodes. Such linkage would assist in addressing a range of questions; for example, the effects of adverse events on the whole of hospital stay.

Appendix 1 Chance matches among all possible links

The measures of match reliability presented in Section 4.2 were calculated using equations derived below.

A1.1 Probability of any chance matches among all possible links

The probability of any chance matches for hospital separations on a given day is examined by looking at the reverse probability; that is, that the probability that no people associated with hospital separations on a particular day have the same birth date as the people associated with candidate residential aged care entries, assuming random distribution of birth dates among people separating from hospital and those entering residential aged care.

When considering chance matching between hospital separations and all candidate entries into residential aged care, the different matching strategies to be used for linking to admissions, RACS hospital leave and RACS social leave need to be taken into account. Consider the example where the distinction between different modes of admission for hospital episodes is ignored, and linkage is undertaken using exact single-date matching to RACS admissions, exact period matching to hospital leave from residential aged care—that is, no allowance is made for time spent in an emergency department or time taken to notify a residential aged care service about a death in hospital—and cover matching to RACS social leave, then the probability of having no chance matches is as follows:

P = *Pr{no separations on a particular day have same birth date as possibly matching RACS entries}*

= *Pr{no separations of a particular length have same birth date as any RACS admissions and any RACS hospital leave returns of the same length and ending on the hospital separation date and RACS social leave returns covering the period, for all lengths of hospital episode}*

$$= \prod_{l=1}^{\max length} Pr(\text{dates of birth of hospital separations length } l \neq \text{dates of birth of possibly matching RACS entries})$$

If linking is occurring 'by chance', then hospital separations and entries into residential aged care can be viewed as independent Poisson processes. Also, for independent processes we have:

Pr(no matches for hospital episodes of length l)

$\sum_{mk}$ = *Pr(no matches for hospital episodes of length l | m hospital separations of length l, k RACS possible matches) Pr(m hospital separations of length l, k RACS possible matches)*

= $\sum_{mk}$ *Pr(no matches for hospital episodes of length l | m, k) Pr(m, k | l)*

= $\sum_{mk}$ *Pr(no matches for hospital episodes of length l | m, k) Pr(m | l) Pr(k | l)*

Using the simplifying assumptions that all residential aged care entries which are candidates for matching have different birth dates, and that birth dates are spread uniformly across a 15-year period, the probabilities can then be derived from the assumption of independent Poisson processes in the manner set out below:

Pr(dates of birth of hospital episodes of length l ≠ dates of birth of possible RACS matches)

$$= \sum_m \sum_k Pr(\text{no matches} \mid l, m, k)\, Pr(m \mid l)\, Pr(k \mid l)$$

$$= \sum_m \sum_k \left(\frac{5{,}475 - k}{5{,}475} \right)^m \frac{\lambda_{Rl}{}^k \exp(-\lambda_{Rl})}{k!} \frac{\lambda_{Hl}{}^m \exp(-\lambda_{Hl})}{m!}$$

$$\cong \left(\frac{5{,}475 - \lambda_{Rl}}{5{,}475} \right)^{\lambda_{Hl}}$$

where λ_{Hl} is the expected daily number of hospital separations of length *l* and λ_{Rl} is the expected daily number of candidate RACs matches (including admissions) for a hospital episode of length *l*. We then get:

P =

$$\prod_{l=1}^{\max length} Pr(\text{dates of birth of hospital separations length } l \neq \text{dates of birth of possibly matching RACS entries})$$

$$\cong \prod_l \left(\frac{5{,}475 - \lambda_{Rl}}{5{,}475} \right)^{\lambda_{Hl}}$$

The probability of at least 1 hospital separation linking to a possibly-matching residential aged care entry purely due to the random distribution of birth dates is $Q=(1-P)$.

The above equation can easily be generalised to allow for different strategies for hospital separations with statistical and non-statistical admissions, and for different strategies for entries into residential aged care. Furthermore, since the above equation involves only the expected daily number of hospital separations of different lengths and the expected daily number of candidate RACS matches, it can be used to explore theoretically the effect of different linkage strategies on the reliability of identified links. For example, the length of hospital leave could be ignored when linking to RACS hospital leave so that single-date matching is used rather than period matching. Derivation of the resulting probability of chance matches resulting from using exact single-date matching to RACS admissions and hospital leave and extended cover matching to RACS social leave, ignoring the admission mode of hospital episodes, is presented in detail in Tables A7 and A8 for populations of 10,000 and 35,000, respectively.

A1.2 False match rate

Two numbers are required to estimate the false match rate: the expected number of chance matches and the estimated total number of matches made. When looking at all types of entries into residential aged care, the expected number of chance matches depends on the total number of comparisons being made, taking into account all comparisons between hospital separations and candidate residential aged care admissions, returns from hospital leave and periods of social leave. The total number of expected chance matches is simply the sum of the expected number of chance matches between hospital separations and the different types of aged care entries. This sum is then divided by the estimated total number of matches made (of all types) to get the false match rate.

Again consider the example where the distinction between different modes of admission for hospital episodes is ignored, and linkage is undertaken using: exact-date matching to RACS admissions, exact period matching to RACS hospital leave, and extended cover matching to RACS social leave. Then, using the same notation as above, the false match rate for all possible matches is estimated as follows:

$$F = \frac{\text{Total number of expected chance matches}}{\text{Estimated total number of matches between hospital separations and the different types of aged care entries}}$$

$$= \sum_l (\lambda_{Hl}\lambda_{Rl} / 5{,}475) \Big/ \sum_l \sum_t (\lambda_{Hl}\alpha_{tl})$$

$$= \sum_l (\lambda_{Hl}\lambda_{Rl} / 5{,}475) \Big/ \sum_l \lambda_{Hl} (\sum_t \alpha_{tl})$$

where α_{tl} is the achieved match rate with entries of type *t* for hospital episodes of length *l*.

As for the probability of making any chance matches, this equation can be generalised for other linkage strategies, and therefore can be used to examine the effectiveness of a wide range of strategies.

Calculating the false match rate using the above equation relies on the availability of estimates of achieved match rates under various linkage strategies. However, since the feasibility study is the source of information for minimum achieved match rates between different types of events, detailed information on match rates according to length of hospital episode and linkage strategy used is not available. Therefore it is not currently possible to adjust the expected match rates to reflect precisely the different linkage strategies that might be used when linking to the various RACS entry events. In the analysis in this paper, in order to estimate upper limits for the false match rate it has been assumed that match rates to episodes of RACS hospital leave do not change with the date linkage strategy being used (that is, single-date, period or cover matching) or the length of the period of leave, and that the linkage rate between hospital separations and periods of social leave is zero. That is, the false match rate is estimated by:

$$F = \sum_{l} (\lambda_{Hl} \lambda_{Rl} / 5{,}475) \Big/ \lambda_{H.} (\alpha_{RACS\ admissions} + \alpha_{RACS\ hospital\ leave})$$

The match rates, $\alpha_{\text{RACS admissions}}$ and $\alpha_{\text{RACS hospital leave}}$, used in this equation can be based on either exact-date or 3-date matching (see Section 3.2).

Appendix 2 Tables

Table A1: Expected number of hospital separations and residential aged care entries (or event rates) for an area (SLA group) with 70,000 people aged 65+

	Women	Men	Total
Nationally			
Hospital separations per 1,000 people annually[(a)]	416	501	442
RACS admissions per 1,000 people annually[(b)]	40	28	35
Social leave returns per 1,000 RACS permanent places annually			386
Hospital leave returns per 1,000 RACS permanent places annually			512
Regionally			
SLA group population	**38,889**	**31,111**	**70,000**
Expected hospital separations per day in region	44	43	85
Expected hospital separations per day in region with particular DOB[(c)]	0.00809	0.00781	0.01549
Expected RACS admissions per day in region	4	2	7
Expected RACS admissions per day in region with particular DOB	0.00078	0.00043	0.00121
Expected social returns per day in region[(d)]			4
Expected social returns per day in region with particular DOB			0.00074
Expected hospital leave returns per day in region[(d)]			5
Expected hospital leave returns per day in region with particular DOB			0.00098

(a) Based on 2000–01 hospital separation rates by sex (excludes same-day separations, but includes deaths and statistical discharges (including transfers) (AIHW 2002:Table A26.1). In NSW/ACT in 1999–00 4.8% of separations were for deaths, 4.1% for statistical discharges and 8.5% for transfers to another hospital.

(b) Based on 2001–02 admission rates by sex (AIHW 2003b; ABS 2003).

(c) Assumes people were born in a 15-year period (i.e. 365*15 = 5,475 distinct birth dates distributed uniformly).

(d) Based on 52,629 social leave episodes and 69,937 hospital leave episodes (in 2001–02) and proportional distribution of 136,500 permanent RACS places across regions (using estimated residential population 2002 for population 65+, giving 3,837 places in the region).

Notes

1. Separations or admissions per day in region = rate per 1,000 people x SLA population / 1,000 / 365 = A
2. Separations or admissions per day in region with particular DOB = A/5,475 = B
3. Leave returns per day in region = rate per 1,000 places x RACS permanent places in the region/ 1,000 = C
4. Leave returns per day in region with particular DOB = rate per 1,000 places x RACS permanent places / 1,000 = C/5,475

Sources: ABS 2003; AIHW 2002, 2003b, AIHW analysis of ACCMIS database.

Table A2: Probability of no matches by chance and expected number of chance matches when comparing hospital separations and residential aged care admissions/returns

Hospital separations	RACS admissions/leave returns											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**50**
	Probability of no matches by chance											
1	>0.9995	>0.9995	>0.9995	>0.9995	>0.9995	>0.9995	0.999	0.999	0.999	0.999	0.999	0.991
5	0.999	0.998	0.997	0.996	0.995	0.995	0.994	0.993	0.992	0.991	0.990	0.955
10	0.998	0.996	0.995	0.993	0.991	0.989	0.987	0.985	0.984	0.982	0.980	0.912
15	0.997	0.995	0.992	0.989	0.986	0.984	0.981	0.978	0.976	0.973	0.970	0.871
20	0.996	0.993	0.989	0.985	0.982	0.978	0.975	0.971	0.968	0.964	0.961	0.832
30	0.995	0.989	0.984	0.978	0.973	0.968	0.962	0.957	0.952	0.947	0.941	0.759
40	0.993	0.985	0.978	0.971	0.964	0.957	0.950	0.943	0.936	0.929	0.923	0.693
50	0.991	0.982	0.973	0.964	0.955	0.947	0.938	0.929	0.921	0.913	0.904	0.632
60	0.989	0.978	0.968	0.957	0.947	0.936	0.926	0.916	0.906	0.896	0.886	0.577
70	0.987	0.975	0.962	0.950	0.938	0.926	0.914	0.903	0.891	0.880	0.869	0.526
80	0.985	0.971	0.957	0.943	0.930	0.916	0.903	0.890	0.877	0.864	0.851	0.480
90	0.984	0.968	0.952	0.936	0.921	0.906	0.891	0.877	0.862	0.848	0.834	0.438
120	0.978	0.957	0.936	0.916	0.896	0.877	0.858	0.839	0.821	0.803	0.786	0.333
	Expected number of chance matches											
1	<0.0005	<0.0005	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.009
5	0.001	0.002	0.003	0.004	0.005	0.005	0.006	0.007	0.008	0.009	0.010	0.046
10	0.002	0.004	0.005	0.007	0.009	0.011	0.013	0.015	0.016	0.018	0.020	0.091
15	0.003	0.005	0.008	0.011	0.014	0.016	0.019	0.022	0.025	0.027	0.030	0.137
20	0.004	0.007	0.011	0.015	0.018	0.022	0.026	0.029	0.033	0.037	0.040	0.183
30	0.005	0.011	0.016	0.022	0.027	0.033	0.038	0.044	0.049	0.055	0.060	0.274
40	0.007	0.015	0.022	0.029	0.037	0.044	0.051	0.058	0.066	0.073	0.080	0.365
50	0.009	0.018	0.027	0.037	0.046	0.055	0.064	0.073	0.082	0.091	0.100	0.457
60	0.011	0.022	0.033	0.044	0.055	0.066	0.077	0.088	0.099	0.110	0.121	0.548
70	0.013	0.026	0.038	0.051	0.064	0.077	0.089	0.102	0.115	0.128	0.141	0.639
80	0.015	0.029	0.044	0.058	0.073	0.088	0.102	0.117	0.132	0.146	0.161	0.731
90	0.016	0.033	0.049	0.066	0.082	0.099	0.115	0.132	0.148	0.164	0.181	0.822
120	0.022	0.044	0.066	0.088	0.110	0.132	0.153	0.175	0.197	0.219	0.241	1.096

Notes

1. Assumes all RACS admissions have a different birth date.
2. Assumes that birth dates are over a 15-year period and distributed uniformly.
3. Probability of all hospital separations having different birth dates from all RACS admissions
   = [Probability of a single hospital separation having a different birth date from all RACS admissions]$^{(\text{no. hospital separations})}$
   = [(total number of allowed birth dates)/(total number of possible birth dates)]$^{(\text{no. hospital separations})}$
   = [(5,475 – no.RACS admissions)/5,475]$^{(\text{no. hospital separations})}$
4. Expected number of hospital separations having same dates of birth as a RACS admission/return
   = [Probability of a single hospital separation matching to a RACS admission/return] x [number of comparisons]
   = [Probability of a single hospital separation matching to a RACS admission/return] x [number of hospital separations] x [number of RACS admission/returns]
   = [number of hospital separations] x [number of RACS admission/returns] / 5,475

Table A3: Expected number of hospital separations and residential aged care admissions per day by population size

	Expected hospital separations per day[a]		Expected RACS admissions (respite and permanent) per day[b]	
	Men	Women	Men	Women
Average number per day per 1,000 people	**1.37**	**1.14**	**0.08**	**0.11**
Population size (65+)				
500	0.7	0.6	—	0.1
1,000	1.4	1.1	0.1	0.1
2,000	2.7	2.3	0.2	0.2
3,000	4.1	3.4	0.2	0.3
4,000	5.5	4.6	0.3	0.4
5,000	6.9	5.7	0.4	0.5
10,000	13.7	11.4	0.8	1.1
15,000	20.6	17.1	1.1	1.6
20,000	27.5	22.8	1.5	2.2
25,000	34.3	28.5	1.9	2.7
30,000	41.2	34.2	2.3	3.3
35,000	48.1	39.9	2.7	3.8
40,000	55.0	45.6	3.0	4.4
45,000	61.8	51.3	3.4	4.9

(a) Based on 2000–01 hospital separation rates by sex (excludes same-day separations) (see Table A1).

(b) Based on 2001–02 admission rates by sex (see Table A1).

Source: AIHW estimates derived using data in Table A1.

Table A4: Expected number of hospital separations and residential aged care admissions per day, probabilities of any chance matches, and expected number of false matches, for selected populations

				Probability of any chance matches		Expected number of chance matches	
Population (65+)	Sex group	Daily expected hospital separations	Daily expected RACS admissions	Exact admission date match	3-date match	Exact admission date match	3-date match
4,000	Women	4.6	0.4	<0.1%	<0.3%	<0.0005	0.001
10,000	Men	13.7	0.8	<0.3%	<0.8%	0.002	0.006
25,000	Women	28.5	2.7	<1.6%	4.3%	0.014	0.042
35,000	Women	39.9	3.8	2.9%	7.1%	0.028	0.083

Notes

1. Based on 2000–01 hospital separation rates by sex (excludes same-day separations) (see Table A1).
2. Assumes all RACS admissions on a particular day are for people with a different birth date.
3. Assumes that birth dates are over a 15-year period and distributed uniformly.
4. ‘Probability of a chance match’ is the probability—due solely to the random distribution of birth dates within a gender group—that the date of birth associated with a hospital separation on a given day is the same as a date of birth associated with a residential aged care admission on the same day
   = 1 – [Probability of all hospital separations having different birth dates from all RACS admissions].

Sources: Data in Tables A1, A2 and A3.

Table A5: Expected number of RACS entries for matching, by type of entry and length of hospital episode

Length of hospital episode	Population 10,000 (single sex)			Population 35,000 (single sex)		
	Hospital leave	Hospital leave	Social leave	Hospital leave	Hospital leave	Social leave
Days	[a]Covers hospital episode, and starts on or 1 day before hospital admission	[b]Covers hospital episode, and ends on separation date	[c]Covers hospital episode	Covers hospital episode, and starts on or 1 day before hospital admission	Covers hospital episode, and ends on separation date	Covers hospital episode
1	0.18	0.77	2.2	0.62	2.69	7.71
2	0.13	0.66	1.62	0.47	2.32	5.68
3	0.12	0.59	1.27	0.42	2.07	4.45
4	0.11	0.53	1.04	0.39	1.85	3.66
5	0.10	0.47	0.88	0.35	1.65	3.09
6	0.09	0.42	0.76	0.32	1.46	2.65
7	0.08	0.37	0.65	0.27	1.30	2.29
8	0.06	0.33	0.57	0.22	1.14	1.99
9	0.05	0.29	0.50	0.18	1.02	1.75
10	0.05	0.26	0.44	0.16	0.92	1.54
11	0.04	0.24	0.39	0.14	0.84	1.36
12	0.03	0.22	0.34	0.12	0.77	1.2
13	0.03	0.20	0.30	0.11	0.70	1.05
14	0.03	0.18	0.26	0.10	0.65	0.93
15	0.02	0.17	0.23	0.08	0.59	0.81
16	0.02	0.16	0.21	0.07	0.55	0.72
17	0.02	0.15	0.18	0.07	0.51	0.64
18	0.02	0.14	0.16	0.06	0.47	0.57
19	0.02	0.13	0.14	0.06	0.44	0.50
20	0.02	0.12	0.13	0.05	0.41	0.44
21	0.01	0.11	0.11	0.05	0.39	0.38
22	0.01	0.10	0.09	0.04	0.36	0.33
23	0.01	0.10	0.08	0.04	0.34	0.29
24	0.01	0.09	0.07	0.04	0.32	0.25
25	0.01	0.08	0.06	0.04	0.30	0.21
26	0.01	0.08	0.05	0.03	0.28	0.17
27	0.01	0.07	0.04	0.03	0.26	0.14
28	0.01	0.07	0.03	0.03	0.24	0.10
29	0.01	0.06	0.02	0.03	0.23	0.07
30+	0.07	0.27	0.01	0.02	0.94	0.05
Respite and permanent admissions		0.9			3.3	

(a) $= E(l) + E(l+1)$, where l=length of hospital stay, and E(l)=expected number of hospital leave returns of length l in a day.

(b) $= \sum_{j=l}^{\max} E(j)$ where l=length of hospital stay, and $E(l)$=expected number of hospital leave returns of length l in a day.

(c) $= \sum_{i=0}^{\max-l} \sum_{j=l+i}^{\max} E(j \mid leave\ ends\ i\ days\ after\ hospital\ separation\ of\ length\ l)$

where l=length of hospital stay, and
$E(l)$=expected number of hospital leave returns of length l in a day.

Sources: Table A1 for person admission/leave return rates; AIHW analysis of ACCMIS database for distributions of length of leave for leave ending in 2001–02.

Table A6: Length of episode by type of client movement (days)

	Median	90th percentile	95th percentile
Hospital, all separations (NSW/ACT 1999–00)	5	18	n.a.
Hospital to RACS, acute care separations (NSW/ACT 1999–00)	5	15	n.a.
Hospital to RACS (all separations, NSW/ACT 1999–00)	15	45	n.a.
Hospital to RACS (acute care separations, NSW/ACT 1999–00)	14	37	n.a.
Hospital to community (all separations, NSW/ACT 1999–00)	5	17	n.a.
Hospital to community (acute care separations, NSW/ACT 1999–00)	4	15	n.a.
Hospital leave from RACS	6	26	35
Social leave (claimable) from RACS	2	8	14

n.a. not available from source publication.

Notes

1. Hospital separations are for people aged 65 and over; RACS leave episodes are for all ages (4.5% of permanent residents at December 2000 were aged under 65; AIHW 2001:28).
2. Hospital separations exclude same-day episodes, and those ending in transfer to another hospital, change in care type or death.

Sources: AIHW 2003a:Illustration 4; AIHW analysis of ACCMIS.

Table A7: Expected number of hospital separations per day, possibly matching residential aged care admissions per day, and probabilities of no matches by chance, for an area of 10,000 (single sex)

Length of hospital episode (days)	Hospital daily separation rates	[(a)]RACS daily entry rates for matching	Probability of no chance matches per separation day (%)
1	2.25	3.92	99.84
2	1.37	3.34	99.92
3	1.09	2.99	99.94
4	0.95	2.76	99.95
5	0.80	2.60	99.96
6	0.71	2.47	99.97
7	0.70	2.37	99.97
8	0.58	2.29	99.98
9	0.46	2.22	99.98
10	0.39	2.16	99.98
11	0.33	2.11	99.99
12	0.27	2.06	99.99
13	0.25	2.02	99.99
14	0.24	1.98	99.99
15	0.19	1.95	99.99
16	0.15	1.92	99.99
17	0.12	1.90	>99.995
18	0.11	1.88	>99.995
19	0.09	1.86	>99.995
20	0.09	1.84	>99.995
21	0.10	1.83	>99.995
22	0.08	1.81	>99.995
23	0.06	1.80	>99.995
24	0.05	1.79	>99.995
25	0.05	1.78	>99.995
26	0.04	1.77	>99.995
27	0.04	1.76	>99.995
28	0.04	1.75	>99.995
29	0.04	1.74	>99.995
30+[(b)]	0.45	1.73	99.99
Total	**12.12**		**99.39%**
Probability of making at least 1 match by chance (%)			**0.61%**

(a) Includes all RACS admissions on the day of hospital separation (daily rate of 0.9 for the area), hospital leave that ends on the same day (daily rate of 0.8), and social leave that covers the hospital episode (see also Table A5).

(b) Assumes maximum stay of 30 days: 7.8% and 1.3% of hospital and social leave periods, respectively, are at least 30 days long (2001–02); 13.2% and 2.2% of statistical and non-statistical admission hospital episodes, respectively, are for at least 30 days (NSW/ACT 1999–00). Assumption will lead to an understatement of probability of making no chance matches.

Notes

1. Uses average person separation/admission/leave rates. Hospital separation rates include all non-same-day episodes.
2. Distribution of hospital episodes is based on hospital episodes excluding statistical discharges and transfers to other hospitals.
3. Assumes all RACS admissions have a different birth date.
4. Assumes that birth dates are over a 15-year period and distributed uniformly.
5. Probabilities derived using formulae in text, and daily rates in table.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD.

Table A8: Expected number of hospital separations per day, possibly matching residential aged care admissions per day, and probabilities of no matches by chance, for an area of 35,000 (single sex)

Length of hospital episode (days)	Hospital daily separation rates	[(a)]RACS daily entry rates for matching	Probability of no chance matches per separation day (%)
1	7.88	13.72	98.04
2	4.81	11.69	98.98
3	3.82	10.46	99.27
4	3.32	9.67	99.42
5	2.80	9.10	99.54
6	2.50	8.66	99.61
7	2.45	8.30	99.63
8	2.02	8.00	99.71
9	1.62	7.76	99.77
10	1.36	7.55	99.81
11	1.17	7.37	99.84
12	0.96	7.21	99.87
13	0.86	7.06	99.89
14	0.84	6.94	99.89
15	0.65	6.82	99.92
16	0.52	6.73	99.94
17	0.44	6.65	99.95
18	0.40	6.58	99.95
19	0.33	6.51	99.96
20	0.33	6.45	99.96
21	0.34	6.39	99.96
22	0.26	6.34	99.97
23	0.21	6.30	99.98
24	0.18	6.26	99.98
25	0.17	6.22	99.98
26	0.15	6.18	99.98
27	0.15	6.15	99.98
28	0.15	6.11	99.98
29	0.14	6.08	99.98
30+[(b)]	1.57	6.06	99.83
Total	**42.41**		**92.80%**
Probability of making at least 1 match by chance (%)			**7.20%**

(a) Includes all RACS admissions on the day of hospital separation (daily rate of 3.3 for the area), hospital leave that ends on the same day (daily rate of 2.7), and social leave that covers the hospital episode (see also Table A5).

(b) Assumes maximum stay of 30 days: 7.8% and 1.3% of hospital and social leave periods, respectively, are at least 30 days long (2001–02); 13.2% and 2.2% of statistical and non-statistical admission hospital episodes, respectively, are for at least 30 days (NSW/ACT 1999–00). Assumption will lead to an understatement of probability of making no chance matches.

Note: See notes to Table A7.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD.

Table A9: Sensitivity to assumptions for deriving probability of making any chance matches

Assumptions	Population 10,000 (single sex)	Population 35,000 (single sex)
Across total population	**Probability of making any chance matches per separation day (%)**	
Standard assumptions: • 15 years of birth dates, uniform distribution • same-date matching only of separations to admissions and hospital leave returns • allow separations to match to social leave covering the hospital episode	0.61	7.20
Standard assumptions, excluding matching to social leave	0.38	4.55
Standard assumptions, allowing 3-date matching to admissions	1.02	11.86
Standard assumptions, requiring exact period matching to hospital leave	0.45	5.38
Standard assumptions, requiring exact period matching to hospital leave, excluding matching to social leave	0.22	2.68
Standard assumptions, birth dates spread uniformly over 30 years	0.30	3.66
Standard assumptions, birth dates spread as per RACS admissions over 30 years (see Table A11)	0.40	4.79
Within a 5-year age group		
Standard assumptions, 33.3% in each of three 5-year age groups	within all age groups: 0.20%	within all age groups: 2.46%
Standard assumptions, birth dates spread uniformly over 30 years, 16.7% in each of six 5-year age groups	within all age groups: 0.05%	within all age groups: 0.62%
Standard assumptions, birth dates spread over 30 years, 10,000 (29%) in largest 5-year age group (assumes same age distribution for both hospital separations and RACS entries, and uniform distribution within 5-year age groups)	within largest age group: 0.15%	within largest age group: 1.87%

Note: See notes to Table A7.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD.

Table A10: Distribution of length of hospital episode for separations ending either in a transfer to another hospital or in a statistical discharge (NSW/ACT 1999–00)

Length of hospital episode (days)	Transfers to another hospital		Statistical discharges	
	Per cent	Cumulative per cent	Per cent	Cumulative per cent
0	24.9	24.9	0.6	0.6
1	12.0	36.9	6.0	6.5
2	5.8	42.7	4.1	10.7
3	4.8	47.4	4.9	15.5
4	4.6	52.1	5.1	20.7
5	4.2	56.3	5.4	26.0
6	4.0	60.3	5.9	32.0
7	4.4	64.6	5.6	37.6
8	4.0	68.6	5.3	42.9
9	3.5	72.1	4.5	47.4
10	2.9	75.0	3.9	51.3
11+	25.0	100.0	48.7	100.0
Total	**100.0**		**100.0**	

Source: AIHW analysis of NHMD.

Table A11: Age distribution for hospital separations and residential aged care admissions, 2001–02

	65–69	70–74	75–79	80–84	85–89	90–94	95+	Total	
	Per cent							Per cent	Number
RACS admissions (excludes transfers to care of same type)									
Men	6.28	12.30	21.20	25.83	21.74	10.48	2.18	100.0	30,615
Women	2.96	7.35	15.85	26.10	28.61	15.20	3.94	100.0	55,505
Hospital separations (excluding same-day)[(a)]									
Men	19.50	23.56	24.68	18.15		[(b)]14.11		100.0	516,281
Women	14.52	18.62	22.56	20.65		[(b)]23.65		100.0	567,315

(a) Includes all separations for hospital episodes for stays of at least 1 day, including those ending in death, transfer or statistical discharge.

(b) Figures for 85+.

Sources: AIHW analysis of ACCMIS database; NHMD data cube.

Table A12: Distribution of postcodes by size of their associated SLA groups, 2001

Population size (65+)	Men		Women	
	Number	Per cent	Number	Per cent
0–499	859	32.8	848	32.4
500–999	532	20.3	450	17.2
1,000–1,999	391	14.9	385	14.7
2,000–2,999	204	7.8	172	6.6
3,000–3,999	162	6.2	145	5.5
4,000–4,999	131	5.0	152	5.8
5,000–7,499	146	5.6	188	7.2
7,500–9,999	88	3.4	100	3.8
10,000–14,999	86	3.3	106	4.0
15,000–19,999	11	0.4	53	2.0
20,000–24,999	7	0.3	9	0.3
25,000–34,999	1	—	9	0.3
35,000–49,999	—	—	1	—
Total (postcodes)	**2,618**	**100.0**	**2,618**	**100.0**
Mean size (people)	**2,244**		**2,828**	
Maximum size (people)	**31,079**		**40,701**	

Source: AIHW analysis of ABS demographic statistics.

Table A13: Sensitivity of probability of making any chance matches to population size

Population (single sex, 65+)	Standard assumptions	Standard assumptions, but requiring period matching to hospital leave	Standard assumptions, but excluding matching to social leave	Standard assumptions, but allowing 3-date matching to admissions
	Probability of making any chance matches per separation day (%)			
500	<0.005	<0.005	<0.005	<0.005
1,000	0.01	<0.005	<0.005	0.01
2,000	0.02	0.02	0.02	0.04
5,000	0.15	0.11	0.09	0.26
10,000	0.61	0.45	0.38	1.02
15,000	1.36	1.01	0.85	2.29
20,000	2.41	1.79	1.51	4.03
25,000	3.74	2.78	2.35	6.23
35,000	7.20	5.38	4.55	11.86
50,000	14.14	10.67	9.07	22.72
60,000	19.72	15.01	12.80	31.01

Notes

1. Standard assumptions:
   - 15 years of birth dates, uniform distribution
   - Same-date matching only of separations to admissions and hospital leave returns
   - allow separations to match to social leave covering the hospital episode.
2. See notes to Table A7.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD.

Table A14: Sensitivity to assumptions for deriving false match rate

Assumptions	Population 10,000 (single sex, 65+)		Population 35,000 (single sex, 65+)	
Across total population	**False match rate (%)**			
	Minimum	Maximum	Minimum	Maximum
Standard assumptions: • 15 years of birth dates, uniform distribution • same-date matching only of separations to admissions and hospital leave returns • allow separations to match to social leave covering the hospital episode	0.90	1.68	3.14	5.87
Standard assumptions, excluding matching to social leave	0.56	1.05	1.96	3.66
Standard assumptions, allowing 3-date matching to admissions	1.52	2.83	5.31	9.90
Standard assumptions, requiring exact period matching to hospital leave	0.66	1.24	2.32	4.34
Standard assumptions, requiring exact period matching to hospital leave, excluding matching to social leave	0.33	0.61	1.14	2.14
Standard assumptions, birth dates spread uniformly over 30 years	0.45	0.84	1.57	2.93
Standard assumptions, birth dates spread as per RACS admissions over 30 years (see Table A11)	0.59	1.10	2.07	3.86
Within a 5-year age group				
Standard assumptions, 33.3% in each of three 5-year age groups, within all age groups	0.90	1.68	3.14	5.87
Standard assumptions, birth dates spread uniformly over 30 years, 16.7% in each of six 5-year age groups, within all age groups	0.45	0.84	1.57	2.93
Standard assumptions, birth dates spread over 30 years, 10,000 (29%) in largest 5-year age group (assumes same age distribution for both hospital separations and RACS entries, and uniform distribution within 5-year age groups), within largest age group	0.78	1.46	2.73	5.10

Note: False match range derived using formula in text. This provides an upper limit as any matches to social leave will reduce the rate. See also notes to Table A7.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD.

Table A15: Probability of any chance matches per separation day and maximum false match rate by population size, using base linkage strategy (per cent)

Size of SLA group (single sex, 65+)	Postcodes	Probability of any chance matches per separation day	Maximum false match rate
0–499	32.39	<0.005	0.06
500–999	17.19	<0.005	0.11
1,000–1,999	14.71	0.02	0.23
2,000–2,999	6.57	0.04	0.34
3,000–3,999	5.54	0.07	0.46
4,000–4,999	5.81	0.10	0.57
5,000–7,499	7.18	0.23	0.86
7,500–9,999	3.82	0.41	1.14
10,000–14,999	4.05	0.93	1.71
15,000–19,999	2.02	1.65	2.29
20,000–24,999	0.34	2.56	2.86
25,000–34,999	0.34	4.96	4.00
35,000–60,000	0.04	13.90	6.86
Total	**100.0**	[(a)]**0.15**	**0.40**

(a) Average probability of any chance matches across areas.

Notes

1. Assumes single-date (end-date) matching only of separations to admissions and hospital leave returns, and allows cover matching to social leave.
2. Assumes birth dates are spread uniformly over 22 years (compromise distribution).
3. Estimated maximum false match rate may be too low because the combined size of two SLA groups can be larger than SLA group for one postcode. To minimise this effect, the maximum false match rate is derived using the upper limit of the population group.
4. False match range derived using formula in text. This provides an upper limit as any matches to social leave will reduce the rate.
5. SLA group distribution is that for women in 2001 (Table A12).

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD.

Table A16: Probability of any chance matches per separation day and maximum false match rate by population size using approximation to proposed linkage strategy (per cent)

Linkage strategy	Population (single sex, 65+)	Probability of any chance matches per separation day	Maximum false match rate
• 3-date matching to admissions • Exact-date matching to hospital leave • Cover matching to social leave	499	0.002	0.091
	999	0.007	0.182
	1,999	0.028	0.364
	2,999	0.063	0.547
	3,999	0.112	0.729
	4,999	0.175	0.911
	5,999	0.252	1.094
	8,499	0.505	1.549
	9,999	0.699	1.823
	14,999	1.566	2.734
• Exact-date matching to admissions • Period matching to hospital leave • Cover matching to social leave	10,000	0.307	0.847
	15,000	0.690	1.269
	17,499	0.938	1.481
	19,999	1.223	1.692
	22,499	1.545	1.904
	24,999	1.904	2.115
• Exact-date matching to admissions • Period matching to hospital leave • No matching to social leave	500	—	0.021
	4,000	0.024	0.166
	10,000	0.151	0.416
	15,000	0.340	0.624
	22,000	0.730	0.916
	25,000	0.941	1.041
	27,499	1.138	1.145
	29,999	1.353	1.249
	32,499	1.586	1.353
	34,999	1.840	1.457
	40,000	2.392	1.665
	45,000	3.018	1.873
	50,000	3.713	2.081
	60,000	5.303	2.497
	70,000	7.149	2.914

Notes

1. Assumes birth dates are spread uniformly over 22 years (compromise distribution).
2. Period matching to hospital leave ignores the difference between hospital episodes with and without a statistical admission.
3. False match range derived using formula in text. This provides an upper limit as any matches to social leave will reduce the rate.

Sources: AIHW analysis of ACCMIS database; AIHW analysis of NHMD.

Abbreviations and symbols

ABS	Australian Bureau of Statistics
ACAT	Aged Care Assessment Team
ACCMIS	Aged and Community Care Management Information System
AHMAC	Australian Health Ministers' Advisory Council
AIHW	Australian Institute of Health and Welfare
DOB	Date of birth
NHMD	National Hospital Morbidity Database
RACS	Residential aged care service
SLA	Statistical Local Area
—	nil or rounded to zero

References

ABS (Australian Bureau of Statistics) 2003. Australian demographic statistics. ABS Cat. No. 3101.0. Canberra: ABS.

AIHW (Australian Institute of Health and Welfare) 2001. Residential aged care in Australia 1999–00: a statistical overview. AIHW Cat. No. AGE 19. Canberra: AIHW.

AIHW 2002. Older Australia at a glance (3rd edition). AIHW Cat. No. AGE 25. Canberra: AIHW.

AIHW 2003a. Interface between hospital and residential aged care: feasibility study on linking hospital morbidity and residential aged care data. AIHW Cat. No. AGE 31. Canberra: AIHW.

AIHW 2003b. Residential aged care in Australia 2001–02: a statistical overview. AIHW Cat. No. AGE 29. Canberra: AIHW.

Gibson D 2002. The interface between hospital and residential aged care. Australian Health Review 25(5):132–5.