Chapter 3: Results – Record Linkage

Box 3: Overview of Chapter 3

Structure of this chapter

A summary of the output from the record linkage process is presented in Section 1, followed by estimates of the sensitivity, specificity and positive predictive value (PPV) of the record linkage (Section 2). A detailed analysis then follows in Section 3 of linkage bias – the difference in probability of a mortality record being linked to a census record by demographic and, most importantly, socioeconomic factors.

Output of record linkage

31,635 of 41,310 mortality records (76.6%) were linked to one of 3,373,896 census records.

Sensitivity, specificity and PPV of the record linkage

The PPV of the record linkage (the percentage of census–mortality links that were true links) was estimated to be approximately 97.5%. Using this overall estimate of the PPV, and the above percentages of mortality records linked, it was possible to estimate the sensitivity (the percentage of census respondents that actually died in the three-year period linked to a mortality record) and the specificity (the percentage of census respondents not actually dying in the three-year period not linked to a mortality record) at approximately 75.9% and 99.98% respectively.

Linkage bias

The percentage of mortality records linked to a census record was lowest for 20–24 year old decedents (49.0%) and highest for 65–69 year old decedents (81.0%). By ethnic group (as given by the NHI mortality file), 63.4%, 57.7%, and 78.6% of Maori, Pacific, and non-Maori non-Pacific, respectively, were linked. There was little difference by sex. 79.3% of deaths within six months of the census were linked compared to 72.8% of deaths 30–35 months after the census.

Controlling for demographic factors in a log-linear regression model among 0–74 year olds combined, decedents from the most deprived decile of small areas were 8% less likely to be linked than decedents from the least deprived decile. Among 25–74 year old males combined, decedents from the lowest occupational class were 6% less likely to be linked than decedents from the highest occupational class. While not satisfying statistical tests for heterogeneity by demographic strata, separate regression models within age (25–44 and 45–64 years) by sex group suggested a greater linkage bias by small area deprivation among males and among 25–44 year olds.

1 Output from the record linkage

1.1 Data flow of mortality and census records

Mortality records were requested from New Zealand Health Information Services (NZHIS), where: deaths occurred between 5 March 1991 and 5 March 1994 inclusive; the decedent would have been 0–74 years on 5 March 1991; and the decedent was a New Zealand resident according to NZHIS information. 42,229 mortality records were thus received.

All but 46 of the NZHIS mortality records were linked to a mortality record on the SNZ Vitals file. 17 NZHIS mortality records were linked to two SNZ Vitals file records. Despite non-New Zealand residents being excluded on the basis of NZHIS domicile codes, the SNZ Vitals file meshblock was coded as 'overseas usual residence' for a further 331 cases – they were excluded. One of the 331 overseas residents was for one of the 34 duplicate mortality records 'created' from the original 17 NZHIS mortality records. A decision was made to retain the 33 (34 minus 1) remaining duplicate mortality records, in case the true link could be established later. (This was not possible, and all 33 duplicate records were eventually discarded). Thus 41,915 mortality records were submitted to the record linkage (42,229 – 331 + 17).

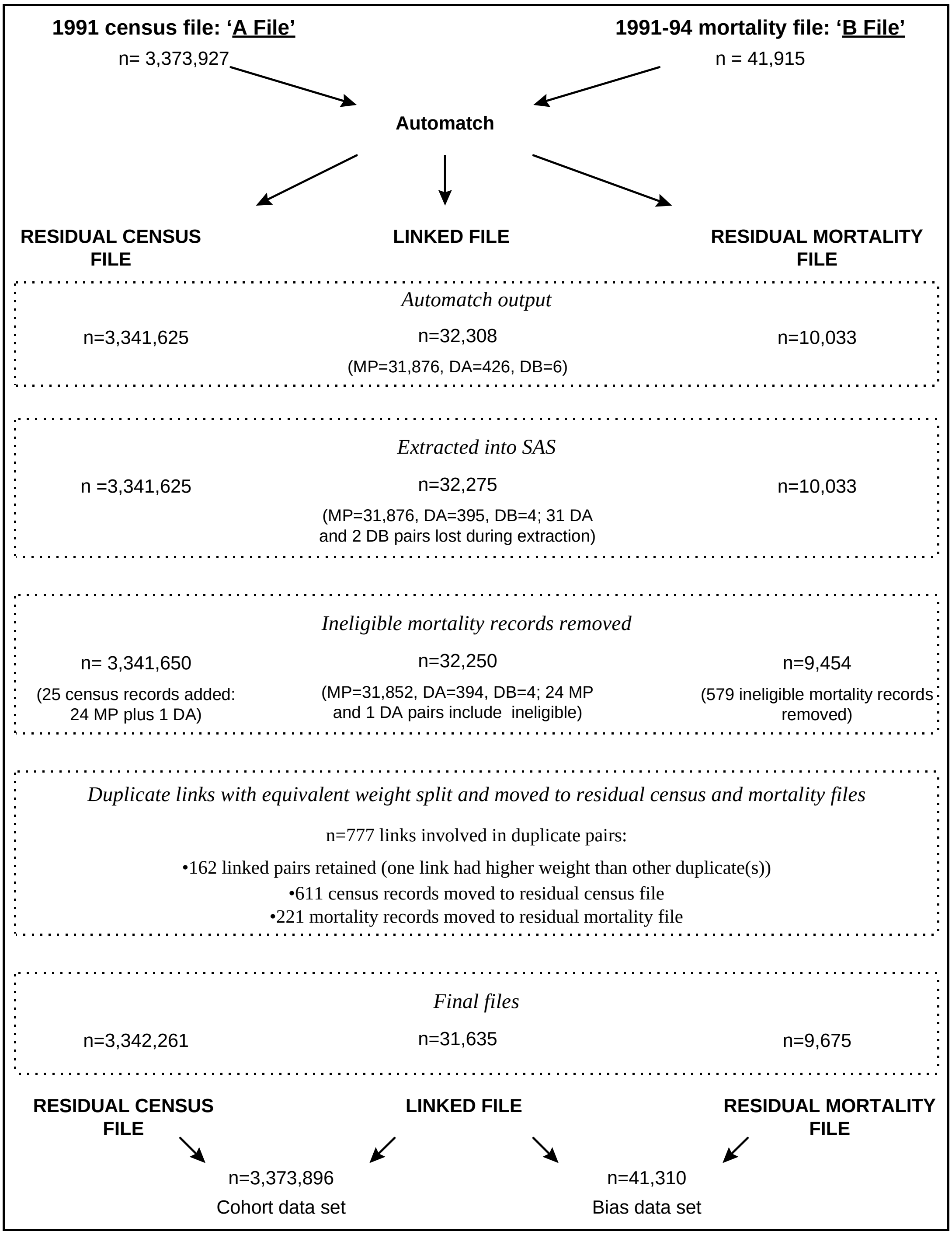
The record linkage involved submitting the mortality and census files to Automatch®, and deriving three output files:

- linked mortality and census records
- unlinked census records
- unlinked mortality records.

The flow of the mortality and census records is shown in Figure 9 below. **MP**, **DA**, and **DB pairs** are Automatch® terms. A MP pair is a pairing of one census and mortality record, and is the highest weight scoring pair for the given mortality record and the given census record (ie, the most likely true link). Occasionally, a mortality record is linked to two census records *above the cut-off*. In this case, one pair is termed the MP pair and the other the DA pair (duplicate A pair – census file was specified as the 'A file' in Automatch®). The MP pair is the highest weight scoring pair or, if both pairs have the same probabilistic weight score, they are randomly assigned as either a MP or DA pair. It is possible to have more than one DA pair associated with one MP pair (eg, three census records linked to one mortality record). DB pairs are the converse of DA pairs (ie, one census record linked to two or more mortality records). As the number of census records greatly outnumbered the mortality records in the NZCMS, DB pairs were much less common than DA pairs.

The total number of observations in the linked file of the Automatch® output (n=32,308) exceeds the number of either unique mortality records or unique census records due to the presence of DA pairs DB pairs. Subtracting the number of DA pairs (n=426) from the total number of links (n=32,308), and adding the number observations in the residual mortality file (n=10,033) gives 41,915 – the total number of mortality records submitted. A similar calculation can be done to give the total number of census records submitted, but instead subtracting the number of DB pairs (n=6).

Figure 9: Flow diagram of census and mortality records in the record linkage process



MP, DA, and DB pairs are Automatch® terminology:

MP = Match pair of one census and one mortality record

DA = Duplicate A pair, that is mortality–census pair of records where the mortality record is already involved in another pair (MP pair, or even MP pair and one or more DA pairs) with a higher or equivalent probabilistic weight score.

DB = Duplicate B pair, the reverse of a DA pair.

During the extraction of data from Automatch® to SAS, 31 DA pairs and 2 DB pairs were 'dropped'. The reason was not determined, and it was not detected until much of the processing of the links had been conducted in SAS. Given the large amount of time and resource that would have been required to re-run the final match-run strategy, and the lack of certainty that the same problem would not recur in any further extraction, these 33 observations were accepted as lost. The overall impact was minor, being 2 out of 41,915 submitted mortality records (0.005%) and 31 of 3,373,927 submitted census records (0.0009%).

The mortality data requested from NZHIS was for people aged 0–74 on census night. However, the data actually included people born up to a year after the census (n=532) – this was detected during the final match-run. Also included (knowingly) in the submitted mortality records were the 33 observations for the 17 NZHIS mortality records with two SNZ Vitals file links, and 38 decedents who actually died on 5 March 1991 (census day). Inspection of records suggested there would be little chance of successfully teasing apart the 17 duplicates. Further investigation also suggested that the likelihood of someone dying on census day having had a census completed by them (or on their behalf) was remote. Therefore, these 603 (532 + 33 + 38) 'ineligible' mortality records were removed from the data. Calculations (not presented here) suggested that inclusion of these 603 ineligible records had no effect on the probability of a true link being found for the remaining eligible mortality records. Therefore, there was no justification to repeat the final match-run of the record linkage.

Automatch® does not allow the DA or DB pair(s) *and* the associated MP pair to all be discarded together – instead one link has to be accepted as the correct MP pair. For DA and DB pairs with an equivalent weight to the associated MP pair, the DA and DB pair(s) *and* the MP pair were to be discarded. (An equivalent weight for two census records linked to one mortality record meant that there was no better than a 50:50 odds of selecting the true link, so both were discarded to maintain a high positive predictive value of the record linkage.) This discarding had to be conducted in SAS. 777 linked pairs were involved in a MP/DA or MP/DB association. 162 MP pairs had a higher match weight than the associated DA or DB pair(s), and were therefore retained as the 'best link'. The remaining 615 links were separated into 611 unique census records, and 221 unique mortality records.

The final size of the linked file was 31,635, and included links from all eight passes of the final match-run strategy. (The eight passes are summarised in the following section.) The sum of the linked file and residual census file records was 3,373,896, or 31 less than the original census file size due to the loss of 31 DA pairs during extraction from Automatch®. The sum of the linked file and residual mortality file records was 41,310, or two less than the number of eligible mortality records due to the loss of 2 DB pairs during extraction from Automatch®.

Forty-eight of the linked mortality and census records had a date of birth on the *census* data that made them older than 74 on census night. (As the record linkage was probabilistic, it was possible to form links when one variable (eg, year of birth) did not agree exactly.) These 48 links were retained in the linkage bias analyses as:

- the linkage bias analyses used mortality record data only
- the exclusion criteria would, obviously, not be applicable to those mortality records unlinked to a census record.

Age on census night according to census data was, however, used to determine subsequent inclusion in cohort analyses by age.

1.2 Record linkage strategy

The previous section presented a summary of the data flow in the record linkage. This section details the underlying record linkage strategy. The matching and blocking variables used in that strategy were described in Methods (Chapter 2). The determination of the best order of passes (each pass is a separate configuration of blocking and matching variables), cut-off weights, and the like is detailed in the Technical Report (Blakely et al 1999). In this section only a summary of the final record linkage strategy is presented.

The final match-run strategy, and number of links by pass, is presented in Table 15. The majority of the linked mortality records were identified on the first meshblock pass (25,311, or 61.27% of the total 41,312 eligible mortality records). For all eight passes, 76.6% of mortality records were linked to a census record. The first five passes were fully automated, whereas the last three passes required clerical review.

Table 15: Final match-run strategy

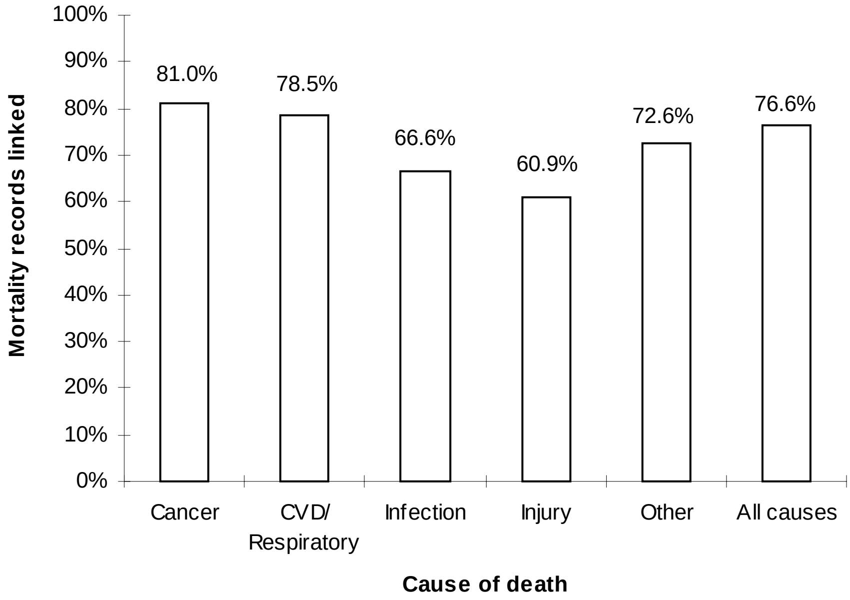
Pass number and blocking variable(s)	Links (% of eligible mortality records) †	
Meshblock	25,311	61.27%
Vitals-CAU, and month of birth	3473	8.41%
Post-CAU, and month of birth	1117	2.70%
Pre-CAU, and month of birth	340	0.82%
NHI-CAU, and month of birth	416	1.01%
Meshblock	429	1.04%
Meshblock	91	0.22%
Vitals-CAU and month of birth	458	1.11%
Total	**31,635**	**76.58%**

The source of the blocking variable for mortality records varies: meshblock = the meshblock from the SNZ Vitals file; Vitals-CAU = the CAU from the SNZ Vitals file; post-CAU = the CAU from the NMDS file for the health event (if any) immediately after the 1991 census; pre-CAU = the CAU from the NMDS file for the health event (if any) immediately before the 1991 census; NHI-CAU = the CAU from the NHI file.

† Links here are those remaining after full data cleaning as depicted in Figure 9.

The percentage of mortality records linked to a census record varied by broad grouping of death as shown in Figure 10.

Figure 10: Percentage of mortality records linked by cause of death



2 Estimates of the sensitivity, specificity, and PPV of the linkage overall

The estimated positive predictive value (PPV) and number of false positives for the first five passes are shown in Table 16. Two methods were used to estimate the positive predictive value: the chance method and the duplicate method. (The two methods are described briefly in Chapter 2 of this report, and in detail in the Technical Report (Blakely et al 1999).)

The overall PPV for the first five passes was estimated to be 97.8% by the chance method, and 98.1% by the duplicate method. The close agreement between the chance and duplicate method allows some confidence in the robustness and accuracy of both methods. It was not possible to estimate the PPV directly for the last three clerical review passes, but it was probably in the range of 80% to 90% based on work undertaken in the development of the clerical review rules. Assuming it was 85% for these three final passes, then the PPV for all eight passes combined was about 97.3% to 97.7%.

For practical purposes of comparison, the eight passes can be divided into three groups:

- very high PPV (greater than 99.5%; pass 1; 80.0% of all linked mortality records)
- high PPV (approximately 90%; pass 2–5; 16.9% of all linked mortality records)
- moderate PPV (80–90%; passes 6–8; 3.1% of all linked mortality records).

Table 16: Positive predictive value (PPV) and expected number of false positives (E[FP]) for passes 1 to 5 of the final match–run

Pass		Link pairs	Chance method		Duplicate method	
			E[FP]	PPV	E[FP]	PPV
1	Meshblock, weight > 30.0	23,000	22	99.9%	48	99.8%
	Meshblock, weight < 30.0	2311	– †	–	37	98.4%
2	Vitals-CAU	3473	365	89.5%	274	92.1%
3	Post-CAU	1117	130	88.4%	134	88.0%
4	pre-CAU	340	52	84.9%	39	88.5%
5	NHI-CAU	416	81	80.5%	41	90.1%
	Total	**30,657**	**687 ‡**	**97.8%**	**573**	**98.1%**

† The chance method can only be used when the majority of links above the weight cut-off are exact links. This did not apply to Pass 1 below a weight cut-off of 30.

‡ The chance method total includes the 37 estimated false positives by the duplicate method below the exact cut-off (30.0) for pass 1 to allow comparability.

The number of false negative links is approximated by the 9677 mortality records not linked to a census record (23.4% of all mortality records). However, this will not be the exact number of missed matches as:

- some decedents would not have been in New Zealand on 1991 census night
- some decedents would not have completed the census despite being in New Zealand on 1991 census night (perhaps 1–2% using the 1996 PES results – see page 20)
- 221 mortality records were linked to a census record, but were rejected as there was a duplicate link with the same weight. (If two census records were linked to one mortality record with the same weight score, then there was no better than a 50% chance of selecting the true link, so both were rejected.).

Taking the above into account, and the fact that 48 linked deaths were not among the 0–74 year old cohort (see page 191), it seemed reasonable to conclude that among the 0–74 year old cohort:

- about 2.5% (ie, 1 – PPV) of the linked census and mortality records were false links (ie, n = 2.5% × (31,635 – 48) = 790)
- about 97.5% (ie, PPV) of the linked census and mortality records were true links (ie, n = 97.5% × (31,635 – 48) = 30,797)
- about 2.5% (ie, a best guess) of the mortality records (n = 2.5% × 41,310 = 1033) might have been for decedents that had not completed a census record due to being absent from the country, simply failed to complete the census, or who were actually recorded with an age of greater than 74 years on the census.
- the residual mortality records (n = 8690, or 20.5% of the mortality records) actually had a true link somewhere in the census data set, but were either discarded due to being a duplicate (n = 221) or were simply missed by the record linkage. Further, there would be some mortality records with the age greater than 74 years on census night (and hence not obtained from NZHIS), but for whom their census-recorded age was less than 75 years – assume there were 100 such mortality records. Further, there would have been some census respondents who emigrated

and subsequently died overseas – assume there were 800 such deaths in the census cohort. Finally, all but approximately 2.5% (failure to complete census) of the mortality records involved in a false link (n = 97.5% × 790 = 770) would have actually had a true census link somewhere else on the census file. Thus, from the perspective of the census cohort, this means that about 10,360 (8690 + 100 + 800 + 770) census respondents might actually have died in the three-year follow-up, but were not detected (ie, false negative links).

Table 17: Best estimate of the two by two table of link / non-link status by vital status for the total 1991 census cohort

		True vital status at the end of follow-up		
		Died	Alive	
Output from record linkage	Linked	30,797	790	31,587
	Unlinked	10,360	3,178,223	3,188,583
		41,157	3,179,013	3,220,170

Taking these best estimates, the above two-by-two table of the likely true vital status of the 0–74 year old census cohort by their linkage status was constructed (Table 17). From this 'best estimate' table, the classification of the mortality outcome by the record linkage in the 1991 census cohort overall was estimated to have the following parameters:

- a sensitivity of 74.8% (30,797/41,157) – little different from the 76.6% figure from simply dividing the number of mortality records linked by the number of mortality records submitted
- a specificity of 99.975%
- and a PPV of 97.5% – as specified.

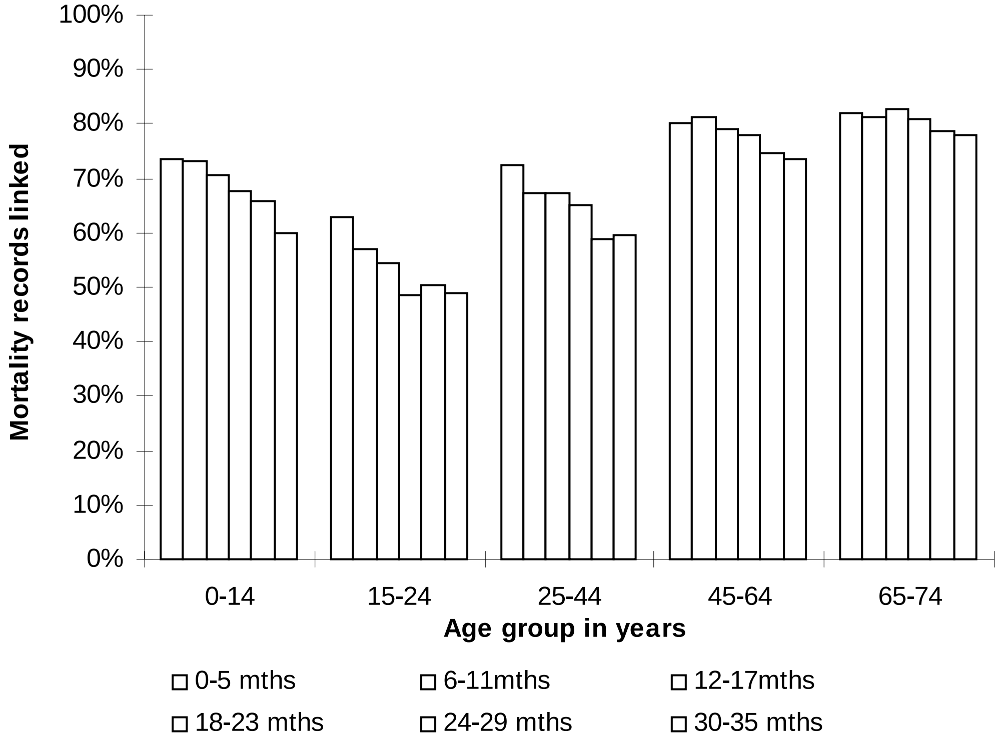
3 Linkage bias

The results of comparing linked with unlinked mortality records are presented for time elapsed following census, demographic variables, and finally the socioeconomic factors available on the mortality data (small area deprivation and occupational class). A detailed comparison of the mortality record linked to those not linked is provided in the Technical Report (Blakely et al 1999).

3.1 Time following census

The percentage of mortality records linked declined steadily with increasing time between census night and death: 79.3% of deaths within six months of the census were linked compared to 72.8% of deaths 30-35 months after the census. This decline was greatest for 15–24 year olds (63.0% to 49.0%), and least for 65–74 year olds (82.1% to 77.9%) – see Figure 11.

Figure 11: Percentage of mortality records linked to a census record by six-month period following the 1991 census by age group



3.2 Sex, age and ethnic group

The percentage of mortality records linked to a census record for each strata of sex by age by ethnic group is shown in Table 18.

Table 18: Percentage of 41,310 mortality records linked to a 1991 census record for deaths occurring during 1991–94, by sex, age, and ethnic group

Sex	% linked	Age group	% linked	Ethnic group	N †	% linked
Male	75.7%	0–14	68.7%	Maori	90	57.0%
				Pacific	36	71.4%
				Non-Maori non-Pacific	372	71.4%
		15–24	52.1%	Maori	198	49.8%
				Pacific	39	44.0%
				Non-Maori non-Pacific	1098	53.0%
		25–44	61.3%	Maori	423	54.1%
				Pacific	99	48.0%
				Non-Maori non-Pacific	2145	63.4%
		45–64	76.8%	Maori	1110	66.0%
				Pacific	264	63.7%
				Non-Maori non-Pacific	7788	78.7%
		65–74	81.3%	Maori	588	62.1%
				Pacific	159	54.7%
				Non-Maori non-Pacific	10809	82.7%
Female	77.9%	0–14	69.7%	Maori	69	60.6%
				Pacific	21	70.0%
				Non-Maori non-Pacific	246	72.1%
		15–24	58.8%	Maori	69	56.5%
				Pacific	18	58.8%
				Non-Maori non-Pacific	345	59.3%
		25–44	71.6%	Maori	267	65.3%
				Pacific	84	60.7%
				Non-Maori non-Pacific	1188	73.8%
		45–64	79.0%	Maori	927	70.2%
				Pacific	174	49.1%
				Non-Maori non-Pacific	4656	81.9%
		65–74	79.7%	Maori	462	61.0%
				Pacific	111	66.4%
				Non-Maori non-Pacific	7452	81.1%

† Number of submitted mortality records in each sex by age by ethnic group strata, random rounded to the nearest multiple of 3 as per SNZ protocol.

Overall, females had modestly increased linkage success (77.9%) compared to males (75.7%). The linkage success by five-year age group is shown in Figure 12. Together, sex and age interacted as predictors of linkage such that 15–44 year old male decedents were less likely to be linked than 15–44 year female decedents, whereas there was little difference by sex for other age groups (Table 18).

Figure 12: Percentage of mortality records linked by five-year age group

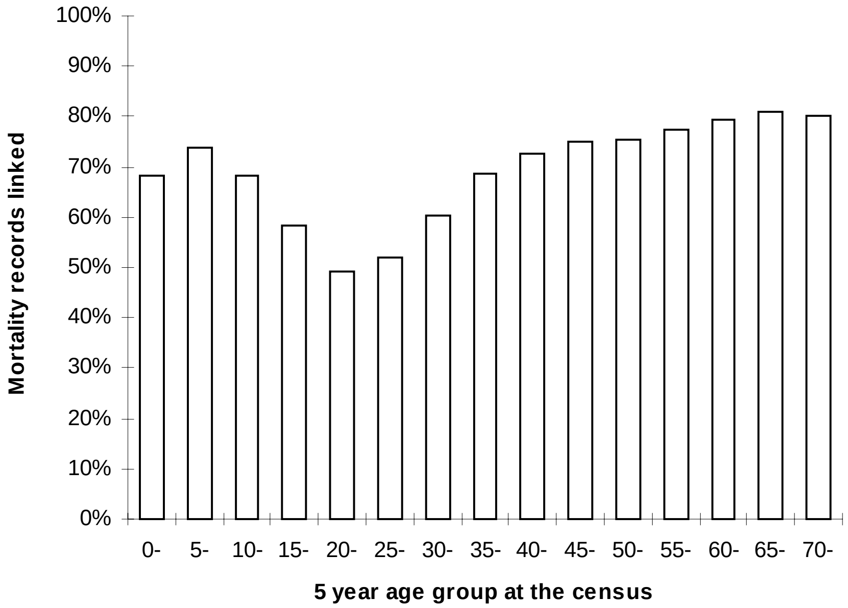
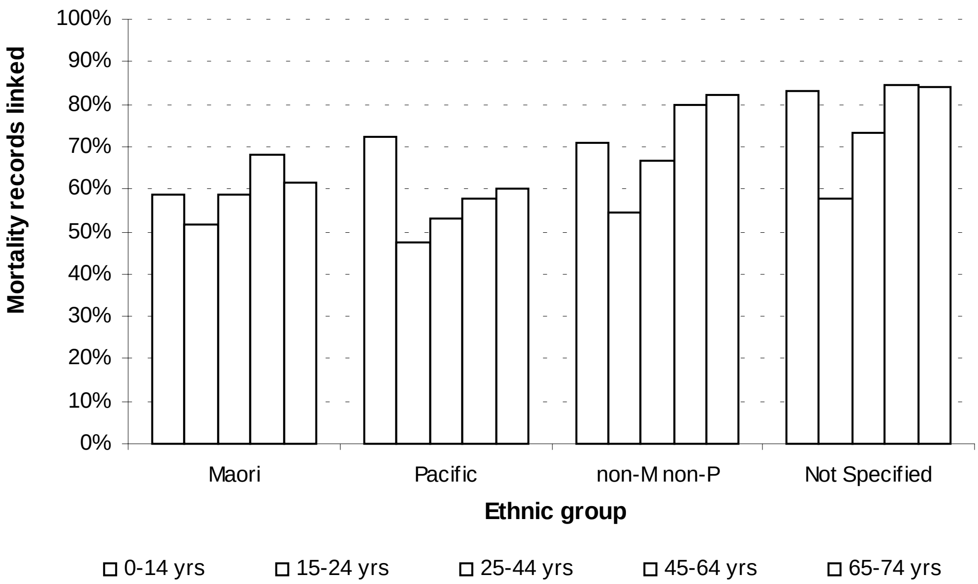


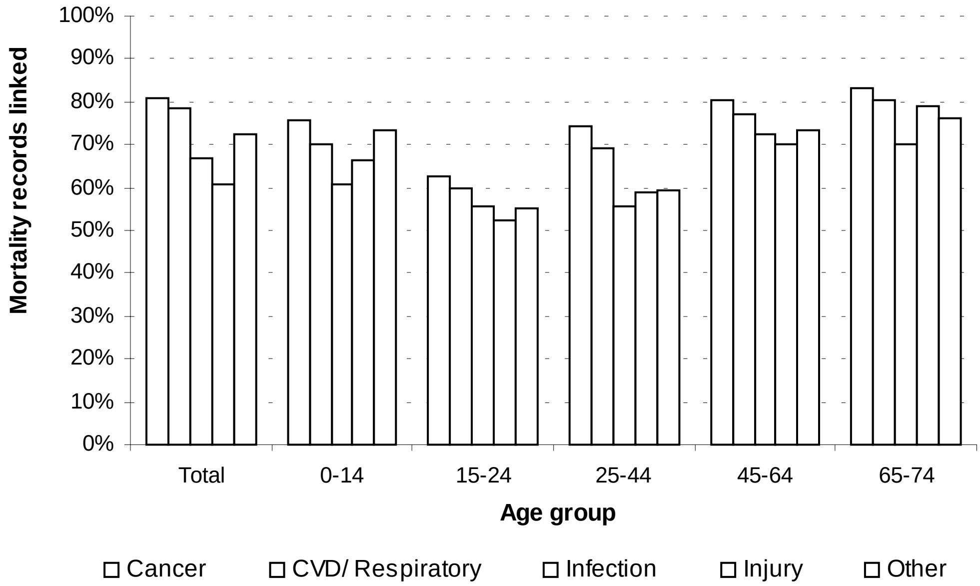
Figure 13: Percentage of mortality records linked to a census record by age group by ethnic group



The record linkage varied by ethnic group stated on the mortality data (NHI ethnicity): 63.4% of 4204 Maori, 57.7% of 1009 Pacific, 78.5% of 34,610 non-Maori, non-Pacific, and 81.9% of 1487 decedents with no specified ethnic group were linked. Age group and ethnic group also interacted as predictors of record linkage – the difference in linkage success for Maori and Pacific decedents compared to all other decedents was greatest among the older age groups (Figure 13).

The record linkage success varied by cause of death as shown in Figure 10. Figure 14 below shows that this variation in linkage rates by cause of death was partly due to age. For example, the discrepancy between the percentage of injury and other deaths linked is less within age groups than for the total population. Perhaps a third to half of the discrepancy in linkage rates by cause of death is explained by age group.

Figure 14: Percentage of mortality records linked to a census record by cause of death by age group



3.3 Linkage bias by socioeconomic position

3.3.1 Small area deprivation

36,927 (89.4%) of the mortality records could be assigned a NZDep91 score. (9.3% of mortality records had only a CAU on the SNZ Vitals file, and therefore could not be assigned a NZDep91 score. A further 1.3% of mortality records had a meshblock code with no assigned NZDep91 score (eg, remote communities). Table 19 shows the risk ratios for linkage by NZDep91 decile, compared to the least deprived decile of small areas, for a log-linear model including all mortality records and interaction terms of age, sex, and ethnicity. There was little difference between deciles 2–9 and decile 1 (the least deprived or highest socioeconomic position decile) for all cause mortality (risk ratios all 0.96 or above), but the risk ratio was 0.92 for decile 10. A similar pattern was evident for cancer, ischaemic heart disease, and unintentional injury (Table 19).

Table 19: Risk ratios (95% CI) by NZDep91 decile for the proportion of mortality records linked by cause of death, controlling for sex, age, and ethnic group †

NZDep91 decile	All causes (n=36,927; 0–74 years)		Cancer (n=12,389; 25–74 years)		IHD (n=8999; 25–74 years)		Unintentional injury (n=2241; 0–74 years)	
1 ‡	1.00		1.00		1.00		1.00	
2	0.98	(0.96–1.01)	1.00	(0.96–1.03)	0.95	(0.91–0.99)	1.00	(0.89–1.13)
3	0.98	(0.96–1.01)	1.00	(0.96–1.03)	0.99	(0.95–1.03)	0.85	(0.73–0.98)
4	0.98	(0.95–1.00)	0.98	(0.95–1.02)	0.97	(0.93–1.01)	0.91	(0.80–1.03)
5	0.98	(0.96–1.00)	0.97	(0.94–1.01)	0.96	(0.92–1.00)	0.92	(0.81–1.05)
6	0.97	(0.95–1.00)	1.00	(0.97–1.03)	0.93	(0.89–0.97)	0.98	(0.88–1.10)
7	0.96	(0.94–0.98)	0.98	(0.95–1.01)	0.95	(0.91–0.99)	0.90	(0.79–1.02)
8	0.96	(0.94–0.98)	0.98	(0.95–1.01)	0.93	(0.89–0.97)	0.90	(0.79–1.01)
9	0.96	(0.94–0.98)	0.97	(0.94–1.00)	0.94	(0.90–0.98)	0.86	(0.76–0.98)
10	0.92	(0.90–0.94)	0.94	(0.90–0.98)	0.90	(0.87–0.94)	0.86	(0.76–0.98)

† The risk ratios are from a log-linear model with the interaction products [age group * ethnic group] and [sex * age], in addition to the main effect [NZDep91].

‡ Reference category, and least deprived decile.

The above are summary estimates of the linkage bias by small area deprivation for 0–74 year olds combined. While there was no *statistically* significant evidence (p less than 0.05) for heterogeneity in the linkage bias by NZDep91 by demographic groups, for the purposes of direct use in sensitivity analyses in this report separate regression models were conducted within the four sex by age groups used in the cohort analyses, and with deaths in the first six months excluded. As shown in Table 20, the results by sex and age group are not greatly different from the combined result shown in the first column of Table 19, although there was some suggestion that the linkage bias by small area deprivation was greater among 25–44 year olds and greater among males. Note that, as with the combined analyses, the greatest drop-off in linkage success tended to be for decedents in the most deprived decile.

Table 20: Risk ratios (95% CI) by NZDep91 decile for the proportion of mortality records (all-cause deaths) linked for regression models conducted separately by age and sex, excluding deaths in first six months after census night †

NZDep91 decile	Males				Females			
	25–44 years (n=1926)		45–64 years (n=6846)		25–44 years (n=1125)		45–64 years (n=4371)	
1 ‡	1.00		1.00		1.00		1.00	
2	1.08	(0.93–1.25)	0.98	(0.92–1.03)	0.97	(0.84–1.12)	1.04	(0.98–1.11)
3	0.99	(0.84–1.17)	0.99	(0.94–1.05)	1.01	(0.87–1.17)	0.98	(0.92–1.05)
4	0.96	(0.81–1.12)	0.99	(0.94–1.04)	0.91	(0.77–1.07)	0.99	(0.92–1.05)
5	0.96	(0.82–1.13)	0.95	(0.90–1.01)	0.95	(0.81–1.10)	1.00	(0.94–1.07)
6	0.95	(0.81–1.10)	0.95	(0.90–1.01)	1.02	(0.89–1.18)	1.03	(0.97–1.10)
7	0.93	(0.80–1.09)	0.96	(0.91–1.01)	0.97	(0.84–1.13)	0.94	(0.88–1.00)
8	0.97	(0.83–1.12)	0.96	(0.91–1.01)	0.93	(0.80–1.07)	0.98	(0.92–1.05)
9	0.96	(0.83–1.11)	0.96	(0.91–1.01)	0.90	(0.78–1.05)	0.98	(0.93–1.05)
10	0.87	(0.74–1.01)	0.89	(0.84–0.94)	0.91	(0.79–1.05)	0.96	(0.90–1.02)

† The risk ratios are from a log-linear model with dummy variables for age in five-year groups, and dummy variables for Maori and Pacific.

‡ Reference category, and least deprived decile.

Finally, cause-specific associations of NZDep91 (and other socioeconomic factors) with mortality are reported in this report by sex, but for 25–64 year olds combined. For sensitivity analyses about these later cohort analyses, separate regression models of linkage bias were conducted by sex for four broad causes of death: cancer, cardiovascular disease, unintentional injury, and suicide. (Smaller aggregate levels of cause of death were not considered due to small numbers.) Results are shown in Table 21. There was a suggestion of greater linkage bias by NZDep91 among male injury and suicide deaths.

Table 21: Risk ratios (95% CI) by NZDep91 quintile for the proportion of mortality records (cause-specific deaths) linked among 25–64 year olds combined, by sex, and excluding deaths in first six months after census night

NZDep91 quintile	Cancer		CVD		Injury		Suicide	
Males	(n=2859)		(n=3402)		(n=657)		(n=480)	
1 ‡	1.00		1.00		1.00		1.00	
2	1.00	(0.94–1.06)	0.99	(0.94–1.05)	0.97	(0.79–1.18)	0.99	(0.81–1.21)
3	0.96	(0.91–1.02)	0.95	(0.90–1.01)	0.95	(0.78–1.14)	0.74	(0.58–0.94)
4	0.99	(0.94–1.05)	0.96	(0.91–1.02)	0.93	(0.77–1.11)	0.91	(0.75–1.11)
5	0.93	(0.87–0.99)	0.93	(0.88–0.98)	0.84	(0.69–1.02)	0.84	(0.68–1.04)
Females	(n=2685)		(n=1386)		(n=210)		(n=159)	
1 ‡	1.00		1.00		1.00		1.00	
2	0.99	(0.94–1.05)	0.99	(0.90–1.08)	0.90	(0.69–1.18)	1.02	(0.68–1.54)
3	0.99	(0.95–1.05)	1.00	(0.91–1.09)	0.97	(0.75–1.25)	0.98	(0.76–1.26)
4	0.97	(0.92–1.02)	0.92	(0.84–1.02)	0.95	(0.75–1.20)	0.86	(0.64–1.16)
5	0.99	(0.94–1.04)	0.94	(0.86–1.02)	0.93	(0.75–1.15)	1.12	(0.82–1.53)

The risk ratios are from a log-linear model with dummy variables for age in five-year groups, and dummy variables for Maori and Pacific – except suicide deaths where ethnicity was not included due to inability to fit the model.

‡ Reference category, and least deprived quintile.

3.3.2 Occupational class

Overall analyses by NZSEI occupational class were restricted to the 13,701 male decedents and 2,059 female decedents aged 25–74 years who had an occupation recorded on the death registration form and who died in the second and third year of follow-up (84.2% and 19.6% of 25–74 year old male and female decedents dying in the second and third year of follow-up, respectively). Decedents dying in the first year of follow-up had to be discarded as 1991 was a transition year between 1968- and 1990-base occupational codes. There was a 6% reduced chance of linkage for male decedents in occupational class 6 compared to occupational class 1 (Table 22). As with the results for NZDep91, there was only a substantial drop off in record linkage success for the lowest socioeconomic group, ie, occupational class 6. Results for females must be treated with caution due to small numbers and possible unrepresentativeness, but a similar decline in the probability of record linkage from high to low occupational class was apparent. As with the NZDep91 regression models, there was no evidence of a *statistically* significant interaction of NZSEI occupational class and demographic factors.

Table 22: Risk ratios (95% CI) by NZSEI occupational class for the proportion of mortality records linked by sex for 25–74 year olds, and excluding deaths in first 12 months after census night

NZSEI class	Males (n=13,701) ‡		Females (n=2059) ‡	
1	1.01	(0.97–1.04)	0.97	(0.87–1.08)
2	1.02	(0.99–1.05)	0.97	(0.91–1.03)
3	1.00	(0.97–1.02)	0.98	(0.92–1.05)
4†	1.00	–	1.00	–
5	1.00	(0.97–1.02)	0.98	(0.91–1.05)
6	0.94	(0.91–0.98)	0.91	(0.84–0.99)
Farmers	0.91	(0.88–0.94)	0.86	(0.74–0.99)

† Reference category.

‡ For both sexes, the risk ratios are from a log-linear model with just the interaction product [age group * ethnic group] and the main effect [NZSEI].

Farmers had a lower record linkage success than all six occupational classes. The reason for this lower linkage success was almost certainly that rural decedents were less likely to be assigned a meshblock, and hence their probability of linking with a census record was therefore less than for urban records.

Table 23: Risk ratios (95% CI) by NZSEI occupational class for the proportion of mortality records linked by age group for males, excluding deaths in first 12 months ‡

NZSEI class	25–44 years (n=1302)		45–64 years (n=5325)	
1	1.14	(0.99–1.32)	1.00	(0.95–1.06)
2	1.01	(0.86–1.18)	1.01	(0.96–1.07)
3	0.95	(0.83–1.08)	0.96	(0.92–1.00)
4†	1.00	–	1.00	–
5	1.01	(0.90–1.14)	0.99	(0.95–1.03)
6	0.86	(0.71–1.03)	0.94	(0.88–1.00)
Farmers	0.95	(0.81–1.12)	0.91	(0.85–0.96)

† Reference category

‡ The risk ratios are from a log-linear model with dummy variables for age in five-year groups and dummy variables for Maori and Pacific.

For direct use in sensitivity analyses of the impact of linkage bias on the observed occupational class mortality gradients, results from separate models for 25–44 and 45–64 year old males are shown in Table 23. Likewise, separate regression models by cause of death for 25–64 year old males combined are shown in Table 24. Note that the pattern of linkage bias by occupational class for male unintentional injury and suicide deaths is not consistent with that observed in Table 21 for small area deprivation, although the lack of consistency may be due to small numbers – the 95% confidence intervals mostly include 1.0 and occupational class was available for fewer deaths than was NZDep91.

Table 24: Risk ratios (95% CI) by NZSEI category for the proportion of mortality records (cause-specific deaths) linked for regression models, males aged 25–64 years combined, excluding deaths in first 12 months †

NZSEI class	Cancer		CVD		Injury		Suicide	
Males	(n=2280)		(n=2562)		(n=510)		(n=321)	
1	1.01	(0.94–1.10)	1.06	(0.98–1.15)	0.81	(0.60–1.08)	1.06	(0.72–1.55)
2	1.01	(0.94–1.09)	1.02	(0.94–1.10)	0.95	(0.74–1.21)	1.10	(0.83–1.46)
3	0.99	(0.92–1.05)	1.00	(0.94–1.07)	0.77	(0.60–0.99)	0.98	(0.74–1.30)
4 ‡	1.00		1.00		1.00		1.00	
5	1.02	(0.96–1.08)	1.02	(0.96–1.08)	0.96	(0.80–1.15)	1.02	(0.78–1.34)
6	0.90	(0.81–1.01)	0.96	(0.89–1.05)	0.80	(0.56–1.15)	1.27	(0.91–1.78)
Farmers	0.94	(0.87–1.02)	0.91	(0.83–0.99)	0.88	(0.70–1.10)	1.00	(0.74–1.34)

† The risk ratios are from a log-linear model with dummy variables for age in five-year groups, and dummy variables for Maori and Pacific.

‡ Reference category, and least deprived decile.

3.3.3 Small area deprivation and occupational class considered simultaneously

When both NZDep91 decile and NZSEI occupational class were included in the same log-linear model for males aged 25–74 (n=12,249), the risk ratios between high and low socioeconomic position for both NZDep91 decile and NZSEI occupational class changed little. (See Technical Report for results (Blakely et al 1999).)

3.3.4 Conclusion

Taken together, these results for NZDep91 and NZSEI suggest only a modest linkage bias by socioeconomic position, independent of demographic factors. In particular, this summary linkage bias was almost negligible for the majority of people (ie, NZDep91 deciles 1–9 and NZSEI occupational classes 1–5), and the proportion of mortality records that were linked declined significantly only for the lowest socioeconomic stratum (ie, NZDep91 decile 10 and NZSEI occupational class 6).

The results from separate regression models within sex and 25–44 and 45–64 year old groups, and by broad categories of cause of death, are used in sensitivity analyses of the effect of misclassification bias on the observed risk ratios in the cohort analyses, summarised in the next chapter.