Brief Report

Anonymous linkage of New Zealand mortality and Census data

Abstract

Background: The New Zealand Census-Mortality Study (NZCMS) aims to investigate socio-economic mortality gradients in New Zealand, by anonymously linking Census and mortality records.

Objectives: To describe the record linkage method, and to estimate the magnitude of bias in that linkage by demographic and socio-economic factors.

Methods: Anonymous 1991 Census records, and mortality records for decedents aged 0-74 years on Census night and dying in the three-year period 1991-94, were probabilistically linked using Automatch®. Bias in the record linkage was determined by comparing the demographic and socio-economic profile of linked mortality records to unlinked mortality records.

Results: 31,635 of 41,310 (76.6%) mortality records were linked to one of 3,373,896 Census records. 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, 63.4%, 57.7%, and 78.6% of Maori, Pacific, and decedents of other ethnic groups, respectively, were linked. Controlling for demographic factors, decedents from the most deprived decile of small areas were 8% less likely to be linked than decedents from the least deprived decile, and male decedents from the lowest occupational class were 6% less likely to be linked than decedents from the highest occupational class.

Conclusion: The proportion and accuracy of mortality records linked was satisfactorily high. Future estimates of the relative risk of mortality by socio-economic status will be modestly under-estimated by 5-10%.

(*Aust N Z J Public Health* 1999; 24: 92-5)

Tony Blakely, Alistair Woodward and Clare Salmond

Department of Public Health, Wellington School of Medicine, University of Otago, Wellington, New Zealand

With support and permission from Statistics New Zealand, Wellington, New Zealand (see Acknowledgments)

Socio-economic factors are major determinants of mortality.[1,2] In New Zealand the major body of work on mortality by socio-economic factors is that by Pearce and colleagues.[3,4] These studies have been limited to occupational class for males aged 15-64 years. Census and mortality records have been linked to create longitudinal studies in other countries, including Britain[5] and the United States.[6] This approach has a number of advantages:

1. The range of exposures can be increased to include variables such as income, education and labour force status, and in New Zealand smoking data for some Census years.
2. Both sexes and all age groups can be included in the analyses.
3. A linked study overcomes possible artefact introduced by measuring socio-economic factors and ethnic group from different data sources for the numerator and denominator.
4. Large studies can be conducted at relatively small marginal cost by linking existent datasets.

Statistics New Zealand (SNZ; the Government Department vested with conducting the five yearly Census) has recently sought to increase the benefits that can be obtained from its data by providing micro-data access for approved researchers.

This micro-data access is provided at the discretion of the Government Statistician to allow authoritative research of benefit to the public of New Zealand.

The New Zealand Census-Mortality Study (NZCMS) is the first time the NZ Census has been linked to an administrative dataset for purposes apart from improving the quality of SNZ surveys.

The NZCMS is a short duration cohort study of the 1991 Census night population age 0-74 years, followed up for mortality for three years. The priority goals of the NZCMS are to:

- investigate socio-economic mortality gradients in New Zealand by a wide range of socio-economic factors (labour force status, education, income, occupational class, and asset ownership), for different age groups (0-74 years), and both sexes; and.
- investigate possible variation in the relationship of personal socio-economic factors with mortality by community variables such as deprivation, income inequality, region and social capital.

A longer-term goal is to monitor changes over time in socio-economic mortality gradients, by linking other censuses to mortality data.

The objective of this paper is to describe the methods used in the record linkage, and to present the results of the record linkage in terms of the proportion of mortality records linked, likely accuracy, and bias.

Submitted: July 1999
Revision requested: October 1999
Accepted: November 1999

Correspondence to:
Dr Tony Blakely, Department of Public Health, Wellington School of Medicine, University of Otago, PO Box 7343, Wellington, New Zealand. Fax: +64 4 389 5319; e-mail: tblakely@wnmeds.ac.nz

Methods

Privacy

The record linkage was conducted with anonymous data only – no names or text addresses are available on Census data. Staff of SNZ, with strategic advice from TB, conducted the record linkage. Access to the data used in this study was provided by Statistics New Zealand in a secure environment designed to give effect to the confidentiality provisions of the Statistics Act (1975).

Data

Mortality data was obtained from New Zealand Health Information Services (NZHIS) for deaths that occurred between 5 March 1991 (Census night) and 5 March 1994, for decedents aged 0-74 on Census night. Census data was extracted from the 1991 Census master file.

Record linkage

The record linkage was conducted using a commercially available software package, Automatch.[7] Probabilistic record linkage involves assigning agreement and disagreement weights (or odds) for each value of each matching variable.[7-10] The matching variables common to both mortality and Census data were geocodes (meshblock and Census area unit), sex, ethnic group, country of birth, and date of birth (disaggregated to day of birth, month of birth, and year of birth). As the number of matching variables was relatively few in this project, the available information was increased by utilising multiple recordings of the same variable on the mortality data. First, five different geo-codes were available for mortality data: one meshblock code derived from the address on the death registration form (median population size 96), and four Census area codes taken from various NZHIS files for the same decedent (median population size 2,000). Second, date of birth and ethnic group were taken from both the National Health Index file and the death registration form – these two sources of data are independently collected and entered, the former by hospital clerks and the latter by funeral directors and SNZ.

Geocodes were used to 'block' the mortality and Census files – comparisons of records only occurred when the records first agreed on these geocodes. Meshblock, as the most discriminatory geo-code separating the New Zealand population into 30,000 blocks, was used as the blocking variable in the first 'pass' of the record linkage. Subsequent passes used other geo-codes as blocking variables. Using basic probabilities and the observed number of duplicate links (one mortality record linked with two or more Census records), methods were developed to estimate the number of false-links for any given pass (unpublished observations). The final record linkage strategy was a balance of maximising the number of links obtained, but minimising the estimated percentage of false-links. For duplicate links, the highest scoring duplicate link was accepted, or if the scores were tied both (or all) links were discarded.

Bias in the record linkage

Using only mortality data, the proportion of mortality records linked by sex, age, ethnic group, elapsed time in six-month periods between Census and death, occupational class and small area deprivation were determined. The New Zealand Socio-Economic Index (NZSEI) was used to assign an occupational class to mortality records.[11,12] The 1991 New Zealand Index of Deprivation (NZDep91) was used to assign small areas (usually one or two meshblocks) to deciles of deprivation, one being the least deprived and 10 the most deprived decile of small areas.[13,14] The variation in linkage success by socio-economic status (occupational class and NZDep91) was determined by a generalised linear model with a log link and binomial errors (referred to as log-linear hereafter), conducted in SAS version 6.12.[15] Wald Type III test statistics were used to guide backward selection of interaction products. A fuller account of the methods used in the record linkage, and results, can be found in a report published at http://www.wnmeds.ac.nz/academic/dph

Results

Some 41,310 eligible mortality records and 3,373,896 eligible Census records were processed by Automatch. From this sample, 31,635 mortality and Census records were anonymously linked – 76.6% of the eligible mortality records. The percentage of accepted links that were estimated to be true links was high (in excess of 95%).

The percentage of mortality records linked declined steadily with increasing elapsed time between the Census and death: 79.3% of deaths within six months of the Census were linked compared with 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%).

Table 1 shows the percentage of mortality records linked to a Census record for each strata of sex by age by ethnic group. Overall, females had modestly increased linkage success (77.9%) compared with males (75.7%). 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. The record linkage varied by ethnic group stated on the mortality data: 63.4% of 4,204 Maori, 57.7% of 1,009 Pacific, 78.5% of 34,610 non-Maori, non-Pacific, and 81.9% of 1,487 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 with all other decedents was greatest among the older age groups (see Table 1).

To account for these interactions of the three fundamental demographic factors (sex, age, and ethnic group), log-linear regression analyses with dummy variables for sex cross-classified with age group (0-14, 15-24, 25-44, 45-64, 65-74), and age group cross-classified with ethnic group (Maori, Pacific people, 'rest') were used to estimate the independent effect of socio-economic status on the record linkage. 36,927 (89.4%) of the mortality records could be assigned a NZDep91 score. Table 2 shows the risk ratios

Table 1: Percentage of 41,310 mortality records for deaths occurring during 1991-94 anonymously linked to a 1991 Census record by sex, age and ethnic group.

Sex	% linked	Age group	% linked	Ethnic group	n[a]	% linked
		0-14	68.7	Maori	90	57.0
				Pacific	36	71.4
				Rest	372	71.4
		15-24	52.1	Maori	198	49.8
				Pacific	39	44.0
				Rest	1,098	53.0
Males	75.7	25-44	61.3	Maori	423	54.1
				Pacific	99	48.0
				Rest	2,145	63.4
		45-64	76.8	Maori	1,110	66.0
				Pacific	264	63.7
				Rest	7,788	78.7
		65-74	81.3	Maori	588	62.1
				Pacific	159	54.7
				Rest	10,809	82.7
		0-14	69.7	Maori	69	60.6
				Pacific	21	70.0
				Rest	246	72.1
		15-24	58.8	Maori	69	56.5
				Pacific	18	58.8
				Rest	345	59.3
Females	77.9	25-44	71.6	Maori	267	65.3
				Pacific	84	60.7
				Rest	1,188	73.8
		45-64	79.0	Maori	927	70.2
				Pacific	174	49.1
				Rest	4,656	81.9
		65-74	79.7	Maori	462	61.0
				Pacific	111	66.4
				Rest	7,452	81.1

Note:
(a) Number of mortality records in each sex by age by ethnic group strata, random rounded to the nearest multiple of three as per SNZ protocol.

for linkage by NZDep91 decile, compared with the least deprived decile of small areas. There was little difference between deciles 2-9 and decile 1 (the least deprived or highest socio-economic status 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 2). Regression models indicated that the bias in the record linkage for all-cause mortality by NZDep91 decile did not vary by demographic factors (sex, age, and ethnicity), and elapsed time between Census and death.

Analyses by NZSEI occupational class were restricted to the 13,701 male decedents and 2,059 females 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, resulting in substantial inaccuracy.) There was a 6% reduced chance of linkage for male decedents in occupational class 6 compared with occupational class 1. Results for females must be treated with caution due to small numbers and possible unrepresentativeness, but a similar decline in the chance of record linkage from high to low occupational class was apparent. 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 socio-economic status for both NZDep91 decile and NZSEI occupational class changed little.

Discussion

This project is the first time that anonymous New Zealand Census records have been linked to anonymous records from an

Table 2: Risk ratios by NZDep91 decile for the proportion of mortality records linked by cause of death, controlling for sex, age and ethnic group.[a]

	All causes (n=36,927; 0-74 yrs)		Cancer (n=12,389; 25-74 yrs)		IHD (n=8,999; 25-74 yrs)		Unintentional injury (n=2,241; 0-74 yrs)	
NZDep91 decile	Risk ratio	95% CI	Risk ratio	95% CI	Risk ratio	95% CI	Risk ratio	95% CI
1[b]	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

Notes:
*(a) 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].*
(b) Reference category, and least deprived decile.

external dataset, and as such is an important precedent. The linked dataset provides the opportunity to carry out a cohort study following the entire 0-74-year-old New Zealand population, thus exceeding both the UK Office of Population and Census and Surveys (OPCS) Longitudinal Study[5] and US National Longitudinal Mortality Study[16] in size. The major weakness of the linked dataset is that the record linkage can only viably be conducted for a few years after the Census as it is heavily dependent on usual residence at the time of the Census being the same as that at the time of death. Subsequent cohort analyses may be affected by health selection effects (follow-up for three years only) and follow-up bias (the mortality records not linked back to a Census record may systematically differ from those linked by socio-economic status). Health selection effects will be investigated in subsequent publications by analysing changes over time in the observed socio-economic mortality gradients. Quantifying the possible follow-up bias, however, is the major objective of this paper.

First, 76.6% of mortality records were anonymously linked to a Census record – a satisfactorily high follow-up rate. However, for young people, Maori and Pacific people, the percentage was less than 70%.

Second, the biases in the record linkage by the key variables of interest – socio-economic factors – were both *small* and *uniform* across all strata of demographic factors and elapsed time between Census and death. For either small area deprivation or occupational class, the proportion of mortality records linked for the lowest socio-economic strata was no more than 8% less than that for the highest socio-economic strata. The bias by socio-economic status in the record linkage was most marked for the lowest socio-economic status groups – for the majority of socio-economic strata differences in the proportion of decedents linked were smaller again. These analyses of bias in the record linkage suggest that subsequent cohort analyses will tend to under-estimate the true socio-economic mortality gradients by about 5-10%. While mortality risks between demographic strata are not directly comparable due to varying success in record linkage, unbiased comparisons can be made between demographic groups in the *relative* risks of mortality by socio-economic factors.

Acknowledgments

Staff of Statistics New Zealand, including: Paul Willoughby, Keith McLeod, Sandra McDonald, Tracey Gilmour, Sharleen Forbes, John Cornish, and the Government Statistician Len Cook.

Co-investigators on the New Zealand Census Cohort Mortality Study: Cindy Kiro, Peter Davis, and Neil Pearce.

The New Zealand Census-Mortality Study is funded by the New Zealand Health Research Council.

Tony Blakely is funded by a New Zealand Health Research Council Training Fellowship.

References

1. Black D, Morris J, Smith C, Townsend P, 1980. The Black Report. In: Townsend P, Davidson N, Whitehead M, editors. *Inequalities in health*. Hammondsworth: Penguin, 1992.
2. Sorlie P, Backlund E, Keller J. US mortality by economic, demographic, and social characteristics: the national longitudinal mortality study. *Am J Public Health* 1995; 85: 949-56.
3. Pearce N, Marshall S, Borman B. Undiminished social class mortality differences in New Zealand men. *N Z Med J* 1991; 104: 153-6.
4. Pearce N, Pomare E, Marshall S, Borman B. Mortality and social class in Maori and non-Maori New Zealand men: changes between 1975-7 and 1985-7. *N Z Med J* 1993; 106: 193-6.
5. Fox A, Goldblatt P. *Longitudinal study 1971-1975: Socio-demographic mortality differences*. London: Her Majesty's Stationary Office: Office of Population Censuses and Surveys, 1982.
6. Rogot E, Sorlie P, Johnson N. Probabilistic methods in matching Census samples to the national death index. *J Chron Dis* 1986; 39: 719-34.
7. *Automatch* [generalised record linkage system]. Version 4.2: User's Manual. Kennebunk(Maine): MatchWare Technologies, 1998.
8. Newcombe H. *Handbook of record linkage: methods for health and statistical studies, administration and business*. Oxford: Oxford University Press, 1988.
9. Baldwin J, Acheson E, Graham W. *Textbook of medical record linkage*. Oxford: Oxford University Press, 1987.
10. Jaro M. Probabilistic linkage of large public health data files. *Stat Med* 1995; 14: 491-8.
11. Davis P, McLeod K, Ransom M, Ongley P. *The New Zealand Socioeconomic Index of Occupational Status (NZSEI)*. Wellington: Statistics New Zealand, 1997.
12. Davis P, McLeod K, Ransom M, et al. The New Zealand Socioeconomic Index: developing and validating an occupationally-derived indicator of socio-economic status. *Aust N Z J Public Health* 1999; 23: 27-33.
13. Crampton P, Salmond C, Sutton F. NZDep91: A new index of deprivation. *Soc Policy J N Z* 1997; 9: 186-93.
14. Salmond C, Crampton P, Sutton F. NZDep91: A New Zealand index of deprivation. *Aust N Z J Public Health* 1998; 22: 835-7.
15. *SAS System* [statistical program]. Version 6.12 .Cary(NC): SAS Institute, 1996.
16. Rogot E, Sorlie P, Johnson N, Schmitt C. *A mortality study of 1.3 million persons by demographic, social, and economic factors: 1979-1985 follow-up*. Bethesda: National Institutes of Health (publication No. 92-3297), 1992.

Appendix: Statistics New Zealand Summary Security Statement

The full security statement is in a technical report to be published by the Wellington School of Medicine in hardcopy and at http://www.wnmeds.ac.nz/academic/dph. The New Zealand Census-Mortality Study (NZCMS) is a study of the relationship between socio-economic factors and mortality in New Zealand, based on the integration of anonymous population Census data from Statistics New Zealand and mortality data from the New Zealand Health Information Service. The project was approved by Statistics New Zealand as a Data Laboratory project under the Microdata Access Protocols in 1997. The datasets created by the integration process are covered by the Statistics Act and can be used for statistical purposes only. Only approved researchers who have signed Statistics New Zealand's declaration of secrecy can access the integrated data in the Data Laboratory. For further information about confidentiality matters in regard to this study please contact Statistics New Zealand.