The New Zealand Census–Mortality Study

Socioeconomic inequalities and adult mortality 1991–94

Published in June 2002
by the Ministry of Health
PO Box 5013, Wellington, New Zealand

ISBN 0-478-27048-8 (Book)
ISBN 0-478-27049-6 (Internet)

Blakely T. *The New Zealand Census–Mortality Study: Socioeconomic inequalities and adult mortality 1991–94*. Wellington: Ministry of Health, 2002.

This document is available on the Ministry of Health's website: http://www.moh.govt.nz/phi
and the NZCMS website: http:/www.wnmeds.ac.nz/nzcms-info.html

Foreword

The New Zealand Census - Mortality Study (NZCMS) is the principal instrument by which the Ministry of Health monitors social inequalities in health. This report provides the technical background to the study, as well as illustrative substantive results.

The NZCMS is hosted by the Wellington School of Medicine and Health Sciences, University of Otago, and is led by Dr Tony Blakely. The study has been funded by the Health Research Council since 1998, and co-funded by the Ministry of Health since 2001.

The study involves anonymous and probabilistic linkage of census and mortality records, thereby creating cohort (or follow-up) studies of the entire New Zealand population for the three years following each census.

The NZCMS is conducted in conjunction with Statistics New Zealand. The linking of the two datasets was undertaken by Statistics New Zealand staff, and access to the linked data was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975.

As a census-based cohort study, the NZCMS offers several advantages over other study designs:

- direct linkage of exposures to outcomes at the individual level
- rich exposure data
- great statistical power (because of the large 'sample' size).

The New Zealand Census, conducted by Statistics New Zealand, collects data on numerous social factors, allowing the NZCMS to 'map' mortality by (among other variables) ethnic group, education, occupation, income, asset ownership, family type, labour force status, region and small area deprivation.

The NZCMS, therefore, has great potential to both further our understanding of the determinants of health in New Zealand and to assist agencies such as the Ministry of Health to plan healthy public policy, public health services and health care services.

The purpose of this report is to provide users of NZCMS information with access to the technical infrastructure of the study, as well as illustrative substantive content: specifically, results for adults for the period 1991–94 (ie, following the 1991 Census).

The specific objectives of the study for the next five years may be summarised as follows:

Objectives of the NZCMS for the period 2001 to 2005

1 To extend the NZCMS database by **linking** mortality records to each of the 1981 and 1996 censuses (linkage to the 1991 and 1986 censuses has already been completed).

2 For each of the four censuses (1981, 1986, 1991, 1996), set-up and maintain:

   a the **cohort** data set (ie, including full census data set and linkage status, for analyses of socioeconomic and ethnic differences in mortality)

   b the **bias** data set (ie, including only mortality records and whether they were linked, to allow quantification of linkage bias)

   c the **unlock** data set (ie, including only linked mortality records, with both their mortality and census ethnicity codes, to allow quantification of numerator–denominator bias).

3 Document in a series of **technical reports**:

   a the linkage process, data flow and linkage bias for each census cohort

   b the unlocking of the numerator–denominator bias for ethnicity between census and mortality data for each census cohort, including the development of adjustment ratios to correct this bias

   c the weighting procedure to overcome linkage bias for each census cohort.

4 Using the cohort data sets, analyse, monitor and report:

   a socioeconomic mortality gradients for a **range of individual-level socioeconomic factors**, including income, education, car access, socioeconomic deprivation and labour force status

   b socioeconomic mortality gradients by **cause of death** (including '**avoidable**' mortality)

   c **trends** in socioeconomic mortality gradients over the time period captured by the census cohorts

   d socioeconomic mortality gradients for all **lifecycle** stages (including children and youth, as well as adults and older people up to 74 years of age), by **sex**

   e socioeconomic mortality gradients for the three major **ethnic** groups (Maori, Pacific, non-Maori, non-Pacific) and the contribution of socioeconomic factors to ethnic inequalities

   f **multilevel analyses** of contextual/ecological effects of variables such as small area deprivation, social capital and income inequality

   g **regional variation** in socioeconomic mortality gradients (including possible application of appropriate mapping techniques).

5 Using the cohort data sets, determine:

   a **absolute mortality risks** by socioeconomic and demographic strata

   b **lifetable** statistics

   c measures of **premature mortality** (such as years of life lost).

6 To conduct **cross-national comparisons** of socioeconomic mortality gradients in New Zealand with those described for other developed countries (to the extent possible), both at given points in time and for trends over time.

7 Using the 1981 and 1996 census cohorts, measure the contribution of **tobacco smoking** to socioeconomic mortality gradients.

8 To continue and extend the **unlocking** of the bias in ethnicity by:

   a demonstrating the **extent of bias** in routinely published mortality rates by ethnicity caused by numerator–denominator bias

   b applying these adjustment ratios to regenerate and report **corrected time series** for ethnic-specific mortality rates.

9 To pursue more complex analyses (both for a given census and over time) of the **independent, joint and interactive effects** of (several) socioeconomic factors and demographic factors on mortality, using simple (eg, multiple-dimensional stratification) and complex (eg, non-parametric, multivariate and structural equation modelling) methods.

10 To **collaborate** with other research groups on specific projects.

11 To investigate extending the NZCMS to include the **linkage of morbidity data** (such as hospitalisation and cancer registration data) and census data.

We are very fortunate indeed to have this ongoing cohort study available, providing us with timely and regularly updated information about social inequalities in health of a scope and quality that few, if any other, countries can match, and at a very small marginal cost.

More detailed technical reports, including a copy of Dr Blakely's PhD thesis *Socioeconomic factors and mortality among 25–64 year olds: The New Zealand Census–Mortality Study* (on which this report is based), can be found at the NZCMS website:

http://www.wnmeds.ac.nz/nzcms-info.html

For further information on the NZCMS or its application to policy, please contact either:

Tony Blakely at tblakely@wnmeds.ac.nz

or Martin Tobias at martin_tobias@moh.govt.nz.

Martin Tobias
Public Health Physician
Public Health Intelligence
Public Health Directorate

Preface and Acknowledgements

There are numerous people who must be acknowledged for their support and guidance in the development of the NZCMS.

Len Cook, Government Statistician at SNZ until 2000, agreed in principle to the anonymous and probabilistic linkage of census and mortality records. Without his agreement and vision the NZCMS would never have begun.

I have been ably supported and encouraged by my co-investigators: Professor Alistair Woodward, Mrs Clare Salmond, Professor Peter Davis, Dr Cindy Kiro, and Professor Neil Pearce. In particular, I am indebted to Alistair and Neil for their supervision of my PhD thesis on which this report is based. I am also particularly indebted to Alistair and Clare for their representations to Statistics New Zealand on my behalf, particularly when I was overseas for a period.

Within Statistics New Zealand, John Cornish and Dr Sharleen Forbes have provided managerial oversight of the NZCMS. Sandra McDonald has managed and overseen all access to the linked census cohort data in the Data Laboratory at SNZ. Paul Willoughby and Keith McLeod conducted the anonymous and probabilistic linkage of the 1991 Census and 1991–94 mortality records. During 2000 Jonathan Briggs and Victoria Wilcox took over the day-to-day implementation and oversight of the NZCMS. Numerous other staff of SNZ have contributed to the development and implementation of the NZCMS, including Tracey Gilmour, Robert Didham, Robert Templeton, Chris Zingel, Frances Krsinich, Brenda Colville and Robyn Bishop.

From the New Zealand Health Information Service, Barbara Bridger, Tracey Vandenberg, Jim Fraser, and Liz Mooney have all been of great assistance.

Beyond the co-investigators directly involved in the NZCMS, many other academic colleagues have provided advice, encouragement, or committed themselves to future collaboration on the NZCMS, including Philippa Howden-Chapman, Peter Crampton, Des O'Dea, Sunny Collings, Bridget Robson, Paparangi Reid and Andrew Sporle. I am also appreciative of the support from colleagues working in policy institutions who have patiently waited for results from the NZCMS, in particular Martin Tobias at the Ministry of Health.

In late 2000, June Atkinson has joined the NZCMS as the Data Manager / Bio Statistician, and has assisted with the extraction of some of the final tables in this report.

Tony Blakely
New Zealand Census Mortality Study
Department of Public Health
Wellington Medical School

Abbreviations

CAU	census area unit
dd	day of birth
DOB	date of birth
LS	Longitudinal Survey (the OPCS LS unless stated otherwise)
mm	month of birth
NHI	National Health Index
NLMS	(US) National Longitudinal Mortality Study
NMDS	National Minimum Dataset
NPV	negative predictive value
NZCMS	New Zealand Census–Mortality Study
NZDep91	small area deprivation index, based on 1991 census data (Salmond et al 1998)
NZSEI	New Zealand Socioeconomic Index (used to derive occupational classes)
OPCS	(UK) Office of Population Censuses and Surveys
PES	Post Enumeration Survey (conducted for the first time following the 1996 census in New Zealand to determine the census undercount)
PPV	positive predictive value
RII	relative index of inequality (Mackenbach and Kunst 1997)
RTC	road traffic crash
yy	year of birth

Contents

List of Tables

List of Figures

List of Boxes

Executive Summary

Objective

To measure the association of deprivation, education, occupational class, housing tenure, car access and income with mortality among 25–64 year olds during 1991–94 in New Zealand.

Methods

Anonymously and probabilistically linking census and mortality records created a cohort study of the New Zealand population in 1991 followed up for mortality for three years. Gradients of mortality by socioeconomic factors were determined using logistic regression.

Results

Seventy-six point six percent of eligible mortality records were linked to a census record. There was a modest linkage bias in the record linkage, whereby lower socioeconomic decedents were five to ten percent less likely to be linked to a census record than higher socioeconomic decedents.

Controlling for age and ethnicity, strong associations of each socioeconomic factor with mortality were observed, with people from lower socioeconomic groups having approximately twice the mortality risk of their higher socioeconomic counterparts. For example, among 45–64 year olds the odds ratios of all-cause mortality for those living in a low-income household (equivalised household income less than $10,000) compared to those living in a high-income household (greater than or equal to $70,000) were 2.05 for males and 1.58 for females. Similarly, the odds ratios among 25–44 year olds with no educational qualification compared to those with tertiary qualifications were 2.13 for males and 2.00 for females. All-cause mortality gradients by education tended to be stronger among 25–44 year olds, and by income tended to be stronger among 45–64 year olds. Strong mortality gradients were evident for all specific causes of death, other than non-lung cancers.

Health selection bias – reverse causation whereby health influences socioeconomic position – accounted for a large part of the observed high mortality risk among the non-active labour force, possibly some underestimation of occupational class mortality gradients, and probably some overestimation of income mortality gradients (particularly for cancer and for males).

Multivariate analyses including education, income, car access and labour force status demonstrated a particularly strong association of unemployment with suicide, persistent independent associations of education and car access with all-cause mortality and large reductions to the null of the income–mortality gradient following control for labour force status for cancer and cardiovascular disease. The latter reduction of the income–mortality gradient to the null was, in part at least, probably due to health selection.

Conclusion

Large socioeconomic mortality gradients existed during 1991–94 among 25–64 year old adults in New Zealand for all socioeconomic factors and nearly all causes of death. There was a notably strong association of unemployment with suicide death. Similar linkage of the 1981, 1986 and 1996 New Zealand Censuses to mortality data in the future will allow an investigation of whether these socioeconomic mortality gradients are changing over time in New Zealand.

Statistics New Zealand's Security Statement

Dr Tony Blakely and his co-researchers from the Wellington School of Medicine, University of Otago, initiated the New Zealand Census–Mortality Study. It was approved by the Government Statistician as a Data Laboratory project under the Micro data Access Protocols.

Requirements of the Statistics Act

Under the Statistics Act 1975 the Government Statistician has legal authority to collect and hold information about people, households and businesses, as well as the responsibility of protecting individual information and applying limits to the use to which such information can be put. The obligations of the Statistics Act 1975 on data collected under the Act are summarised below.

1 Information collected under the Statistics Act 1975 can be used only for statistical purposes.

2 No information contained in any individual schedule is to be separately published or disclosed to any person who is not an employee of Statistics New Zealand, except as permitted by sections 21(3B), 37A, 37B and 37C of the Act.

3 This project was carried out under section 21(3B). Under Section 21(3B) the Government Statistician requires an independent contractor under contract to Statistics New Zealand, and any employee of the contractor, to make a statutory declaration of secrecy similar to that required of Statistics New Zealand employees where they will have access to information collected under the Act. For the purposes of implementing the confidentiality provisions of the Act, such contractors are deemed to be employees of Statistics New Zealand.

4 Statistical information published by Statistics New Zealand, and its contracted researchers, shall be arranged in such a manner as to prevent any individual information from being identifiable by any person (other than the person who supplied the information), unless the person owning the information has consented to the publication in such manner, or the publication of information in that manner could not reasonably have been foreseen.

5 The Government Statistician is to make office rules to prevent the unauthorised disclosure of individual information in published statistics.

6 Information provided under the Act is privileged. Except for a prosecution under the Act, no information that is provided under the Act can be disclosed or used in any proceedings. Furthermore, no person who has completed a statutory declaration of secrecy under section 21 can be compelled in any proceedings to give oral testimony regarding individual information or produce a document with respect to any information obtained in the course of administering the Act, except as provided for in the Act.

Census data

The Population Census is the most important stocktake of the population that is carried out. The statistics that are produced provide a regular picture of society. Results are used widely in making decisions affecting every neighbourhood. They are used in planning essential local services and they also help to monitor social programmes ranging from housing to health.

Traditionally, census data is published by Statistics New Zealand in aggregated tables and graphs for use throughout schools, business and homes. Recently, Statistics New Zealand has sought to increase the benefits that can obtained from its data by providing access to approved researchers to carry out research projects. Microdata access is provided, at the discretion of the Government Statistician, to allow authoritative statistical research of benefit to the public of New Zealand.

This project used anonymous census data and mortality data, which were integrated using a probabilistic linking methodology, to create a single dataset that allows the researchers to undertake a statistical study of the association of mortality and socioeconomic factors. This is the first time that the census has been linked to an administrative dataset for purposes other than improving the quality of Statistics New Zealand surveys. The project has been closely monitored to ensure it complies with Statistics New Zealand's strict confidentiality requirements.

Further information

For further information about confidentiality matters in regard to this study please contact either:

Chief Analyst, Analytical Support Division or
Project Manager, Data Laboratory
Statistics New Zealand
PO Box 2922
Wellington
Telephone: +64-4-495 4600
Facsimile: +64-4-495 4610

Chapter 1: Introduction

Box 1: Overview of Chapter 1

Structure of this Chapter

Section 1 details the objectives of this report and of the NZCMS. Section 2 presents a brief overview of social epidemiology and models of the socioeconomic determinants of health and concludes with a framework for use in this report. Sections 3 and 4 deal with study designs and sources of error, respectively.

Objectives of this report

The New Zealand Census–Mortality Study (NZCMS) has been established to measure and monitor the association of socioeconomic factors with mortality in New Zealand. The study data is formed by anonymously and probabilistically linking census and mortality records. This report represents the early stages of the NZCMS, with an exclusive focus on socioeconomic mortality gradients among 25–64 year olds in the 1991 census followed up for mortality for three years. A major objective of this report is to report the association of small area deprivation, education, car access, household tenure, labour force status, occupational class and equivalised household income with all- and cause-specific mortality. A particular and important focus of this report is to quantify the effect of different sources of error (ie, selection bias, health selection, linkage bias and confounding) in the NZCMS.

Social epidemiology and models

Each incremental improvement in socioeconomic position tends to be associated with an incremental reduction in mortality risk. Thus, mortality is distributed as a gradient by socioeconomic position, with little evidence of threshold effects. Social epidemiology is a rapidly expanding field of research attempting to elucidate the causal mechanisms behind this gradient using epidemiological methods. It is important to incorporate layers of causation in social epidemiology, from distal (eg, government economic policy) to proximal (eg, individual health behaviours).

Study designs

Common study designs used to measure socioeconomic mortality gradients:

- unlinked census–mortality studies
- linked census–mortality studies (eg, the NZCMS)
- non-census cohort studies
- ecological and multilevel studies.

Sources of error

Sources of error that affect the measurement of socioeconomic mortality gradients, with a particular focus on those likely to affect a linked census–mortality study such as the NZCMS:

- confounding
- misclassification bias of socioeconomic exposures and confounders
- misclassification of the mortality outcome
- health selection.

(A more detailed discussion of sources of error is included in Appendix B.)

Socioeconomic gradients in health, and more specifically mortality, are pervasive. Poor people die earlier than rich people (Backlund et al 1996; Davey Smith et al 1998c; Kaufman et al 1998; Sorlie et al 1995), poorly educated people die at a higher rate than highly educated people (Feldman et al 1989; Fox and Goldblatt 1982; Kunst and Mackenbach 1994b; Shkolinov et al 1998; Sorlie et al 1995), people from lower social classes die at a higher rate than those from high social classes (Fox and Goldblatt 1982; Marmot et al 1984; Pearce et al 1991; Sorlie et al 1995) and so on. Furthermore, there is no apparent threshold. As a rule, each incremental improvement in socioeconomic position is associated with an incremental decrease in mortality risk. Thus, we are required to consider and research socioeconomic mortality *gradients* that span right across society's socioeconomic hierarchy, not just the health inequalities between those in poverty and those not in poverty (or some other arbitrary threshold) (Howden-Chapman et al 2000).

While socioeconomic inequalities in health are common to all populations that have been investigated, they are not unchangeable. The steepness of the gradient (in relative terms) has tended to increase in the recent decades (Diderichsen and Hallqvist 1997; Feldman et al 1989; Goldblatt 1989; Marang-van de Mheen et al 1998; Marmot and McDowall 1986; Pappas et al 1993; Valkonen 1993), and there is some (although perhaps less than previously thought) variation between countries in the size of the mortality gradient (Kunst et al 1999; Kunst et al 1998d; Kunst and Mackenbach 1994a; Kunst and Mackenbach 1994b; Mackenbach et al 1997). Behind all-cause mortality, the incidence of cardiovascular disease (and tobacco smoking) has shown large changes in its association with socioeconomic position, being initially a disease (lifestyle behaviour) of higher socioeconomic groups before concentrating among lower socioeconomic groups as the epidemic has progressed. Thus, socioeconomic mortality gradients are an appropriate area of focus in public health as the mortality differences are large; everyone in society is included on the socioeconomic ranking, and the gradient appears (in part at least) to be variable over time and between countries. What causes the socioeconomic mortality gradient, and the mutability of the gradient, is the subject of a rapidly expanding international research effort. The New Zealand Census–Mortality Study aims to contribute to this research effort. This report presents the first steps of the larger NZCMS project. The NZCMS aims to measure socioeconomic mortality gradients in New Zealand for all people aged 0–74 years, for a range of socioeconomic factors: small area deprivation, education, labour force status, occupational class, income and wealth (eg, car access). As such, this represents a significant increase in the information on socioeconomic mortality gradients available in New Zealand – previously information was available only for occupational class (15–64 year old males) (Pearce et al 1991) and small area deprivation (Salmond and Crampton 2000).

1 Objectives of this report and of the NZCMS

The objectives of this report are to:

- consider the various study designs and sources of error encountered when measuring socioeconomic mortality gradients, with a particular focus on linked census–mortality studies
- review the literature on socioeconomic mortality gradients
- briefly describe the record linkage methodology used in the NZCMS to link 1991 census records to mortality records for 1991 to 1994, and the output of that record linkage
- present the univariate and multivariate results for the association of small area deprivation, education, labour force status, occupational class, housing tenure, car access and income with mortality (all-cause and cause-specific) for 25–64 year olds, by sex, for the 1991 linked census–mortality study
- interpret the results with regard to causal inference and policy relevance.

This report may be viewed as summarising the establishment, processes and initial results of the larger NZCMS project. As such, the NZCMS will build on the first wave of results presented in this report by guiding analyses:

- for each of the four linked census–mortality studies (1981, 1986, 1991 and 1996 census cohorts), thereby allowing a measurement of trends in socioeconomic mortality gradients over time
- for children (0–14 year olds) and youth (15–24 year olds), and the elderly (65–74 year olds), in addition to 25–64 year olds
- comparing socioeconomic mortality gradients in New Zealand with other countries (eg, the cross-national comparison studies by Kunst, Mackenbach, and other colleagues involved in the European Union Working Group on Socioeconomic Inequalities in Health).
- of socioeconomic mortality gradients within specific ethnic groups
- of the numerator–denominator bias for ethnicity between census and mortality data
- of smoking (1981 and 1996 censuses) as a mediator between socioeconomic position and mortality in New Zealand
- of the association of ecological socioeconomic measures (eg, small area deprivation, income inequality and social capital) with mortality controlling for personal socioeconomic factors.

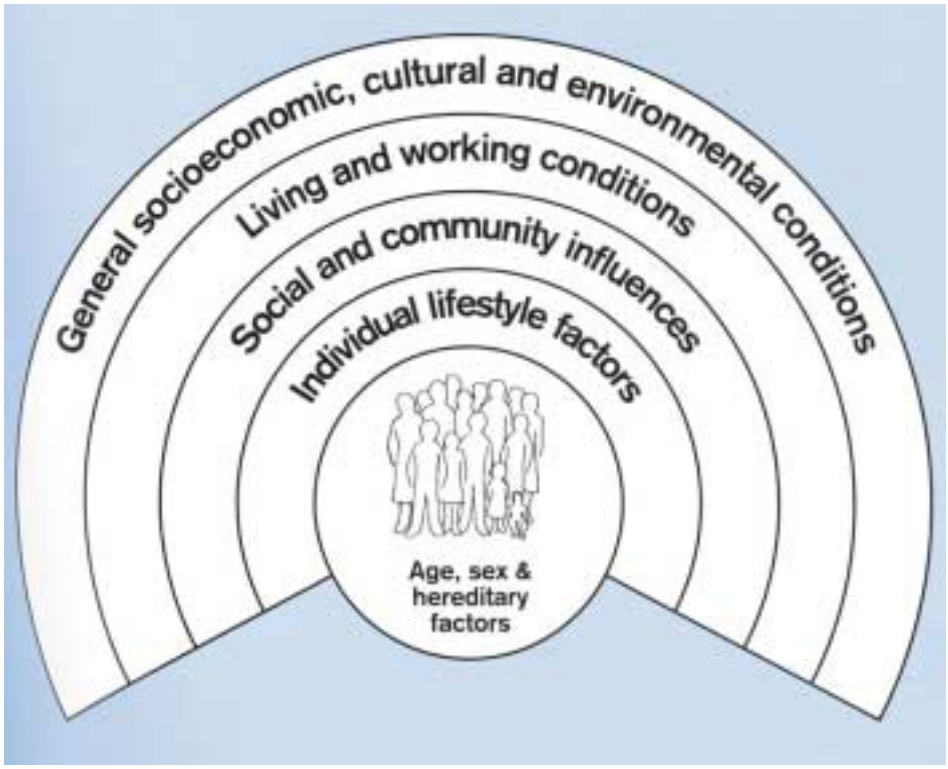
This report is based on a PhD thesis (Blakely 2001). Compared to the original thesis, this report has moved much of the technical information (ie, sources of error and sensitivity analyses) to appendices and focuses more on the process of the NZCMS and early substantive results.

2 Epidemiological theory and models of the socioeconomic determinants of health

Epidemiological theory (explicit or implicit) frames the questions we ask as epidemiologists and the analyses we conduct (Krieger and Zierler 1996). The dominant, if rarely explicitly acknowledged, epidemiological theory post World War II is the lifestyle theory of disease causation (Krieger 1994; Susser and Susser 1996a). Under this theory, it is assumed that individuals directly control their lifestyles and, hence, risk of disease – the fact that behaviours are socially patterned is overlooked. An alternative 'ecosocial' (Krieger 1994; Krieger 1999) theory of disease causation attempts to incorporate how social phenomena (eg, discrimination, violence, income distribution, residential segregation, welfare policies, personal socioeconomic position) affect human biology, both directly and indirectly through mediated pathways (including risk factors and lifestyles). This alternative social theory of disease causation is the epidemiological theory that frames the research questions we ask in social epidemiology.

A relatively simple way to conceptualise (or model) the socioeconomic determinants of health and disease is to think of layers of causation or analysis. While similar to many other representations of the layers of causation, the model proposed by Dahlgren and Whitehead (1991) and used by Benzeval et al (1995) has been popular in New Zealand. This model is shown in Figure 1.

Figure 1: A layered model of the socioeconomic determinants of health

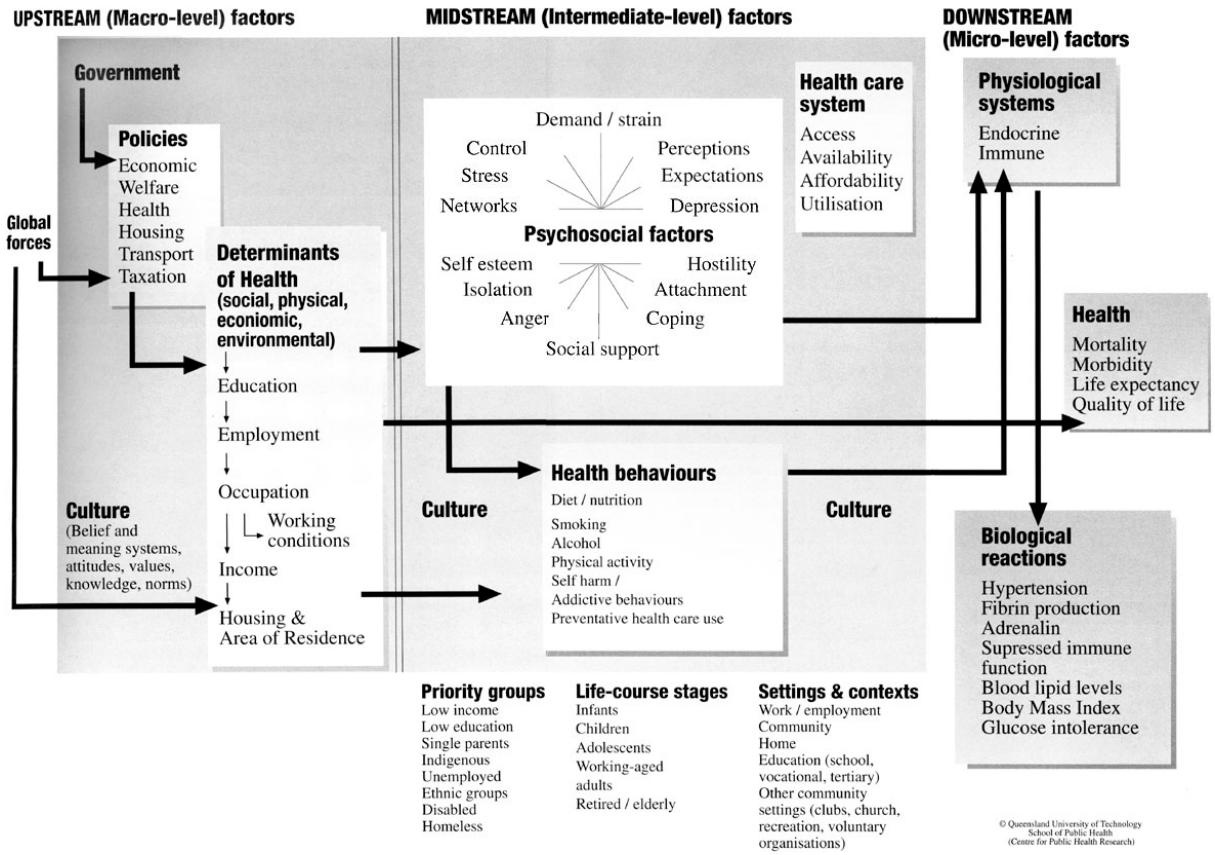


Source: Dahlgren and Whitehead 1991

A more detailed model of the socioeconomic determinants of health and disease has been provided by Turrell et al (1999) (Figure 2). This Australian model usefully presents:

- upstream, midstream and downstream factors in one model, without dichotomising social forces versus lifestyle risks
- global forces, culture and 'government policies' as determinants of the distribution of personal socioeconomic factors ('determinants of health')
- a direct effect of the 'social determinants' on health, *and* an indirect effect mediated through psychosocial factors and health behaviours (ie, the biopsychosocial model is incorporated)
- the importance of priority groups, life-course stages, and settings and contexts as organising structures over which the midstream factors coalesce and interact to affect health.

Figure 2: Framework of the socioeconomic determinants of health



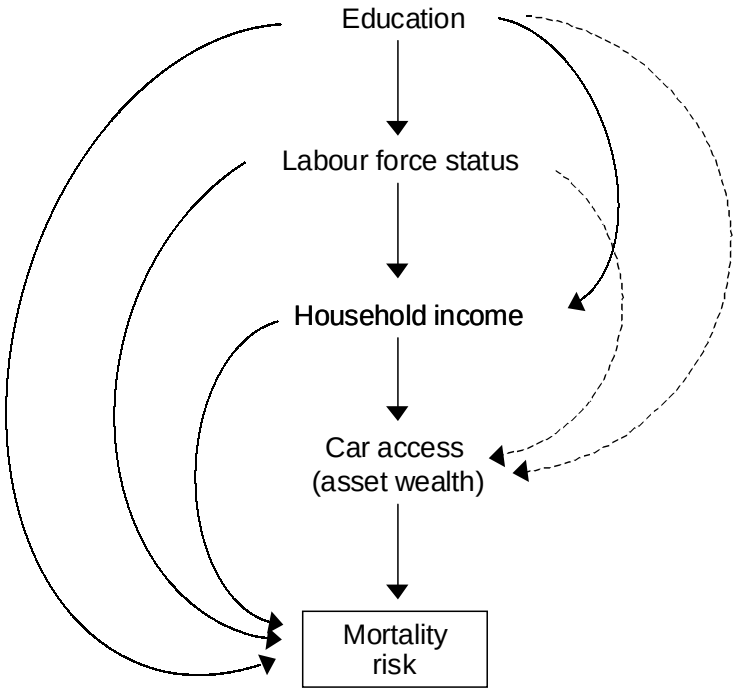
Source: Turrell et al 1999

A limitation of the Turrell et al model in the New Zealand setting is the lack of prominence afforded to ethnicity. Also, other than income and housing, material factors (eg, crowding, financial difficulties, asset wealth, capital, diet, exposure to harmful substances) are arguably underemphasised as either upstream determinants or midstream determinants of health. In another model, the MacArthur Network on SES and Health posits education, occupation and income at the most distal level, and 'environmental resources and constraints' (eg, material resources) and 'psychological influences' at the next level closer to individuals (Adler and Ostrove 1999). 'Exposure to carcinogens and pathogens', 'behaviour', and 'CNS and endocrine response' are the most proximal level to individual health and illness in this MacArthur Network model.

In the NZCMS, information is only available for the 'Determinants of Health' represented in the upstream panel of the Turrell et al framework (Figure 2), (ie, small area deprivation, education, income, occupational class, labour force status and asset ownership). (As the NZCMS progresses to include linkage of other censuses to mortality data, we will also be able to look at how 'policies' may alter socioeconomic mortality gradients by comparing results across separate linked census–mortality cohorts. Also, the 1981 and 1996 censuses provide smoking data allowing an examination of one health behaviour as a pathway to mortality.) While it is not possible in this report to explore other pathways that mediate the effect of socioeconomic factors on mortality risk, it is important to have some understanding of these possible downstream pathways in order to place the results of the NZCMS in context.

A reduced framework or causal diagram is necessary to guide the analyses in this report, and the NZCMS more generally – an initial one is presented in Figure 3. The direction of the arrows in Figure 3 represents the chronological ordering of the socioeconomic factors for the 25–64 year old cohort: education usually precedes entry to the labour force; labour force status affects income; and car access (as a measure of asset wealth) is a function of income. Education, labour force status and income are all represented as having a direct association with mortality (ie, the arrows directly from each socioeconomic factor to mortality), as well as an association with mortality mediated by other socioeconomic factors. An arrow is shown from education to income, suggesting that education has an influence on household income over and above whether or not you have employment. Dotted arrows are also shown from education and labour force status to car access (as a measure of asset wealth), suggesting there might be an association independent of that via income. For example, education may impart a propensity to invest in the future regardless of income, and labour force status may be a proxy for attributes (eg, self-motivation) that may also be determinants of asset acquisition independent of income.

Figure 3: Initial framework of the association of socioeconomic factors with mortality used in this report



The framework in Figure 3 is, obviously, a simplification of reality, and includes only variables that are available in the NZCMS. However, simplifications in causal diagrams are essential in any epidemiological analysis (Greenland et al 1999), and it is better to be explicit about those necessary simplifications of reality rather than leave them implicit. Even just considering additional variables available in the NZCMS, occupational class and housing tenure (another measure of asset wealth) could have been included. However, the New Zealand Census uses current occupation to assign occupational class, meaning that occupational class is missing for a large percentage of the cohort – so it has been excluded from the analytical framework.

Further, in reality each socioeconomic factor is not compartmentalised at one point in a causal diagram – income affects asset wealth which, if invested wisely, may further enhance income. The framework is specific to 25–64 year olds in the NZCMS. For example, over the long-term parental income may influence children's education – but the framework in Figure 3 assumes that education is relatively fixed by age of entry into the 25–64 year old age group, and that we are interested in those individual's mortality risk, not their children's.

Based on the univariate results and sensitivity analyses presented later in this report, the framework in Figure 3 had to be revised before conducting the multivariate analyses (see revision at commencement of Section 9, Chapter 4). For example, labour force status was that at census night, and as such would have been a strong correlate of health status on census night. (People in poor health are often not in the active labour force.) Controlling for labour force status, therefore, risked over-adjusting the association of any socioeconomic factor with mortality, as (other than injury deaths) health status is an intermediary step between socioeconomic factors and mortality. Health selection effects also made analysis of income–mortality gradients that controlled for labour force status difficult to interpret.

3 Measuring socioeconomic mortality gradients – study designs

3.1 An overview of study designs to measure socioeconomic mortality gradients

It is helpful to think of six types of study that can be used to measure socioeconomic mortality gradients – they are shown in Table 1. The first four study types have been described previously by Valkonen (1993). Cohort studies have been more commonly used than case-control studies to measure socioeconomic mortality gradients – particularly for all-cause mortality. Multilevel studies combine measures of 'average' socioeconomic position obtained in an ecological study with individual-level data on socioeconomic factors obtained in any one of the five other studies in Table 1.

Table 1: Types of studies used to measure socioeconomic mortality gradients

Study type	Description
1 Unlinked census–mortality study	The distribution by socioeconomic position is determined separately for the numerator deaths (ie, mortality data) and denominator population (ie, census data), then compared at an aggregate level.
2 Linked census–mortality study	Mortality records are linked to census records, then mortality rates for each socioeconomic group are determined directly as per a cohort study design.
3 Non-census cohort study	A typical cohort study design, using a non-census data set as the study base.
4 Ecological study	Comparison by groups (eg, regions) of a measure of 'average' socioeconomic position and mortality rates.
5 Case-control study	A study involving the comparison of the distribution of socioeconomic factors among people who die (cases) with that among a suitably selected sample of living people (controls).
6 Multilevel study	In social epidemiology, a study that includes measures of individual-level and contextual socioeconomic factors. Thus, a multilevel study combines features of an ecological study and one of the four other study types.

In this section, describes the first four study types listed in Table 1, and include a brief description of multilevel studies with the description of ecological studies.

3.2 Unlinked census–mortality studies

An unlinked census–mortality study design is usually associated with studies of occupational class mortality gradients. However, the underlying method is the same as that used in all routinely published mortality data by sex, age, ethnic group, small area deprivation and region – for example, those found in *Our Health, Our Future* (Ministry of Health 1999).

Unlinked studies of occupational class have a long history. The series of Decennial Supplements on Occupational Mortality published by the Registrar-General of England and Wales dates back to the 1921 census for occupational/social class analyses, and back to 1851 for specific occupational analyses (Valkonen 1993). Essentially, social class (more correctly occupational class) is derived from the occupation stated on death certificates (numerator) and census data (denominator), and class-specific mortality rates are then calculated. By ordering the occupational classes from highest (eg, professionals) to lowest (eg, labourers), the size of the socioeconomic mortality gradient can be determined by comparing mortality rates between the highest and lowest occupational classes or by summarising the gradient across the ordered occupational classes.

The major body of work on socioeconomic mortality gradients in New Zealand is that by Pearce and colleagues, using an unlinked census–mortality study design to measure occupational class mortality gradients about the times of the 1976 and 1986 censuses. (Results using the 1996 census have been submitted). All analyses were restricted to 15–64 year old males. Analyses for females were not possible due to the relatively low involvement of females in the labour force, and analyses for the elderly were not possible as the New Zealand censuses elicit current occupation only. The major analyses included:

- all-cause socioeconomic mortality gradients, 1974–78 (Pearce et al 1983a) and 1985–87 (Pearce et al 1991)
- cause-specific socioeconomic mortality gradients, 1974–78 and 1985–87 (Kawachi et al 1991; Marshall et al 1993; Pearce and Bethwaite 1997; Pearce et al 1983b; Pearce and Howard 1986)
- ethnic-specific socioeconomic mortality gradients, 1974–78 (Pearce et al 1984; Pearce et al 1985) and 1985–87 (Pearce et al 1993).

A well-known limitation of unlinked studies is possible numerator–denominator bias, as the mortality and census occupational data are collected by different systems. For example:

- occupation on the death certificate (actually the 'death registration form' (BDM28 form) in New Zealand) is collected by the undertaker from the next of kin, whereas it is self-identified on the census. Thus, the occupation elicited may differ on census data and mortality data. This misclassification bias could be random or non-differential, but a systematic or differential bias is possible (eg, there may be a tendency for next of kin to 'promote the dead')
- there may be variation between census and mortality data as to whether current, most recent or main lifetime occupation is elicited. In New Zealand the undertaker elicits from the next of kin the deceased's '*usual* occupation, profession or job'. However, the 1991 census elicits information about *current* occupation only. (Note that in other countries the census question differs (eg, the 1971 England census asked about most recent occupation if not currently working).)
- the amount of information collected varies between mortality and census data. For example, in New Zealand just occupational title is collected on mortality data, but both occupational title and tasks/roles are collected on census data. Thus, there will be misclassification biases of occupation codes (eg, New Zealand Standard Classification of Occupations) within each of census and mortality data. This misclassification could be either non-differential or differential, and the amount of misclassification may vary between census and mortality data.

Comparative evidence from the Office of Population Censuses and Surveys (OPCS) Longitudinal Study suggests that, in England and Wales at least, the numerator–denominator bias was not substantial in the Decennial Supplements that used an unlinked study design (Fox and Goldblatt 1982). Moreover, evidence is emerging that unlinked census–mortality studies may be superior to linked census–mortality studies when only current occupation is available on census data (as is the case in New Zealand) as opposed to current or last occupation (as in England and Wales) (Kunst 1997; Kunst et al 1998b; Martikainen and Valkonen 1999).

3.3 Linked census–mortality studies

A linked census–mortality study takes routine census data and links it to routine mortality data for deaths occurring in some time period following the census. Thus, it is a prospective cohort study of the entire census population, followed up for mortality. The linkage aims to match a decedent's mortality record to his or her actual census record; the linkage is *not* just statistically matching 'like' mortality records with 'like' census records. An inventory of linked studies by country is shown in Table 2.

Table 2: Linked census–mortality studies by country

Country (n, approx †)	Linkage description	Percentage-linked and accuracy of linkage	Example of research outputs
Canada ‡ (n=47,935, all ages)	Sample of 1986 census records from Manitoba Probabilistically and anonymously linked to health insurance records Mortality in three years following census determined from health records	74% of census records linked to a health insurance plan record 95.5% of links were estimated to be true links (Houle et al 1996)	Methods (Houle et al 1996) Income and education (Mustard et al 1997)
Denmark (3 million)	1970–86 censuses Some linkage with tax files Linkage to mortality and cancer registry data	>99% of deaths ascertained (Kunst et al 1996) Accuracy not stated in accessible publications, but presumably good due to use of personal identifier	Unemployment and cause-specific mortality (Iversen et al 1987) Unemployment and cancer mortality (Lynge and Andersen 1997) (Relatively few Danish-specific studies using the linked data were found by Medline searches, but the data set has been used for cross-national comparisons – see Table 3.)
Finland (3 million)	Five-yearly censuses (since 1970) Linkage between censuses Linkage of each census to mortality data for the following five years	>99% of deaths ascertained (Martikainen and Valkonen 1998a) Accuracy not stated – but probably high	Education, income, occupational class and labour force status, and all- and cause-specific mortality (Martikainen and Valkonen 1998a) (Martikainen 1990; Martikainen and Valkonen 1996; Valkonen 1993) Socioeconomic mortality gradients among the elderly (Martelin 1994; Martelin et al 1998) Socioeconomic mortality gradients by gender (Koskinen and Martelin 1994) Education and health expectancy (Valkonen et al 1997) Diabetes (Koskinen et al 1996) Contribution of alcohol to socioeconomic mortality gradient (Makela et al 1997) Health selection, and labour force status and occupational class (Martikainen and Valkonen 1998b; Martikainen and Valkonen 1999)
France (800,000)	2–3% sample of 1975 census (Kunst et al 1996) Linkage to 1975–89 mortality data	Not available (Kunst et al 1996)	(No French-specific studies found in the English language journals by Medline – although there is a limited number of publications in French (eg, Desplanques 1984). The data set has been used for cross-national comparisons – see Table 3.)

Country (n, approx †)	Linkage description	Percentage-linked and accuracy of linkage	Example of research outputs
Italy (36 million)	1981 census cohort Linked to mortality data for the following six months	75–78% of mortality records linked to a census record (Faggiano et al 1995; Kunst et al 1996) Not stated	Education and cancer mortality (Faggiano et al 1995) Occupational class and cancer mortality (Costantini et al 1994) (Relatively few Italian-specific studies were found in the English-language journals by Medline searches, but the data set has been used for cross-national comparisons – see Table 3.)
Italy (Turin) (700,000)	1981 census cohort Linked to 1981–89 mortality data	>99% of mortality records linked to a census record (Kunst et al 1996) Not stated	Education and cancer mortality (Rosso et al 1997) Socioeconomic status (generally) and mortality (Cardano et al 1999) Relatively few Italian-specific studies found in the English-language journals by Medline searches, but the data set has been used for cross-national comparisons – see Table 3.)
Norway (2.6 million)	1970 and 1980 censuses Linkage between censuses Linkage to 1980–90 mortality data	Not stated in accessible publications, but presumably good due to use of personal identifier	Health selection, and occupational class (Dahl 1993) Breast cancer mortality among fishermen's wives (Lund and Bona 1993) (Relatively few Norwegian-specific studies were found by Medline searches, but the data set has been used for cross-national comparisons – see Table 3.)
Sweden (5 million)	1960 to 1985 censuses Linkage between censuses, and between census and other routine data (eg, tax files) Linkage of each census to mortality data for the following 5 or 10 years	>98% of deaths ascertained (Diderichsen and Hallqvist 1997; Nordic Statistical Secretariat 1988; Statistics Sweden 1982) Accuracy not stated – but probably high	Occupation and cause-specific mortality, males (Diderichsen and Hallqvist 1997) Unemployment and all-cause mortality (Stefansson 1991) Childhood and adulthood occupational class, and all-cause and IHD mortality (Vagero and Leon 1994) Occupational class and avoidable mortality (Westerling et al 1996) Provision of data on adult socioeconomic position (income, education, car access and occupational class) for a study of the association of fetal growth rate (cohort born 1915–29) and adult IHD (Leon et al 1998)
United Kingdom (OPCS Longitudinal Survey; 500,000 all ages)	1% sample of 1971 census, replenished by 1981 census Linkage between censuses Ongoing linkage to mortality data	>95% of deaths among cohort in 1971–75 were ascertained <1% of links incorrect (Fox and Goldblatt 1982)	Initial technical report and results (Fox and Goldblatt 1982) Labour force status (Bethune 1996; Moser et al 1984; Moser et al 1986; Moser et al 1987) Social class (Fox et al 1985; Goldblatt 1989; Harding 1995) Occupation (Fox and Adelstein 1978; Moser and Goldblatt 1991) Asset ownership (Filakti and Fox 1995; Goldblatt and Fox 1979) Suicide (Lewis and Sloggett 1998)
USA (National Longitudinal Mortality Study; NLMS; 1.3 million all ages)	Respondents to the Current Population Survey (CPS) between 1979 and 1985 Probabilistically linked to national mortality data up to 1989	>93% of deaths ascertained Perhaps 1% of links being incorrect (Calle and Terrell 1993; Rogot et al 1992b)	Initial technical report (Rogot et al 1992b) Record linkage methods (Rogot et al 1986) Labour force status (Sorlie and Rogot 1990) Income (Backlund et al 1996) Education (Elo and Preston 1996) Labour force status, education, income and social class (Backlund et al 1999; Rogot et al 1992a; Sorlie et al 1995) Ethnicity (Sorlie et al 1993; Sorlie et al 1992a; Sorlie et al 1992b) Stroke (Howard et al 1997; Howard et al 1994; Howard et al 1995) Neighbourhood and personal income (Anderson et al 1997) Children and youth (Singh and Yu 1996a; Singh and Yu 1996b)
USA (1960 Matched Records Study)	1960 census population Manually linked to mortality records for deaths in the four months after census	262,966 of 340,033 (77%) of mortality records linked to census record Accuracy not estimated, but probably high	Kitagawa and Hauser (1973) *Differential Mortality in the United States* – a full report of socioeconomic mortality gradients by education, income, and occupational group, by age, sex, race and cause of death (Kitagawa and Hauser 1973).

† Numbers of 20–74 year olds unless stated otherwise, taken from (Kunst et al 1996).

‡ Results of the record linkage are reported as the percentage of census records linked to a health insurance plan record, contrary to reporting in other studies. The underlying record linkage project was of all census records and all health insurance plan records – for privacy reasons the analytical sample was limited to 20,000 households. As the health insurance plan is universal, the percentage of health insurance plan records linked to a census record would also have been about 74%. Separate linkage results are not provided for mortality events only.

The Scandinavian countries set the standard in linked census–mortality studies. The introduction after World War Two of unique personal identifiers for each citizen that are routinely recorded on different administrative data sets makes record linkage relatively easy. Consequently, linkage between census and mortality records is near complete and highly accurate (Table 2). Furthermore, census data sets are routinely linked to previous/subsequent census data sets, and non-health data is also linked to census data (eg, tax data allowing an accurate measurement of net income). Much research on socioeconomic mortality gradients has emanated from Sweden and Finland, but surprisingly little (given the quality of data) from Norway and Denmark (Table 2).

An exemplary study of how thorough the linkage of data is in Sweden is that by Leon et al (1998). They retrospectively assembled a birth cohort of 14,611 people born in one region of Sweden during 1915–29. For the period prior to World War Two the researchers followed the cohort by a statutory system of parish archives. After 1950, follow up was simply conducted by using personal identification numbers, allowing incorporation of census socioeconomic data, immigration/emigration data, and finally linkage to mortality data. Thus, antenatal and perinatal data, childhood socioeconomic position (parental occupational class), adult socioeconomic position (at several censuses) and mortality outcomes were combined for each individual, allowing an elegant study of whether birth weight was associated with adult ischaemic heart disease mortality independent of life-course socioeconomic position (Leon et al 1998). (The conclusion was that increased birth weight was associated with decreased IHD mortality.)

Perhaps the best known linked census–mortality study in the English-speaking world is the Office of Population Surveys and Censuses (OPCS) Longitudinal Study (LS). In this study, a 1% sample of the 1971 census population has been followed up for mortality, and linked with 1981 census data. In the absence of unique identification numbers in England and Wales, linkage to mortality and immigration/emigration data was conducted manually (Fox and Goldblatt 1982). However, the presence of names and text addresses on both census and mortality data has ensured that linkage is near complete and accurate.

The US equivalent is the National Longitudinal Mortality Study (NLMS). Strictly speaking, the NLMS is not a census–mortality linkage study as the 'census' records are actually Current Population Survey (CPS) records. The CPS is similar to New Zealand's Household Economic Survey and Household Labour Force Survey. The NLMS pooled the CPS subjects aged 25 years and older surveyed between 1979 and 1985 to form the cohort study base. In other respects, the NLMS is similar to the other linked studies – hence the inclusion in Table 2. The NLMS also uses probabilistic record linkage methods to link the CPS and mortality records – the same underlying record linkage method used in the NZCMS. (Probabilistic linkage methods are described in this chapter. Briefly, in the absence of a unique identifier, and with many records, the most accurate and efficient way to link records is probabilistically. 'Probabilistically' refers to the use of agreement and disagreement weights for each comparison of matching variables. The weights are derived from the probability of a match purely by chance when the comparison pair is not a true link or from the probability of a match when the comparison pair is actually a true link.)

Being a probabilistic record linkage study, the NLMS requires scrutiny of the completeness and accuracy of the record linkage. At least 93% of the deaths were estimated to be successfully ascertained – this equates to the sensitivity of the record linkage as a 'test' to detect actual death outcomes. Regarding the accuracy of the ascertained links, perhaps 1% were false links – this equates to a positive predictive value (PPV) of 99% for the record linkage. To estimate the PPV, Calle et al submitted a test sample of 15,000 participants of the American Cancer Society cohort with known vital status (stratified random sample: 5000 dead and 10,000 alive) to an automated and probabilistic record linkage with the US National Death Index (Calle and Terrell 1993). While not an exact examination of the accuracy of the record linkage in the NLMS, this PPV estimate is probably a reasonable estimate of the PPV in the NLMS given the same underlying record linkage methodology (personal communication, Paul Sorlie, July 1999).

A classic record linkage study is the 1960 Matched Records Study in the US, reported by Hauser and Kitagawa (Kitagawa and Hauser 1973). This study involved taking all 340,033 deaths occurring in the four months following the 1960 census, and assigning each mortality record an enumeration district code on the basis of the address on the death registration form. For each mortality record, a manual search was conducted of the 1960 census data set for the same individual within the coded enumeration district using the name, address, and other personal details of the decedent – a huge task! Seventy-seven percent of the mortality records were linked back to their census record. However, only 25% of the total census population was required to complete questions on socioeconomic factors such as family income, education, and occupation. Thus, the number of census records and decedents in each category of socioeconomic factor (by sex, age, and race) was determined by: estimating the denominator number of census records from the 25% census sample and estimating the numerator number of deaths from the 25% sample of the 77% of mortality records linked to the complete census data set. Regarding this latter estimation, considerable effort was undertaken to adjust for any bias between the 23% of mortality records not linked to a census record, and the 77% of mortality records that were linked, by sending a questionnaire to the next of kin of 9541 decedents (either linked or unlinked to a census record) to elicit sociodemographic characteristics. Based on this questionnaire, the following biases in the record linkage were estimated:

- Over 80% of decedents aged 55 years and over were linked back to their census record, but only 65% of 15–35 year old and less than one year old decedents were linked. Kitagawa and Hauser (1973) attributed this lower success to a higher residential mobility of young adults.
- Married decedents were more likely to be linked than widowed, divorced or never married decedents – also probably a function of residential mobility.

A relatively small sample of mortality and census records have been probabilistically linked in Manitoba, Canada (Houle et al 1996) (Table 2). Due to privacy requirements similar to those in New Zealand, names and text addresses were not used in the record linkage. Thus the record linkage used anonymous data (eg, geocodes, date of birth, sex). Probabilistic record linkage methods were used. The results were reported for linkage of census to health insurance records, not mortality records to census records (Houle et al 1996). Seventy-four percent of the selected census records were linked to a health record, with 95.5% of these links estimated to be true links based on examination of names and addresses for a subset (n=1000) of the nearly 50,000 total census records.

The complete Italian 1981 census data set has been linked to mortality records for the six months following that census (Faggiano et al 1995). Similar to the Canadian experience (although reported as the percentage of mortality records linked to census records), 75–78% of mortality records were linked to a census record (Faggiano et al 1995; Kunst et al 1996). English language publications do not state whether probabilistic linkage methods were used (although it seems highly likely), and whether the accuracy of the linkage was determined.

The small Canadian and the national Italian linkage studies approximate the NZCMS more closely than the remaining studies in Table 2, as:

- anonymous mortality and census data were used
- geocodes (based on the unavailable text address) were critical in the record linkage process
- the absence of unique identifiers and/or names and addresses meant the percentage of mortality records linked to a census record was substantially less than 100%.

The studies shown in Table 2 have been used extensively in cross-national comparisons. Selected examples of these cross-national studies are shown in Table 3. By far the greatest contribution to cross-national comparisons has been provided by Kunst, Mackenbach, and other colleagues involved in the European Union Working Group on Socioeconomic Inequalities in Health.

Table 3: Selected examples of cross-national comparisons of socioeconomic mortality gradients ‡

Authors	Socioeconomic factor, disease grouping and countries included
Occupational class	
Vagero and Lundberg 1989	All-cause mortality, males: *England and Wales*, and Sweden
Kunst et al 1998c	All-cause mortality, males: Denmark, Finland, France, England and Wales, *Ireland*, Italy (Turin), Norway, *Portugal*, Sweden, *Switzerland*, *Spain*
Kunst et al 1998d	Cause-specific mortality, males: 11 European countries above
Kunst et al 1999	Ischaemic heart disease mortality, males: 11 European countries above, and the United States
Kunst et al 1998a	Stroke, males: 11 European countries above, and the United States
Education	
Kunst and Mackenbach 1994b	All-cause mortality, males: Denmark, Finland, France, Great Britain, Italy (Turin), Netherlands, Norway, Sweden, and the United States
Mackenbach et al 1999	All- and cause-specific mortality, males and females: *Czech Republic*, *Estonia*, Finland, *Hungary*, Italy (Turin), Norway, and United States
Sihvonen et al 1998	Health expectancy, males and females: Finland and Norway
Occupational class and education	
Lahelma and Valkonen 1990	All cause-mortality, males and females: Denmark, England and Wales, Finland, *Hungary*, Norway, and Sweden
Mackenbach et al 1997	All-cause mortality, males: Denmark, Finland, France, Great Britain, Italy, Norway, Sweden, *Switzerland*, and *Spain*

‡ Data from countries in italics were not from l inked census–mortality record studies. For these countries, the data was usually from unlinked cross-sectional studies of occupational class mortality gradients. Data from England and Wales, and the United States, were not from linkage of entire census data sets with mortality data, but from the OPCS Longitudinal Study and the NLMS, respectively.

3.4 Non-census cohort studies

Both the above study designs use routinely collected census and mortality data. More commonly in epidemiology, a study base is assembled for a specific purpose. Much useful information on socioeconomic mortality gradients comes from cohort studies principally examining other exposures (eg, cardiac risk factors) where socioeconomic factors were perhaps measured as potential confounders. Non-census cohort studies typically include variables between socioeconomic position and mortality (eg, behavioural factors, physiological measures and psychosocial variables). Thus, these epidemiological studies are particularly important in the search for explanations of how socioeconomic factors 'get under the skin' and affect health – the midstream factors in Figure 2.

An extension to the basic cohort study design is to include longitudinal or repeated measures. For example, income could be measured at several stages allowing a 'life-course' analysis of the association of income with health (Kuh and Ben-Shlomo 1997). Inclusion of longitudinal data or repeated measures is also feasible with linked record studies (eg, Scandinavian-linked record studies have built up socioeconomic profiles of individuals by linking census to census records, in addition to linking census to mortality records) (see Table 2).

The disadvantages of non-census cohort studies include cost, loss of follow-up, the long period of time between set-up and obtaining results, and (comparatively) few deaths, thus reducing statistical power.

Some non-census cohort studies of socioeconomic mortality gradients of international note include:

- the Whitehall Study I of 18,000 male UK civil servants recruited in 1967–69 (Davey Smith et al 1990; Marmot and Shipley 1996; Marmot et al 1984)
- the West of Scotland collaborative study of 5645 working men aged 35–64 recruited in 1970–73 (Blane et al 1996; Davey Smith et al 1997; Davey Smith et al 1998a; Davey Smith et al 1998b)
- the US Multiple Risk Factor Intervention Trial (MRFIT) of 300,685 white males and 20,224 black males aged 35–57 years recruited in 1973–75 (no personal socioeconomic factor data collected at screening, so median zip code income used) (Davey Smith et al 1996a; Davey Smith et al 1996b)
- the Dutch Longitudinal Study on Socioeconomic Health Differences of 27,070 non-institutionalised adults aged 15–74 years recruited in 1991 (Mackenbach et al 1994; Schrijvers et al 1999).

While not linking census and mortality records, other government sponsored cross-sectional surveys (eg, health and nutrition surveys) have been linked to subsequent mortality data to create a cohort study:

- the US National Health and Nutrition Examination Survey (NHANES) (Feldman et al 1989)
- the US Panel Study of Income Dynamics (McDonough et al 1997).

3.5 Ecological and multilevel studies

3.5.1 Ecological studies

An ecological study uses groups of people as the unit of analysis, not individual people as the unit of analysis (Last 1995). Thus the average/summary score for some exposure measured by ecological units is compared to some average occurrence of disease/health measured for the same ecological units. Examples include:

- correlations between sociodemographic indicators and cancer standardised mortality ratios (SMR) by areas in Boston (Jenkins 1983)
- the association of a composite index of socioeconomic position with mortality by neighbourhoods in Philadelphia (Dayal et al 1986)
- the association of percentage car ownership and percentage non-manual occupational class with the SMR by towns in the UK (Pocock et al 1980; Pocock et al 1987; Shaper 1984)
- the association of socioeconomic position with infant mortality by small urban areas in Australia (Turrell and Mengersen 2000)
- the association of social deprivation with premature mortality by regions in England (Eames et al 1993)
- the association of income inequality with mortality by country (Wilkinson 1992)
- the association of income inequality with mortality by states in the US (Kaplan et al 1996; Kennedy et al 1996)
- the association of metropolitan area-level income inequality with mortality by metropolitan areas in the US (Lynch et al 1998)
- the association of income inequality and social capital with mortality by states in the US (Kawachi et al 1997).

The interpretation of ecological studies may be partitioned by level of inference (Diez-Roux 1998):

- inference is to the individual level
- inference is to the ecological level.

In the first variant, an ecological analysis is used to *ecologically infer* the individual-level association (Beral et al 1979; Norstrom 1988). For example, the association of neighbourhood income with health is assumed to reflect the parallel individual-level association of personal income with personal health.

In the second variant, an ecological analysis is used to measure the effect of ecological-level processes on health. For example, the association of mean income with health at the neighbourhood level may reflect more than just the underlying individual-level association; both one's personal income and the community's income may have independent effects on health. An alternative and less ambiguous example of the second variant of inference in ecological studies is the recent ecological studies on the association of income inequality with mortality (Kaplan et al 1996; Kawachi et al 1997; Kennedy et al 1996; Lynch et al 1998; Wilkinson 1992). There is no parallel for income inequality at the individual level, and thus these studies explicitly aim to measure the effect of a contextual or ecological variable on health.

Table 4: Types of fallacy in multilevel research (taken from Diez-Roux 1998)

Unit of analysis	Level of inference	Type of fallacy
Group	Individual	Ecological
Individual	Group	Atomistic †
Individual: relevant group-level variables excluded	Individual	Psychologistic †
Group: relevant individual-level variables excluded	Group	Sociologistic

† Also called individualistic by some authors.

Each variant of inference (to the individual-level association or the ecological-level association) is prone to error, namely the ecological fallacy (Robinson 1950; Selvin 1958) and the sociologistic fallacy (Riley 1963), respectively (see Table 4). The ecological fallacy arises when both the independent and dependent variables are measured at the ecological level (ie, an ecological study), the inference is to the individual level (ie, ecological inference), but that ecological inference is incorrect (Diez-Roux 1998). There are several reasons why an ecological fallacy may arise (Blakely 1996; Greenland 1992; Greenland and Morgenstern 1989; Greenland and Robins 1994; Morgenstern 1995; Morgenstern 1998), including:

- confounding at the ecological level, whereby some characteristic of the ecological units confounds the association of the exposure and outcome of interest *independently of any individual-level confounding*
- effect modification, whereby the individual-level association of the exposure and outcome of interest varies by ecological unit
- confounding at the individual level within ecological units, whereby some individual-level covariate confounds the association of the exposure and outcome (at the disaggregated individual level) and that covariate varies across ecological units
- model misspecification
- misclassification of variables
- inappropriate standardisation of ecological variables (eg, ecological outcome is age standardised, but the ecological exposure is not).

On the other hand, when the results of an ecological study are used to infer an *ecological-level* association, that inference may be incorrect due to a sociologistic fallacy. For example, it would be a sociologistic fallacy to infer that the association of mean neighbourhood income with health was purely a contextual effect, and not explained at all by the parallel association of individual income with personal health.

There is little need nowadays for ecological studies to infer the individual-level association of socioeconomic factors with health – there is ample evidence from other types of studies. However, there is strong interest in determining whether ecological variables (eg, income inequality and small area deprivation) have any independent effect on health over and above the association of individual-level, socioeconomic factors with health. To avoid a sociologistic fallacy in this latter instance, multilevel studies are required.

3.5.2 Multilevel studies

There is an increasing interest amongst social epidemiologists in how social contexts may shape health status in addition to traditional individual-level risk factors (Krieger 1994; Mackenbach 1995; Pearce 1996; Susser 1994; Susser 1998; Susser and Susser 1996b; Woodward 1996). In particular, many researchers of the socioeconomic determinants of health are beginning to analyse ecological and individual-level socioeconomic factors simultaneously, using a multilevel study design. Thus, an ecological effect (eg, the effect of income inequality on health) can be estimated, while trying to avoid the sociologistic fallacy by including the relevant individual-level covariates.

Multilevel studies are complex – both statistically (Bryk and Radenbush 1992; Burton et al 1998; Goldstein 1995) and with regard to epidemiological issues (defining an ecological effect, and sources of error that may bias any such ecological effect) (Blakely and Woodward 2000b). For example, determining the relevant individual-level covariates to include as confounders (as opposed to mediating variables) is not straightforward (Blakely and Woodward 2000b).

Examples of multilevel studies that examine the association of ecological socioeconomic variables on health, over and above the association of individual-level socioeconomic factors, include:

- the association of neighbourhood socioeconomic position on mortality in the Alameda Study (Haan et al 1987; Yen and Kaplan 1999)
- the association of area-based measures of socioeconomic position with healthy lifestyles (or 'risk behaviours') (Duncan et al 1993; Duncan et al 1996; Duncan et al 1998; Duncan et al 1999; Humphreys and Carr-Hill 1991)
- the association of community deprivation with mortality and other health-related life events (Sloggett and Joshi 1994; Sloggett and Joshi 1998)
- the association of area-based measures of socioeconomic position with long-term limiting illness (Shouls et al 1996)
- the association of neighbourhood median income with mortality (Anderson et al 1997)
- the association of neighbour socioeconomic position with coronary heart disease and cardiac risk factors (Diez-Roux et al 1997)
- the association of income inequality with self-rated health in the US (Blakely and Kawachi 2001; Blakely et al 2000a; Blakely et al in press; Kennedy et al 1998; Soobader and LeClere 1999)
- the association of income inequality with mortality (Fiscella and Franks 1997) (Daly et al 1998; Lochner et al 2001)
- the association of social capital with self-rated health in the US (Kawachi et al 1999)
- the association of inequality in political participation with self-rated health in the US (Blakely et al 2001a).

The NZCMS has much potential for multilevel analyses. For example, it will be possible to examine the association of income inequality at the Territorial Authority level on mortality, controlling for personal socioeconomic factors. However, this lies outside the scope of this report. The focus of this work is to describe and present the output of the record linkage, and conduct the key individual-level analyses (eg, the association of individual education with mortality).

4 Sources of error

All epidemiological studies are prone to systematic error; namely, selection bias, confounding and information bias (Rothman and Greenland 1998). These sources of error must be accounted for before causal inference is attempted on the basis of observed associations. A thorough consideration of sources of error in the study of socioeconomic determinant of health is often absent. Appendix B of this report provides a detailed overview of sources of error encountered measuring socioeconomic mortality gradients.

As a brief example here of sources of error, though, consider a finding of a two-fold excess mortality risk among low-income people compared to high-income people. Does this mean that the causal association of income with mortality is of the same magnitude? Almost certainly not, (Blakely and Woodward 2000a) for the following reasons:

- *Confounding.* Some of the association of income with mortality will be confounded by other factors such as education, smoking and so on. (Conversely, it should be noted that controlling for tobacco smoking will to some extent be 'over-control' as smoking is on the causal pathway from socioeconomic position to health.)
- *Health selection.* Poor health prior to death will probably reduce one's earning capacity, and therefore exaggerate the association of low-income with higher mortality. Such a *drift selection* effect undoubtedly occurs, but most international research suggests it does not cause a large amount of bias. Another form of health selection is *differential health selection.* This form of health selection normally arises when measuring the association of occupational class with mortality, and the occupational class measure is based on *current* occupation. As people from lower occupational classes are more likely to be unemployed than people form higher occupational classes, the effect of using current occupation is to underestimated occupational class mortality gradients. (See Appendix B for a more detailed consideration of health selection.)
- *Misclassification bias (or measurement error).* Misclassification bias is a two-edged sword. All variables in empirical research are mismeasured. For the association of one independent variable with a dependent variable (ie, univariate analyses) the usual result is to underestimate the association. This is definitely the case with income and mortality as income is notoriously difficult to measure. On the other hand, mismeasurement of covariates (or confounders) in multivariate analyses will result in residual confounding. The net effect of these two misclassification biases on the observed association of income with health is uncertain.

- *Linkage bias.* In the NZCMS linkage bias is a misclassification bias of the mortality outcome. In most epidemiological studies the outcome is measured reasonably accurately. However, this was not the case in the NZCMS due to an inability to link back approximately a quarter of the mortality records to their census record. Linkage bias (and its consequent effect on observed associations) is a major focus in the NZCMS. A detailed consideration can be found elsewhere (Blakely et al 1999; Blakely et al 2000b) and in Chapter 3 of this report. Extensive sensitivity analyses to adjust for linkage bias are also presented in Appendix C.
- *Selection bias.* Selection bias occurs when the observed association amongst the study population varies to that amongst the target population. As such, this is not a great problem for most variables during univariate analyses on the full cohort in the NZCMS as the sample and target populations being nearly the same – census records and the total New Zealand population. However, most of the analyses in this report are for a restricted cohort of census respondents with non-missing data, and therefore some selection bias is possible. Again, sensitivity analyses about this possibility are presented in Appendix C.

Chapter 2: Methods

Box 2: Overview of Chapter 2

Structure of this chapter

This chapter is structured systematically according to the key tasks of conducting a cohort study:

- Defining the study exposures (in this case socioeconomic factors) and covariates – Section 1.
- Defining the study outcomes (mortality) – Section 2.
- Following-up the cohort over time for the outcome of interest (ie, the record linkage of census and mortality records) – Section 3.
- Analysis of the cohort study data – Section 4.

Exposures and covariates

The study-base was the 1991 census data set. Socioeconomic 'exposures' derived from this data set included small area deprivation, education, labour force status, car access, hosing tenure and household income. Covariates included age, sex, ethnicity, receipt of a sickness benefit and marital status.

Outcome

Mortality in the three years following census night, for people aged 0–74 years on census night.

Follow-up

The cohort was assembled by anonymously and probabilistically linking 1991 census records to mortality records for 1991–94. Automatch® software was used to conduct the record linkage.

Analysis

First, mortality records that were linked to a census record were compared to those that were not linked to determine the bias in the record linkage by demographic and socioeconomic factors (ie, linkage bias). Stratified analysis and log-linear regression methods were used to measure this bias.

Second, the association of socioeconomic factors measured on the 1991 census (education, small area deprivation, occupational class, housing tenure, car access, and income) with all- and cause-specific mortality were determined. These associations were estimated using logistic regression models within four sub-populations: 25–44 year old males; 45–64 year old males; 25–64 year old females; and 45–64 year old females. Numerous sensitivity analyses were conducted to determine the likely impacts of selection bias, health selection, and linkage bias. Multivariate analyses explored the direct and indirect associations of education, car access, income and labour force status with mortality.

1 Socioeconomic exposures and covariates

The 1991 New Zealand Population Census was used as the study-base. All the independent variables were derived from this census data.

A Census of Population and Dwellings occurs every five years in New Zealand, as mandated under Section 23 of the Statistics Act 1975. Each census undergoes extensive questionnaire development, pre-testing and pilot surveys (Department of Statistics 1992a).

The census in New Zealand has a high response rate – although inevitably some people are missed. The actual undercount and how it varies by demographic groups for the 1991 census (and all previous censuses) is unknown. However, a Post Enumeration Survey (PES) was conducted following the 1996 census (Ewing 1997). It is likely that the results from the 1996 Post Enumeration Survey would approximately apply to the 1991 census. The 1996 PES was a random sample of 25,000 people interviewed two weeks after census night. As the PES 'used more tightly controlled collection procedures, and more highly trained and experienced field staff' than the 1996 census itself, the PES was assumed to be a valid tool to estimate the 1996 census undercount (Ewing 1997). Results from the 1996 PES estimated that 1.2% of the New Zealand population were not counted on census night. The estimated undercount also varied by demographic groups:

- 1.4% for males and 1.0% for females
- 1.4%, 2.1%, 0.9%, and 0.6% for 0–14, 15–29, 30–44 and 45+ year olds, respectively
- 2.9%, 3.1%, and 0.8% for Maori, Pacific peoples, and New Zealand Europeans, respectively
- 1.3%, 1.0%, and 0.8% for northern North Island, rest of North Island, and South Island, respectively.

As these undercounts are all small, they are unlikely to cause the census data to be unrepresentative of the total New Zealand population. The age group included for analysis in this report (25–64 year olds) had a less than 1% overall undercount. These undercounts may, however, cause some variation in record linkage success between demographic groups – a small percentage of decedents in 1991 to 1994 would not have completed the 1991 census, and thus fail to have their mortality record linked to their census record.

The general groupings of questions in the 1991 census were:

- activities and voluntary work
- demographics (name and address, sex, age, ethnicity (self-identified and Maori ancestry), country of birth, marital status)
- education
- employment and labour force status
- income and income support
- residence (usual, night of census, five years ago)
- means of travel to work, industry, occupation, and name of employer
- relationship to occupier and living arrangements.

The relevant individual-level socioeconomic variables that could be derived from the 1991 census questionnaire were:

- education
- labour force status
- occupational class

- housing tenure
- car access
- household income.

In addition, small area deprivation was also derived from meshblock geocodes on the census data set. These socioeconomic factors are presented in more detail under subsequent subheadings, followed by the main covariates (sex, age, and ethnic group) and other indicator variables.

1.1 Small area deprivation

The NZDep91 (1991 New Zealand small area deprivation index (Salmond et al 1998)) was developed using 1991 census data. Approximately 20,000 small areas were formed from about 35,000 meshblocks – meshblocks are Statistics New Zealand's (SNZ) smallest geographic unit, with a median population of about 90. Ten variables (proportions of people/households in the small area) that reflected a lack of something were used to create the index, and reflected seven dimensions of deprivation: income, transport, living space, home ownership, employment, qualifications and support. The factor structure of the first two principal components suggested a single underlying construct of deprivation. The resultant NZDep91 index is the weighted (weights in the first principal component) sum of the ten standardised variables for each small area. The distribution of NZDep91 scores is highly skewed – there is much discrimination among the more deprived small areas, but little discrimination among the least deprived small areas.

Small area deprivation was not directly elicited by the census questionnaire, but by later assigning the NZDep91 score to census records by use of the usual residence meshblock code. Decile and quintile categories of NZDep91 were used in the analysis.

Small area deprivation mortality gradients have already been described in New Zealand (Salmond and Crampton 2000). The purpose of analysing deprivation mortality gradients in the NZCMS was therefore not to report new findings, but to:

- measure bias in the record linkage
- measure possible selection biases in the cohort analyses when the cohort has to be restricted to just those records with complete information
- measure possible health selection effects in the cohort analyses when the cohort has to be restricted to just those respondents in the labour force on census night.

1.2 Education

Educational attainment was elicited with two questions on the 1991 census personal questionnaire. The first (Question 16) was 'What is your highest school qualification?', with seven mutually exclusive answers:

- no school qualification
- School Certificate in one or more subjects
- Sixth Form Certificate or University Entrance in one or more subjects
- Higher School Certificate or Higher Leaving Certificate
- University Bursary or Scholarship
- overseas qualification (such as United Kingdom GCE)
- other school qualification (please state).

The second question (Question 17) was 'What educational or job qualifications have you obtained since leaving school?', with 11 answer options (multiple selections permitted):

- no qualifications since leaving school
- still at school
- trade certificate or advanced trade certificate
- nursing certificate or diploma
- New Zealand certificate or diploma
- technicians certificate
- teachers certificate or diploma
- university certificate or diploma below Bachelor level
- Bachelors Degree
- postgraduate degree, certificate or diploma
- other qualifications (such as ACA, local polytechnic certificate or diploma) (please state).

From these two questions SNZ derived three separate measures of educational attainment: highest school qualification, tertiary qualifications, and highest gained qualification (ie, highest school or tertiary). Note that the census questions allow a measure of educational attainment, not academic ability or performance.

As many people had no tertiary qualification, the variable 'tertiary qualifications' was not suitable for ranking individuals by education in the NZCMS. A selection between highest school qualification and highest gained qualification at both school and tertiary institutes was required. Highest school qualification had some initial appeal in that everyone is exposed to a similar schooling experience. However, a substantial number of people had an overseas school qualification that was unable to be ranked compared to the New Zealand school qualifications. Moreover, while many of the 'other' school qualifications would have been equivalent to school certificate, or alternative technical qualifications, the ranking was not clear. An empirical advantage of the 'highest gained qualification' variable regarding these two categories was that many people with 'other' or 'overseas' school qualifications obtained a higher tertiary qualification that was more readily ranked.

A further disadvantage of 'highest school qualification', compared to 'highest gained qualification', was that the group with no school qualifications was large (45.7% of 25–64 year old males, and 42.4% of 25–64 year old females). This group with no school qualifications was also heterogeneous with regard to subsequent technical and trade training. Thus, the 'highest gained education' variable categorised many of those individuals with no school qualification to a trade or technical tertiary qualification, reducing the size of the groups with no qualifications for the 'highest gained education' variable (32.3% of 25–64 year old males, and 36.6% of 25–64 year old females).

Finally, the 'highest gained education' variable was more comparable to educational attainment variables used in international research.

Table 5: Categories of the highest gained education variable used in the NZCMS

Categories	%	Description
1 Graduate and postgraduate	7.9	Postgraduate degrees (eg, Masters and PhD) and diplomas, and Bachelors degrees.
2 Undergraduate, technical, and teaching	13.4	Undergraduate certificates and diplomas, technician certificates, teachers certificate and diploma, nursing certificate and diploma.
3 Trade certificates, other tertiary	21.4	Trade certificates, other tertiary.
4 11–12 years of school	8.1	University bursary, university scholarship, higher-school leaving certificate, sixth form certificate, and university entrance.
5 10 years of school	12.0	School certificate.
6 Other school qualification	2.7	Other New Zealand school qualification, and overseas school qualification.
7 Nil	34.5	No qualifications.

Note: Percentages are of the combined 25–64 year old male and female population at their usual residence on census night, where the dwelling was a private dwelling population.

The categories of this educational variable are shown in Table 5. The ordering of the variables is approximately from highest to lowest educational attainment. An underlying assumption in the ranking was that any post-school qualification was higher than any school-based qualification.

1.3 Labour force status

Labour force status was elicited with a sequence of questions on the personal questionnaire. The first screening question (Question 21) was 'Do you work in a job, business, farm or profession?', with answers of yes or no. Those answering no were then directed to three further questions:

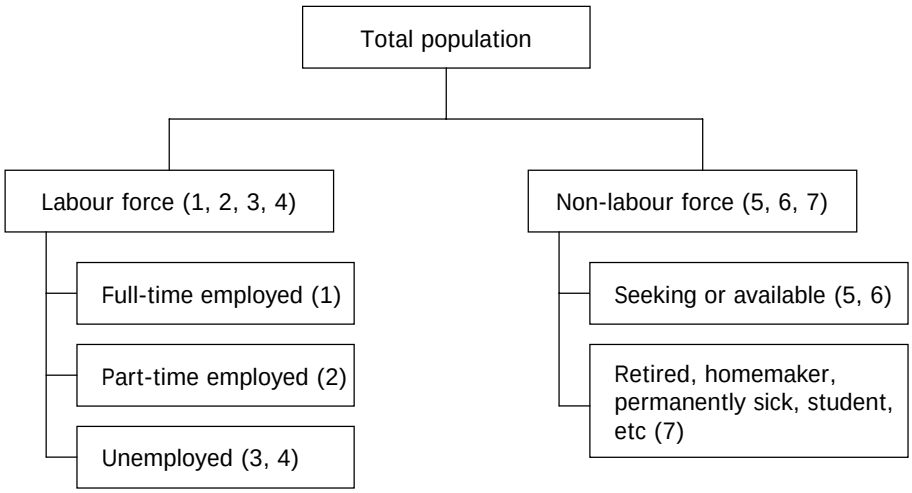
- 'Did you look for paid work in the last four weeks?', with three possible answers:
  - no
  - yes – looked for full-time work (full-time work is 30 or more hours per week)
  - yes – looked for part-time work (part-time work is less than 30 hours per week).
- 'What methods did you use to look for paid work?', with five possible answers (multiple answers allowed):
  - looked at job advertisements in newspapers
  - contacted the Department of Labour's New Zealand Employment Service
  - wrote, phoned or applied in person to an employer
  - contacted friends or relatives for help in finding a job
  - other methods (such as contacted a private employment agency, took steps to set up own business).
- 'If a job had been available, would you have started last week?', with answers of yes and no.

On the basis of these three questions, and a question of those in paid work of the number of hours they worked, SNZ assigned all people to one of the seven following categories of labour force status:

1 employed full time
2 employed part time
3 unemployed and actively seeking full-time work
4 unemployed and actively seeking part-time work
5 not working, seeking work but not available for work
6 not working, available for work but not seeking work
7 not working, not seeking work nor available for work.

Several levels of aggregation were used for labour force status in the NZCMS (Figure 4). The highest level was simply to dichotomise the population into categories 1 to 4 and categories 5 to 7 listed above. These two categories are considered by internationally accepted definitions to be the 'labour force' and the 'non-labour force' (Department of Statistics 1992a). (This report will usually refer to the former as the '*active* labour force' and the latter as the '*non-active* labour force' to further reduce the possibility for confusion.) *Note that the unemployed were included in the active labour force, not the non-active labour force.* At the second level of aggregation, five labour force status groups were identified: full-time employed, part-time employed, and unemployed within the labour force, and 'seeking or available' and 'retired, homemaker, permanently sick, student, etc' among the non-labour force. The unemployed and 'seeking or available' subsumed categories 3 and 4, and 5 and 6 list above, respectively – they were relatively small categories with little reason to differentiate between them. Third, some of the multivariate cohort analyses specifically sought to assess the role of confounding/mediation of the association of unemployment with particular causes of death (eg, suicide). For these specific analyses, three categories of labour force status were used: the unemployed, employed and non-labour force.

Figure 4: Grouping of labour force status used in the NZCMS



Note: Numbers in parentheses refer to the seven categories listed above for labour force status.

Unfortunately, the New Zealand Census does not specifically determine whether people were out of the labour force for health reasons.

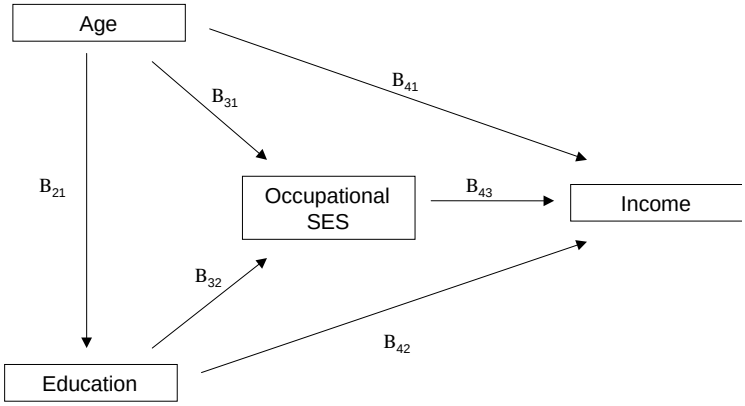
1.4 Occupational class

Information about each individual's occupation was elicited by the questions 'In your main job: (a) What is your occupation? (b) What tasks or duties do you spend the most time on?'. On the basis of the handwritten answers to these questions, coders within SNZ assigned both a NZSCO68 (1968 New Zealand Standard Classification of Occupations) and NZSCO90 (1990) code to each employed individual. (The 1991 census was a transition year from the 1968 to 1990 NZSCO codes, so both codes were assigned.) Occupational class was assigned to each individual using the NZSCO codes with two different measures of occupational status: the New Zealand Socioeconomic Index (NZSEI) and the Elley-Irving scale.

1.4.1 NZSEI occupational class

The NZSEI was developed using 1991 census data (Davis et al 1997; Davis et al 1999b). The NZSEI is premised on the proposition used by Ganzeboom et al (Ganzeboom et al 1992) in their development of the International Socioeconomic Index of Occupational Status, 'that there exists a fundamental relationship between cultural capital or resources (education) and access to material rewards (income), and that this relationship is mediated through the occupational structure' (Davis et al 1997, p.19). Figure 5 below is the path model used in the construction of the NZSEI, with regression coefficients accompanying each arrow. Statistically, one way to fulfil the preceding proposition is to minimise the regression coefficient (B_{42}) directly linking education and income in the path model shown in Figure 5. In effect, this causes the contribution of education (human capital) to income (material rewards) to be channelled as much as possible through occupational socioeconomic position – an indirect causal path. The NZSEI was calculated by an alternating least squares linear regression algorithm that minimised B_{42}. At each iteration, new values of occupational socioeconomic position were assigned to each of the 97 NZSCO90 minor groups. All variables were expressed as standardised continuous variables ('years of education' for the education variable). The output of this process was a scaled occupational socioeconomic position score between 10 and 90 for each of the 97 NZSCO90 minor groups.

Figure 5: Representation of the NZSEI path model



Source: Davis et al, 1997, p.20.

Davis et al (1997) proposed aggregating the NZSCO codes into six occupational classes on the basis of the calculated occupational socioeconomic position score (Table 6 below). Davis et al stated that their division into these six occupational classes was a starting point only – the categorisation has been modified in the NZCMS (Table 6).

Table 6: Alternative classifications of 'occupational class' from NZSEI scores

Class	Range of NZSEI scores (% of 20–69 year old population 1991 census †)				% of 15–64 year old males by Elley-Irving
	Davis et al, 1997		NZCMS modification		Class, 1986 census ‡
1	75–90	5.8%	70–90	10.1%	6.4
2	60–75	17.4%	60–70	13.1%	12.1
3	50–60	20.6%	No change		23.3
4	40–50	22.6%	No change		27.9
5	30–40	16.3%	No change		21.0
6	10–30	17.3%	10–30	8.2%*	9.3
Farmers #	–		22.4, 25.1	9.1%	–

† Derived from Appendix C of Davis et al (1997), giving similar but not identical results to that shown in Table 3.8 of Davis et al (Davis et al 1997).

‡ Taken from Pearce et al 1991 (Pearce et al 1991).

NZSCO90 code 611(market farmers and crop growers) with an NZSEI score of 22.4; NZSCO90 code 612 (market oriented animal producers) with an NZSEI score of 25.1.

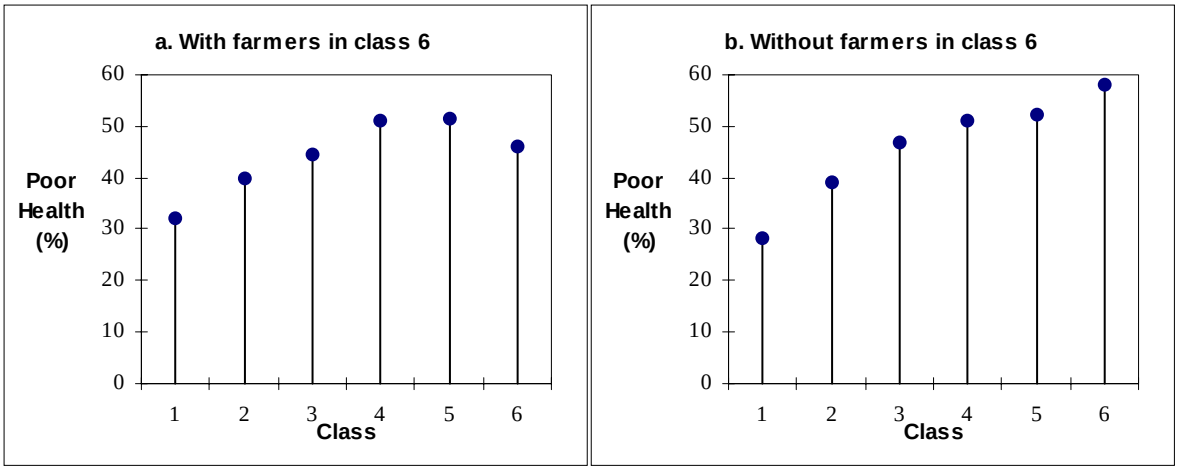
* Excluding farmers.

There were three reasons why the modified NZSEI occupational class classification was preferred in the NZCMS. First, close inspection of NZSCO90 codes allocated by Davis et al to occupational class 2 disclosed a bimodal distribution of NZSEI scores – NZSEI scores were either between 60 and 65, or between 70 and 75. Occupations with a score between 60 and 65 were nursing and midwifery professionals, administrative associate professionals, power generating plant operators, protective service workers, railway engine drivers, primary and early childhood teaching, archivists and librarians, safety and health inspectors, special-interest organisation administrators, physical science and engineering technicians, government associate professionals, and general managers. Occupations with a score between 70 and 75 were business professional, architects and engineers, ship and aircraft controllers, and computing professionals. These latter occupations arguably had more in common for socioeconomic position with the occupations with NZSEI scores above 75 (social and related science professionals, secondary teaching, other teaching professionals, tertiary teaching, life science professionals, physicists and chemists, senior government administrators, mathematicians and statisticians, legislators, legal professionals, senior business administrators and health professionals) than the former occupations with a NZSEI score between 60 and 65.

Second, there are problems with ranking the socioeconomic position of farmers that probably argue for the removal of farmers from the ordinal ranking of occupational classes to be considered as a separate 'special' group. The NZSCO90 classification has just two minor codes for farmers: NZSCO90 code 611(market farmers and crop growers) with an NZSEI score of 22.4; NZSCO90 code 612 (market-oriented animal producers) with an NZSEI score of 25.1. Within these two groups, there is no distinction between farm owners, farm managers, farm supervisors, and farm workers, and hence a wide distribution of socioeconomic position (Davis et al 1997). Moreover, farm owners are also self-employed, a group known to have a low declared income compared to similar status occupations in New Zealand (Clemance 1985). Occupational class indices used in Europe commonly separate farmers into a separate occupational class (eg, (Kunst et al 1998c)). As generated by the path model used to develop the NZSEI score (Davis et al 1997), the two farming occupation codes both fell within occupational class 6 and 4. Furthermore,

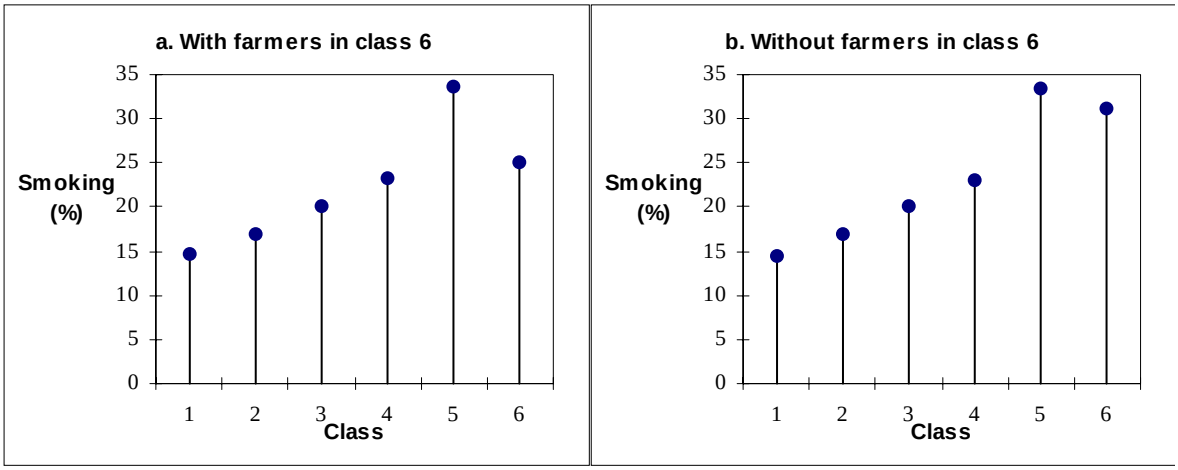
farmers comprised more than half of the proposed occupational class 6 classified by Davis et al. Such 'misclassification' is likely to result in underestimation of adverse health effects for occupational class 6: results for poor self-reported health and smoking prevalence in Figure 6 and Figure 7 demonstrate this effect.

Figure 6: Poor self-reported health in the 1992–93 Household Health Survey by NZSEI occupational class, using the classification proposed by Davis et al, but excluding farmers from occupational class 6 in Figure b



Source: Previously unpublished results, kindly forwarded by Keith McLeod, Statistics New Zealand, 1998.

Figure 7: Smoking prevalence in the 1992–93 Household Health Survey by NZSEI occupational class, using the classification proposed by Davis et al, but excluding farmers from occupational class 6 in Figure b



Source: Previously unpublished results, kindly forwarded by Keith McLeod, Statistics New Zealand, 1998.

A similar problem probably also exists for the Armed Forces (Davis et al 1997), but they are relatively few in number (0.6% of employed people), were allocated to a 'middle' occupational class (class 3), and so were not separated out. It would be possible to derive NZSEI scores at a lower level of aggregation than the 97 minor occupation groups, that is either for the 260 unit groups, the 563 groups, or some hybrid combination. Such further work may be worthwhile in terms of precision obtained. For example, farmers may be successfully separated and differentiated.

Third, the percentage distribution for the modified classification (excluding farmers) is more symmetric than that proposed by Davis et al, with roughly comparable percentages in occupational class pairs 3 and 4, 2 and 5, and 1 and 6. The percentage of people in occupational class 1 increases from 5.8% to 10.1%, making a more robust comparison group if highest occupational class is used as the reference category. Also, the distribution (excluding farmers) is closer to that for the Elley-Irving scale.

1.4.2 Elley-Irving occupational class

The Elley-Irving socioeconomic index was originally developed using 1966 census data, then revised using 1971 census data (Elley and Irving 1976). The Elley-Irving index per se has not been revised since – the NZSEI should be seen as superseding the Elley-Irving. The Elley-Irving index was developed using education and income, but in a more basic manner than the NZSEI. The median educational attainment (school) and income (personal) were simply determined for each specific occupation, standardised to a common scale, and combined to give a summary score. On the basis of this summary score, six occupational classes were formed.

Davis et al (1997, pp.49–50) compared the Elley-Irving and NZSEI occupational classes, having first derived NZSEI scores for NZSCO68 codes (rather than the NZSCO90 codes). While there was an obvious association of the two indices, there was some overlap in the NZSEI scores between each Elley-Irving occupational class – especially in the middle occupational classes. Davis et al suggest that differences may have arisen due to the different underlying statistical models, and changes in the population between the time the scales were derived. In the NZCMS the NZSEI occupational class was derived from NZSCO90 codes and the Elley-Irving from NZSCO68 codes. As the NZSCO90 codes are a *skills*-based classification, and the NZSCO68 are a *task*-based classification, this generated another potential discrepancy.

1.5 Housing tenure

Housing tenure was measured at the household level, and was elicited on the dwelling questionnaire with the prompt 'Do the occupants ...', with options of:

- own this dwelling with a mortgage?
- own this dwelling without a mortgage?
- occupy this dwelling rent free?
- rent or lease this dwelling?

If the last option was selected, the occupier was directed to a further question 'Who is it rented or leased from?', with answers of:

- private person
- real estate agency or business organisation
- Housing Corporation
- other government department or corporation, ministry or state owned enterprise
- local authority.

On the basis of these answers, each household in the NZCMS was assigned to one of five categories:

1 owned with a mortgage
2 owned without a mortgage
3 private rental
4 public rental
5 rental, but not further specified.

Those with free tenancy, or with unspecified tenancy, were assigned as missing values.

1.6 Car access

Car access is measured at the household level, and was elicited with the question 'How many motor vehicles are available for private use by persons in this dwelling?', with answers of none, one, two, three, four, and five or more. The question was on the dwelling questionnaire, completed by the 'occupier'. Occupiers were instructed not to include motor cycles or scooters. In the NZCMS, the answers were aggregated to three categories: none, one, two or more. Unlike income, the number of cars was not equivalised for the number of occupants.

1.7 Equivalised household income

Equivalised total household income was used as the measure of income for each individual in the NZCMS. Thus, each individual in the same household was assigned the same income. The derivation of this variable is justified and described under the following subheadings: household; total household income; equivalisation; and concluding comments.

1.7.1 Household

A household in the 1991 census 'refers to a group of persons, whether related or not, who live together and who normally consume at least one meal together daily or at least share the same cooking facilities' (Department of Statistics 1992b). Households were used as the unit for income in the NZCMS, rather than individuals, as:

> *'It is within households that income and wealth are pooled, and consumption and savings decisions made. An individual's standard of living is determined not by their income, but by the resources available as a whole to the household in which that individual lives. Thus in examining differences in the standard of living, the natural unit of study is the household.'* (Statistics New Zealand 1999, p.51)

In the majority of instances (approximately three-quarters) households were comprised of one family. In the remainder of instances households were single person households, or non-family multi-person households. The use of households, as opposed to families, includes a greater proportion of census respondents.

1.7.2 Total household income

Personal income was determined by question 15 on the 1991 census: 'What will be your total income, including income support, before tax for the year ended 31 March 1991?', with 13 possible options:

1 Nil income or loss
2 $2500 or less per year
3 $2501 to $5000 per year
4 $5001 to $7500 per year
5 $7501 to $10,000 per year
6 $10,001 to $15,000 per year
7 $15,001 to $20,000 per year
8 $20,001 to $25,000 per year
9 $25,001 to $30,000 per year
10 $30,001 to $40,000 per year
11 $40,001 to $50,000 per year
12 $50,001 to $70,000 per year
13 $70,001 and over per year.

Further instructions on the questionnaire were 'Include income from all sources', with prompts of: wages, salary and commission, business or farming income (less expenses), income support, accident compensation weekly payments, interest, dividends, rent, superannuation and pension payments.

A total household income was *unable* to be calculated by SNZ in three circumstances (Department of Statistics 1992a):

- when there were no persons in the dwelling aged 15 years and over
- when there were persons aged 15 years and over absent from their usual residence on census night, *and* the combined income of the persons aged 15 years and over present in the dwelling was less than $70,000. (If the combined income of the persons present was greater than $70,000, then regardless of the absent person's/persons' income the household had already met the requirements to be in the top household income category)
- when there were persons aged 15 years and over in the dwelling who did not specify their personal income, *and* the combined income of the remaining persons aged 15 years and over present in the dwelling was less than $70,000.

In the remaining instances, the total household income was calculated by summing the midpoint income for each individual in the household (ie, $22,500 if category $20,001–$25,000 selected, and $70,001 if top category selected). This total household income was then recategorised to the same 13 categories as the personal income variable.

1.7.3 Equivalisation

There are economies of scale in households. A family of four does not require four times the income of a single person to enjoy the same standard of living, as, for example, rental costs increase marginally rather than directly proportional to the number of people. Equivalisation is a procedure to adjust household incomes such that they are comparable between households of different size and composition. The revised Jensen Index is a commonly used method to equivalise incomes in New Zealand (Jensen 1988). The Jensen Index assigns equivalences to any combination of adults and children (age less than 18) using the formula:

$$I_{a,c} = (a + wc)^u / 2^u$$

where:

$I_{a,c}$ = income equivalence of a family household of 'a' adults and 'c' children
w = the weight for a child compared to an adult
u = the power parameter to be estimated.

Based on a review of a range of international equivalence scales, and the original Jensen Index (Jensen 1978), Jensen selected two anchor points (values of $I_{a,c}$) for the 1988 revised Index (Jensen 1988): 0.65 for a household of one adult and no children, and 1.75 for a household of two adults and four children. With these two anchor points, values were estimated for w (0.73) and u (0.62). Table 7 shows the thus calculated equivalences for any combination of 1–4 adults and 0–6 children. For example, to have an equivalent standard of living as a two adult household with an income of $40,000, a household of two adults and two children needs and income of 1.41 × $40,000 = $56,400.

Table 7: The revised Jensen Index

Number of adults	Number of children						
	0	1	2	3	4	5	6
1	0.65	0.91	1.14	1.34	1.52	1.69	1.85
2	1.00	1.21	1.41	1.58	1.75	1.91	2.06
3	1.29	1.47	1.65	1.81	1.96	2.11	2.25
4	1.54	1.71	1.87	2.02	2.16	2.30	2.44

The revised Jensen Index is not an internationally recognised scale. Perhaps the most commonly used equivalisation procedure internationally (termed the Luxembourg Income Study (LIS) (Atkinson et al 1995) scale here) is simply to divide the household income by the square root of the number of household members (regardless of age). Note that a household income calculated according to the revised Jensen Index is always greater than that calculated according to the LIS scale. For example, a fixed total household income of $40,000 for a two adult household corresponds to an equivalised household income of $40,000 by the Jensen Index, but an equivalised household income of only $28,284 according to the LIS scale. What is important, though, is whether this relative difference between the Jensen Index and LIS scale varies according to household composition. Table 8 shows the ratio of the equivalised household income calculated according to the Jensen Index compared to that according to the LIS scale. Broadly, the relative size of the Jensen to LIS equivalised household income is unchanged with varying numbers of children, but decreases with increasing numbers of adults. Put another way, equivalisation according to the revised Jensen Index 'assumes' less economies of scale for each extra adult in the household than does the LIS scale.

Table 8: Ratio of revised Jensen Index to LIS scale equivalised household income

Number of adults	Number of children						
	0	1	2	3	4	5	6
1	1.54	1.55	1.52	1.50	1.47	1.45	1.43
2	1.41	1.43	1.42	1.41	1.40	1.39	1.38
3	1.35	1.36	1.36	1.36	1.35	1.34	1.34
4	1.30	1.31	1.31	1.31	1.31	1.31	1.30

In the NZCMS, an equivalised household income was calculated by dividing the midpoint of each total household income category by the appropriate value of the revised Jensen Index. For households in the top total household income category ($70,001 and over per year), the 'mid-point' was taken as $99,300 based on data from the 1991 Household Economic Survey. When there were more than four adults or more than six children in a household, the value of the Index selected was that for four adults and six children, respectively. When all household members were younger than 18 (but at least one member older than 15), one child was reclassified as an adult.

1.7.4 Concluding comments

As is apparent from the above description, there are a number of possible sources of misclassification. First, the census only uses one question to elicit personal income (rather than multiple questions of salary/wages, interest, dividends, rent, income support, etc), requiring each individual to take a 'best-guess' at their total income. Second, the measure is not disposable income (allowing for taxes and transfers), but gross income. Third, midpoints are assigned to income categories at two stages: to personal income categories to calculate total household income, and to household income to calculate the equivalised household income. Fourth, while '$70,001 plus per year' in 1991 was a reasonably high top *personal* income category, the same may not be true for *household* income, resulting in some loss of discrimination among high-income households. Finally, there is an inevitable arbitrariness of the selected equivalisation procedure – in this instance, the revised Jensen Index.

1.8 Covariates: sex, age, and ethnicity

1.8.1 Sex

All analyses were conducted separately by sex.

1.8.2 Age

Most analyses were conducted separately for 25–44 year and 45–64 year olds. Within those 20–year age groups, a categorical variable of five-year age groups (eg, 25–29, 30–34, 35–39, and 40–44 years) was included in all analyses.

1.8.3 Ethnicity

The focus of this report was on socioeconomic mortality gradients, not ethnic inequalities in health or the variation in the socioeconomic mortality gradients between ethnic groups. Both are extremely important issues for research, and the NZCMS offers huge potential to unravel some of the overlap between ethnicity and socioeconomic position. Ethnicity and socioeconomic position will be a substantive focus in the NZCMS, but that focus is beyond the scope of this report.

> The analyses in this report neither measure ethnic inequalities in mortality in New Zealand, nor measure socioeconomic mortality gradients within ethnic groups.

However, ethnicity is strongly associated with mortality in New Zealand. As the distribution of socioeconomic factors is correlated with ethnicity, and ethnicity is not on the causal chain between socioeconomic position and mortality, ethnicity is a likely confounder of the association of socioeconomic position with mortality (Rothman and Greenland 1998). Accordingly, ethnicity is included as a covariate in all analyses. (Moreover, controlling for ethnicity minimised linkage bias.)

Ethnicity was elicited in the 1991 census with the question 'Which ethnic group do you belong to?' *(tick the box or boxes which apply to you)*, with answers of: New Zealand European, New Zealand Maori, Samoan, Cook Island Maori, Tongan, Niuean, Chinese, Indian, and Other (please state). SNZ then derived a hierarchical classification:

- New Zealand Maori ethnic group (New Zealand Maori as one of the self-identified ethnic groups)
- Pacific Island Group (any Pacific Island group as one of the self-identified ethnic groups, but not where the individual also self-identified as New Zealand Maori)
- Non-Maori non-Pacific.

In the NZCMS, those not specifying an ethnic group were classified as non-Maori non-Pacific.

1.9 Other indicator variables

In addition to the socioeconomic exposures and covariates of interest, there were a number of other indicator variables that, for example, allowed an estimation of the likely health selection.

1.9.1 Usual residence

This variable identified all the individuals at their usual residence on census night. As the household socioeconomic exposures can only be derived for individuals at their usual residence on census night, most cohort analyses excluded individuals not at their usual residence on census night.

1.9.2 Dwelling type

This variable identified whether the census night dwelling was a private (eg, separate house, caravan, flat) or non-private dwelling (eg, motel, rest home, hospital, prison). Household socioeconomic exposures were also not available for non-private dwellings, so most cohort analyses excluded non-private dwellings.

1.9.3 Sickness beneficiary

Receipt of income support was elicited on the personal questionnaire, with one possible answer being 'Sickness or Invalid's Benefit'. One way to determine the socioeconomic mortality gradient unbiased by health selection is to conduct analyses only upon those individuals that were healthy at the outset of the cohort study. Excluding individuals who received a sickness or invalids benefit allowed a test of possible health selection. However, two opposing biases make this test of uncertain accuracy. First, the sickness benefit is only provided to people below a certain income. This bias would mean that the mortality risk for poor healthy people would be correctly determined, but it would be overestimated for rich healthy people due to residual inclusion of rich sick at the outset. This first bias would cause an underestimate of the socioeconomic mortality gradient. Second, not everyone who is unhealthy (for the purposes of causing a health selection effect) would apply for a sickness benefit. The net effect of these two biases was uncertain.

1.9.4 Hospitalisation before the 1991 census

For the decedents it was possible to determine who had been hospitalised between about 1988 and census night. (Most hospitals began using the NHI number in 1988, allowing death events to be linked to hospital events after than time.) An ideal test of health selection would have been to exclude all members of the *cohort* with a hospitalisation in the three years up to census night, thus allowing analyses on a relatively healthy population at initiation of the cohort study. However, to do so would require linking all people with a hospitalisation event (not just deaths) to the census – an unrealistic exercise. Thus, analysis in the NZCMS could only determine the difference between including all deaths versus including only deaths where the decedent had not been hospitalised in the three years preceding census night.

2 Mortality outcome

Mortality in the three years following census night (5 March 1991) was determined by linking mortality records to census records (see next section for record linkage methods). The mortality records were obtained from New Zealand Health Information Services (NZHIS). In addition to analyses of the association of socioeconomic factors with all-cause mortality, analyses of the association of socioeconomic factors with cause-specific mortality were also conducted. Groupings of cause of death were based on that in the Global Burden on Disease study (Murray and Lopez 1996), with consideration of modifications proposed by Tobias and Christie (1998), as shown in Table 9.

Table 9: ICD codes for groupings of cause-specific deaths used in the NZCMS

Cause of death	ICD codes
Cancer	140–209
Colorectal	153–154
Lung	162
Breast	174
Prostate	185
Cardiovascular disease	410–414, 390–409, 415–459
IHD	410–414
Cerebrovascular	430–438
Infection and pneumonia	001–139, 320–323, 390–392, 460–466, 480–487, 590, 595, 614–616, 680–686, 711, 771
Respiratory	470–478, 490–519
COPD	490–492, 495–496
Unintentional injury	800–949
Road traffic crash	810–825
Other unintentional	800–809, 826–949
Suicide	950–959, 980–989
Homicide, intentional injury	960–979, 990–999
Other	Remaining ICD codes

3 Record linkage

A detailed description of the methods (and results) of the record linkage of mortality and census records can be found in a Technical Report (Blakely et al 1999). Only a summary of the record linkage is included in this chapter, emphasising an epidemiological perspective. In particular, record linkage is presented in terms of a *misclassification bias of the mortality outcome.*

Record linkage has a language of its own. As such, a glossary of terms is provided on page 253. The first use of each term that appears in this glossary is in **bold**.

3.1 Probabilistic record linkage methods

Newcombe (1988) and Baldwin et al (1987) provide good basic introductions to probabilistic record linkage methods (Baldwin et al 1987; Newcombe 1988). The method has been used in the United States to link mortality records to the Current Populations Survey database (Rogot et al 1986). In the last decade, Jaro has developed an advanced software package, Automatch®, for probabilistic record linkage (Jaro 1995; MatchWare Technologies 1998). Automatch® was the software package used in this research.

Humans searching two files for the same individual intuitively do two things. First, they look for agreement or disagreement on variables common on both files (matching variables). Second, they assign varying importance to different variables. For example, a match on a social security number (or some other unique identifier) just about guarantees the records in the two separate files are for the same person. But a match on sex adds only a small amount of discriminatory information. Probabilistic record linkage formalises these intuitive processes, using probability ratios and taking advantage of the processing capacity of computers.

The object of record linkage is to find **matches** between **records** from two (or more) **files**, where a match consists of two records from different files *for the same person*. In the NZCMS, the comparison files were mortality and census records. To achieve this objective, **pairs** of mortality and census records are compared by the variables common to both files – the **matching variables**. It is not possible in any record linkage project to determine exactly which comparison pairs are (correct) matches and non-matches. Rather, pairs are categorised as **links** or **non-links**. It is intended that the majority (hopefully, the vast majority) of links are matches (**true links**), and few matches are falsely assigned as non-links (**false non-links**). A two-by-two table of link/non-link status by match-non-match status is shown in Table 10.

Table 10: Two by two table of link/non-link status by the match/non-match status for comparison pairs in a record linkage project

	Match	**Non-match**
Link	True links (or true positives)	False links (or false positives)
Non-link	False non-links (or false negatives)	True non-links (or true negatives)

At the heart of probabilistic record linkage are **agreement frequency ratios** and **disagreement frequency ratios**, determined by the ***m* probabilities** and ***u* probabilities**. Consider the variable day of birth (dd), and the value 9. The *m* probability is the probability among the linked records that when dd is 9 for one of the records (eg, mortality), dd is also 9 for the record from the other file (eg, census). The linked records are, to the best of one's knowledge, correctly matched. But it is neither necessary that each linked record be a correct match, nor necessary that all correct matches be included among the linked records, for the estimation of *m* probabilities to be accurate provided the numbers of false links and false non-links are small. Note that there is a bootstrap problem here – the *m* probability is calculated among the linked records, but before a set of linked records can be created in a probabilistic record linkage process we need to have *m* probabilities. To get around this problem *m* probabilities are initially specified by the operator on the basis of a best guess, and in subsequent iterations of the record linkage the *m* probabilities are updated on the basis of the last set of linked records.

The *u* probability is similar to the *m* probability, except it applies to the non-linked records: the *u* probability is the probability among the non-linked records that when dd is 9 for one of the records (eg, mortality), dd is also 9 for the record from the other file (eg, census). That is, the *u* probability is the probability that for any random comparison pair the given matching variable agrees. This is closely approximated by the frequency of each specific value of each matching variable in the two files.

Having obtained *m* and *u* probabilities, agreement and disagreement frequency ratios (or odds) are next calculated for each possible comparison of matching variables. For example, Table 11 gives frequency ratios for agreement and disagreement on dd. Assume that among the linked records, 95% agree on day of birth – the *m* probability. (The other 5% would disagree principally as a result of coding errors.) Further, assume among the non-linked records 3% agree on day of birth. This 3% is simply estimated by the inverse of the number of possible values of dd, approximately 1/30. That is, among non-links we expect 3% to agree on dd purely by chance. The agreement frequency ratio of 32 to 1 for an observed match on dd is the odds of [the probability of dd agreeing among links] to [the probability of dd agreeing among non-links]. That is, dd is 32 times more likely to agree among links than among non-links. Conversely, the disagreement odds of dd not agreeing among links versus non-links is 1 to 19.

Table 11: Example of agreement and disagreement frequency ratios and weights for comparison by the matching variable 'day of birth'

Comparison outcome	Proportion/frequency		Frequency ratio	Weight
	Links	Non-links		
Agreement	0.95 (m)	0.03 (u)	32/1 (m/u)	4.98 [ln(m/u)/ln(2)] †
Disagreement	0.05 (1–m)	0.97 (1–u)	1/19 (1–m/1–u)	-4.28 [ln(1–m/1–u)/ln(2)] †

† The divisor, ln(2), transforms the natural logarithm to a base 2 logarithm.

Having determined the frequency ratios for each matching variable (and each value of each matching variable), the next step is to calculate the **combined frequency ratio** for any given comparison pair. The combined frequency ratio is the product of the agreement and disagreement frequency ratios for all matching variables, taking the agreement frequency ratio when the matching variable agrees and the disagreement frequency ratio when it disagrees. But the magnitude of the combined frequency ratio quickly becomes very large (for multiple agreements on the matching variables) or very small (for multiple disagreements), and it is easier to use the **combined weight**. The combined weight is the sum of the **agreement** and **disagreement weights**. The agreement or disagreement weight for each matching variable (or value of the matching variable) is the logarithm to base two of the global (or specific) frequency ratio – the formula is given in Table 11. Using logarithms to base two is not necessary, but was the precedent set by researchers involved in pioneering record linkage in the Oxford Record Linkage Study (Baldwin et al 1987). A convenience of using logarithms to base two is that each increase in the weight by one represents a doubling of the overall odds in favour of the comparison pair being a match. The combined weight is then used to allocate, by means of a cut-off, each possible comparison pair to either a set of highly probable pairs (ie, links) or a set of unlikely pairs (ie, non-links).

An additional, and crucial, step in probabilistic record linkage is the **blocking** of records. Blocking involves partitioning the records in both files by a common variable, and then only conducting comparisons of records between files *within these blocks*. For example, two files of 1000 records each could be blocked by age in years, resulting in approximately 10 records in each block in each file. This dramatically reduces the number of comparisons from 1,000,000 without blocking, to $100 \times 10 \times 10 \times = 10{,}000$ with blocking by age ([blocks] × [records in each block in first file] × [records in each block in second file]. Blocking is thus computationally efficient, and (as described subsequently in the Section 3.3) reduces the number of false links.

3.2 Record linkage in the NZCMS

The record linkage in the NZCMS was anonymous (ie, no names or text address variables were available).

3.2.1 Blocking variables

Five geocodes were extracted from the mortality data (one meshblock code, and four census area units (CAU) codes) that together with the census usual residence meshblock code and CAU code made five possible combinations of blocking variables. These five combinations are shown in Table 12. Meshblocks are the smallest administrative geographic unit in New Zealand, with a median number of 96 people in each meshblock. CAUs are aggregates of meshblocks, containing about 2000 people each.

Table 12: Geocode variables ('blocking' variables) used in the record linkage

Mortality	Census †	Comment
Meshblock (SNZ vitals)	Meshblock	*Mortality data.* 90.7% of the mortality records were assigned a meshblock code for their usual residence at time of death, by merging the NZHIS mortality records to the SNZ Vitals file. The SNZ Vitals meshblock was derived from the address on the death registration form. *Census data.* A usual residence meshblock was routinely assigned to all census records by SNZ.
Vitals-CAU	CAU	*Mortality data.* The CAU containing the SNZ Vitals File meshblock for the above 90.7% of mortality records. While a meshblock was not able to be assigned to the remaining 9.3% of mortality records, the vast majority were directly assigned a CAU. *Census data.* The CAU containing the usual residence meshblock. This census CAU was used for blocking with each of the four mortality data CAUs.
NHI–CAU	CAU	*Mortality data.* The NHI file includes an automatically assigned CAU–code on the basis of the text address entered by hospital clerks that maintain the NHI File. Thus it is an independent source of geocode data to the SNZ Vitals.
Post–CAU	CAU	*Mortality data.* Some decedents were hospitalised after census night, but before the hospitalisation associated with the death event. By linking the NHI and NMDS files at NZHIS, the 'post-CAU' code was derived for the stated usual address at the time of this hospitalisation for some mortality records. The rationale was to obtain a CAU-code for a point in time closer to the census than when the decedent died.
Pre–CAU	CAU	*Mortality data.* Some decedents were hospitalised between 1988 (when the NHI file was established) and census night. By linking the NHI and NMDS files at NZHIS, the 'pre-CAU' code was derived for the stated usual address at the time of the last hospitalisation before census night for some mortality records. The rationale was to obtain a CAU-code for a point in time closer to the census than when the decedent died.

† Census geocodes are for the usual residence address on census night.

The multiple possible geocodes for the mortality data arose due to the existence of a NHI file (National Health Index file; usually entered by hospital clerks), the NMDS Death Event file (National Minimum Data set; built up from the SNZ-Vitals file and the death registration form), and possibly one or more NMDS Health Event files (hospitalisations; entered by hospital clerks) for each decedent. These NHI and NMDS files could be linked together by NZHIS using the NHI number recorded on all files. The benefit from these multiple and independent sources of data for the mortality records was that CAU codes for the usual residence at time of death *and* points in time closer to census night could be obtained. These multiple CAU codes increased the chance of correctly linking mortality and census records.

Only one meshblock code, however, was available from the mortality data – that on the SNZ Vitals file. NZHIS does not store the meshblock code on the NMDS Death Event file, despite SNZ deriving a meshblock code for the majority of decedents prior to forwarding the data to NZHIS. However, the meshblock derived from the death registration form was easily retrieved. By combining the death registration office, year and number variables a unique identifier was created for each death. Using this unique identifier, the meshblock codes (and Vitals-CAU codes) were transferred from the SNZ Vitals file to the NZHIS mortality data.

3.2.2 Matching variables

The matching variables common to both mortality and census data were sex, ethnic group, country of birth, and date of birth. The latter was disaggregated to day of birth (dd), month of birth (mm), and year of birth (yy). All of these matching variables, except country of birth, were available from both the NMDS Death Event file and the NHI file for each decedent. Thus, the census sex variable could be compared to both the NMDS Death Event file and NHI file sex. The advantage of this double comparison was that if there was a coding error for sex on one of these two mortality files, it was highly unlikely that a coding error would have occurred on both files. These two sources of demographic data for the record linkage therefore increased the discriminatory power of the record linkage process.

3.2.3 Record linkage strategy

The way that these matching and blocking variables were used in the record linkage is described in detail in the Technical Report (Blakely et al 1999). Briefly, the geocodes were used to 'block' the two files, and comparisons of records by the personal 'matching' variables only occurred when the geocodes agreed. Meshblock, as the variable with the greatest number of values, was used as the blocking variable in the first **pass** of the record linkage. Subsequent passes used CAU codes as the blocking variable. In all, eight passes were used in the record linkage.

The final record linkage strategy – in particular the cut-offs and clerical review – was a balance of maximising the number of links obtained (maximising sensitivity), but minimising the estimated percentage of false links (maximising positive predictive value). Three methods were available to estimate the positive predictive value – two specifically developed for the NZCMS. They are complex and not presented in this report – instead, they are described in detail in the Technical Report (Blakely et al 1999, pp.42–59). Briefly, two of the methods were used to estimate the number of false positive links in this report: the chance method and the duplicate method. The chance method is essentially a method for estimating the number of exact links that would occur purely by chance, using the *u* probabilities. The duplicate method uses the number of duplicate pairs (DA pairs; one mortality record linkage to two or more census records) and combinatorial probabilities to estimate the number of false positive links.

For duplicate links, the highest scoring duplicate link was accepted, or if the scores were tied both (or all) links were discarded.

3.3 Probabilistic record linkage as (mis)classification of the mortality outcome, and the effect of blocking

In this section, an example is used to illustrate record linkage in terms of a screening test or tool to ascertain the mortality outcome.

Assume there are 3 million census records and 40,000 deaths in the census cohort in the given follow-up period – the approximate numbers in the NZCMS. Further, assume that the mortality records available for record linkage were these same 40,000 deaths. Thus there was a correct census record match for all 40,000 mortality records somewhere in the census file. Assume that 35,000 of these 40,000 mortality records were correctly linked to their census record – the true links. That leaves 5000 false non-links that were missed due to, for example, coding errors on either file. These numbers and the total of 40,000 matches (or death events) are shown in the first column of Table 13. The linked/non-linked status is the *observed* status of each comparison pair, and the match/non-match status is the *actual* status of each comparison pair.

Table 13: Two by two table of link/non-link status by the match/non-match status in a hypothetical record linkage example *without* blocking

	Match	**Non-match**	
Link	35,000 (true links)	1,200,000 (false links)	1,235,000
Non-link	5000 (false non-links)	1.1999876×10^{11} (true non-links)	$1.19998765 \times 10^{11}$
	40,000	1.1999996×10^{11}	1.2×10^{11}

The total number of possible comparison pairs for these 40,000 mortality records and 3 million census records is $40{,}000 \times 3{,}000{,}000 = 1.2 \times 10^{11}$. Subtracting the 40,000 matches leaves 1.1999996×10^{11} non-matches. Assume that the matching variables were as in the NZCMS – dd, mm, yy, sex, ethnicity, country of birth – and that (for simplicity in this example) exact agreement was required on each of these variables. (Probabilistic record linkage allows for some disagreement on some matching variables. See the Technical Report for details of how this was specified in the NZCMS (Blakely et al 1999).) The probability of any randomly selected non-match pair agreeing on all these variables is the product of the *u* probabilities, in this case approximately $1/30 \times 1/12 \times 1/60 \times 1/2 \times 2/3 \times 2/3 = 0.00001$. (Here, 1/60 is an approximate 'average' *u* probability for yy, and 2/3 is an approximate 'average' *u* probability for ethnicity and country of birth.) Thus, $0.00001 \times 1.1999996 \times 10^{11} = 1{,}199{,}999.6 \approx 1.2$ million of the non-matches would be categorised as links – false links in the top right cell of the two-by-two Table 13.

In total, therefore, there would be 1,235,000 links in this record linkage example – the top row total in Table 6. The percentage of these links that were true links is only 2.8% (35,000/1,235,000), an appallingly low positive predictive value in screening terms that will cause ruinous bias in any cohort study! However, **blocking** can substantially improve the positive predictive value. Assume that there exists a geocode with 30,000 values that blocks the census and mortality files into blocks containing, on average, 100 census records and 1.33 mortality records in each block (ie, similar to the meshblock used in the NZCMS). With any blocking strategy, the cost is that if the true matches disagree on the blocking variable, that match is missed in the record linkage and becomes a false non-link – a problem known as 'skipping'. Assume in our example that skipping reduces the number of true links to 30,000 as shown in Table 14. Thus the introduction of a blocking variable reduces the sensitivity of the record linkage in this example from 87.5% (35,000/40,000) to 75% (30,000/40,000).

Table 14: Two by two table of link/non-link status by the match/non-match status in a hypothetical record linkage example *with* blocking

	Match	**Non-match**	
Link	30,000 (true links)	40 (false links)	30,040
Non-link	10,000 (false non-links)	3,989,960 (true non-links)	3,999,960
	40,000	3,990,000	4,030,000

In addition to the gain in computing efficiency brought about by blocking due to fewer necessary comparisons, there is also a substantive gain in the PPV. In this example, there are now $30{,}000 \times 1.33 \times 100 = 3{,}990{,}000$ possible comparison pairs (ie, [number of blocks] × [average number of mortality records in each block] × [average number of census records in each block]). As above, 0.00001 of these comparison pairs will be categorised as links due to a purely chance agreement on the matching variables, ie, $0.00001 \times 3{,}990{,}000 = 39.9 \approx 40$. The PPV now in this example is a very respectable 99.9% (30,000/30,040). The cost of obtaining this improvement in PPV was a drop in sensitivity. Note that the specificity remained unchanged between Tables 13 and 14 ($1.1999796 \times 10^{11}/1.1999996 \times 10^{11} = 3{,}989{,}960/3{,}990{,}000 = 0.9999$).

This above example is simplistic. For example, in Table 13 there are only 40,000 mortality records yet over 2 million links (ie, each mortality record is linked to 50 census records on average). Further, the above example does not allow for varying block sizes, duplicate links (ie, one mortality (census) record linked to two or more census (mortality) records), and partial disagreements on the matching variables. These issues are considered in detail in the Technical Report (Blakely et al 1999). However, the above example demonstrates:

- how the record linkage process is analogous to a screening test for the mortality outcome
- and how blocking increases the PPV by essentially increasing the prevalence of matches in the population of comparison pairs.

4 Data analysis

The data analysis was in two parts: analysis of the linkage bias, and the cohort analysis.

4.1 Linkage bias

In this section, methods are described which are used to quantify the **linkage bias**, where linkage bias is defined as *the biases by demographic and socioeconomic factors in the proportion of mortality records linked to a census record*. (A more detailed description of the methods used in the analysis of linkage bias can be found in the Technical Report (Blakely et al 1999).) A description of linkage bias as a misclassification bias of the mortality outcome in included in Appendix B. It would have been possible to use correction formulas (eg, Copeland et al 1977) to quantify the likely amount of bias of the risk ratios observed in the cohort analyses due to misclassification of the mortality outcome. To do so would first require estimating the sensitivity and specificity of the record linkage for the mortality outcome, including by demographic strata. However, it was simpler, more direct, and probably more accurate, to directly estimate the linkage bias by socioeconomic position (described in the remainder of this section), and then use these estimates to adjust the risk ratios subsequently determined in the cohort analyses. *This adjustment constitutes the sensitivity analyses in the NZCMS of the impact of misclassification bias of the mortality outcome on the observed risk ratios in the cohort analyses.*

Also, note that the linkage bias was determined by comparing the mortality records linked to a census record to those mortality records unlinked to a census record. As such, the mortality records submitted to the record linkage were the total population of analysis. This population approximates, but would not have been exactly the same as, the actual deaths in the census cohort (see Appendix B).

4.1.1 Univariate and stratified analyses

The variation in the proportion of mortality records linked to a census record was determined by demographic factors (sex, age, ethnic group (NHI file), and time between census night and death) and socioeconomic factors (small area deprivation (NZDep91) and occupational class (NZSEI)). Simple categorical methods were used. Estimates of precision (ie, confidence intervals) are not reported for the univariate and stratified analysis – due to the large sample size most differences are statistically significant, and the size of the difference is of greater importance.

4.1.2 Regression analyses

Regression analyses were conducted to determine the 'independent' linkage bias due to socioeconomic factors, controlling for demographic factors. The rationale was that the cohort analyses would be conducted within demographic strata (eg, the association of income and mortality among 45–64 year old females – not the association of income and mortality for all people simultaneously). Thus, *the objective was to quantify the residual linkage bias by socioeconomic position within each demographic stratum or group.*

Such analyses could be conducted simply by stratification rather than regression modelling, deriving the 'actual' linkage rate by strata. However, this approach was limited as:

- the SNZ protocol is that all absolute cell sizes must be random rounded to a multiple of three. Thus the observed percentage linked in small cells would have been inaccurate for sparse strata
- the assigned ethnic group on the mortality data is not equivalent to that on census data (Blakely et al 2001b) – regression modelling to 'smooth' out estimates by strata of ethnic group may be preferable to using actual results from health data ethnic group strata
- census records with either a census night dwelling of 'private hospital', 'public hospital', or 'rest-home', or simply a non-private census night dwelling, were excluded from most of the cohort analyses. But it was not possible to conduct an analysis of bias on a similar restricted set of mortality records as the dwelling type was not recorded in the mortality file. Regression modelling to smooth out estimates by strata may, again, be preferable compared to using actual health data stratum.

Consequently, regression modelling was used to quantify the linkage bias by socioeconomic position. Two different regression-modelling strategies were used. The first strategy attempted to quantify the linkage bias by small area deprivation for all mortality records simultaneously and by occupational class for males and females separately. This is the method described in detail and used extensively in the Technical Report (Blakely et al 1999) – it will only be briefly described here. The second strategy involved simply modelling the linkage bias by small area deprivation and occupational class within each of the four demographic groups used for the cohort analyses in this report – 25–44 and 45–64 year old males, and 25–44 and 45–64 year old females.

The justification of the first modelling strategy was to summarise the linkage bias as much as possible, achieving the most stable estimates possible. However, for the purposes of this report having linkage bias results for each of the four demographic groups allowed direct adjustment for linkage bias in the cohort analyses conducted in this report. Hence, both strategies are described and reported in this report here.

The regression analyses used a generalised linear model with a log link (referred to as log-linear hereafter), conducted in SAS version 6.12. The regression model was:

$$R(x) = \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)$$

where:

$R(x)$ is the average risk of being linked to census record given covariates x

$x_1, x_2, ..., x_n$ are the covariates or interaction products (eg, sex, age group)

$\beta_1, \beta_2, ..., \beta_n$ are the coefficients

α is the intercept.

The log-linear risk models were fitted with a binomial error term, Pearson estimation methods, and quasi-likelihood estimates of the standard error.

A log-linear risk model was preferred over a logistic model, for the following reason. The 'risk' of being linked to a census record was comparatively high. Therefore, the odds ratio of linkage for a stratum with, say, 80% linked compared to a stratum with 60% linked ([0.8/0.2]/[0.6/0.4] = 2.67) is quite different from the risk ratio (0.8/0.6 = 1.33). It is the risk ratio that is of interest, as that will be the 'adjustment' required to the observed risk ratios (approximated by the odds ratios from logistic regression) in the cohort analysis.

4.1.2.1 Regression analyses for demographic strata combined

Regression analyses to determine the linkage bias due to small area deprivation (NZDep91) were conducted for all deaths combined, and separately by sex for 25–74 year olds for occupational class (NZSEI).

Model selection was conducted by using a combination of prior information, and a backward elimination strategy. Prior information was of two main types. First, the univariate and stratified analyses were used as a starting point to consider likely interaction and confounding. Second, each subsequent log-linear model built on previous models. For example, initial models were developed for the effect of sex, age, and ethnic group on the probability of being linked to a census record. This model was then used as the baseline to examine whether the socioeconomic variable of particular interest (eg, NZDep91 score) had any effect over and above the demographic covariates of sex, age, and ethnic group.

The general backward elimination strategy was as follows. As a first step, all main effects and first order interaction products were included in an initial model. The Wald Type III Chi-Square statistic and p value for each first order interaction product were then inspected. If the p value was statistically significant ($p < 0.05$) the interaction product was retained in the second step, otherwise the interaction product was discarded. (If three or more main effects were involved in two or more overlapping and statistically significant first order interaction products (eg, [age]×[sex], and [age]×[ethnic group]), then models with second order interaction products were explored.) In the second step, a model was fitted with the statistically significant first order interaction products from the first step, and all main effects. In the final step, the remaining statistically significant first order interaction terms and any main effects not involved in an interaction term were retained. Note that in the second and final step, main effects were retained even if not statistically significant – all main effects modelled (sex, age, ethnic group, time period between census and death, NZDep91 score, and NZSEI occupational class) were either the exposure of interest, or covariates that had strong prior justification for inclusion.

Often, the iterative estimation of the parameters of the log-linear model failed to converge when a number of main effects and their interaction products were included. In these instances, greater use was made of prior information, and exploratory logistic modelling to determine which interaction products to retain.

4.1.2.2 Regression analyses separately within the four sex by age groups

The previous regression strategy assumes that, unless rejected by statistical tests, the linkage bias by socioeconomic factors is homogenous across demographic strata. To complement this statistical approach, log-linear regression models were also conducted within each of the four sex by age groups used in the cohort analyses in this report (ie, 25–44 and 45–64 year old males, and 25–44 and 45–64 year old females). Further, these analyses excluded deaths in the first six months after census night, thus being more representative of the actual follow-up period for the cohort study. The models simply included dummy variables for five-year age groups and ethnicity (Maori, Pacific) as covariates, and a categorical variable for NZDep91 or NZSEI. No interaction terms were modelled as the analysis was already being conducted within groups of age by sex.

4.2 Cohort analyses

The cohort analyses refer to the association of socioeconomic exposures with the mortality outcome among the 25–64 year old 1991 census cohort. Analyses for 0–24 year olds, and 65–74 year olds are beyond the scope of this report.

The main epidemiological effect measure of interest was the *risk ratio*. The reason for using a *risk* rather than *rate* ratio was that the cohort was of short duration follow-up with a relatively rare outcome (death). Therefore an analysis using person–time in the denominator would give essentially the same result as that using counts in the denominator. (Estimating person–time in the NZCMS would also be done with some error – for example, it was not known which individuals completing a census record subsequently emigrated.) Moreover, it would have entailed a large (and probably not feasible) effort to have assembled the necessary data for a person–time analysis.

The cohort study-base was variously restricted for different analyses. Two main cohorts were used for the analyses: the *full cohort* and the *restricted cohort*. The full cohort consisted of all New Zealand resident and non-absent census records. (The SNZ census data set is hierarchically organised with individuals within dwellings within meshblocks. Some of the individual records are for people away from their usual residence on census night, with values for sex and age only. These records are called 'absentee records'. For each absentee record, one assumes that there is another fully completed census record elsewhere in the census data set within another dwelling for the same person. Thus, restricting the census data set to all non-absentee records should, theoretically, include one complete individual record for every person completing a census form.) The restricted cohort excluded those census records with missing data for any one of household income, car access, education, and labour force status. The major reason for a census record being excluded was missing data for household income. To have a valid household income required that all usual residents of the household aged 15 years and older be at home on census night, unless the total income of those adults actually at home on census night exceeded $70,000.

4.2.1 Univariate analyses

The 'univariate' association of each socioeconomic exposure (small area deprivation, education, labour force status, occupational class, housing tenure, car access, and household income) with all-cause mortality was determined, separately for 25–44 and 45–64 year old males, and for 25–44 and 45–64 year old females. The main univariate results were presented for the restricted cohort. For each of the univariate analyses, the initial summary results are presented *excluding* all census records with a linked death in the first six months. That is, a priori, deaths in the first six months are assumed to be most at risk of health selection effects and were therefore excluded.

Results were presented *crude* using simple categorical analyses, and *adjusted* for age (five-year age groups) and ethnicity using logistic regression. Logistic regression analyses were conducted in SAS versions 6.12 and 8.0. All independent variables were specified as dummy categorical variables. As the outcome was relatively rare, the odds ratio closely approximated the risk ratio. Consequently, the terms risk ratio and odds ratio are used interchangeably in this report.

4.2.2 Sensitivity analyses of the univariate results

A range of sensitivity analyses were conducted for the univariate results to assess the likely impact of selection bias, health selection and misclassification of the mortality outcome. (The majority of these sensitivity analyses are reported in Appendix C of this report.) Age and sex adjusted logistic regression analyses were used for the majority of all-cause mortality sensitivity analyses. However, crude risk ratios were used for some of the sensitivity analyses (particularly cause-specific mortality analyses) to reduce the number of logistic regression models required. The use of crude data also allowed the use of 'floating' risk ratios (ie, whole population risk as reference risk). These floating risk ratios were sometimes easier for the interpretation of trends compared to logistic regression where a greater than expected change in the reference category risk may distort all other odds ratios. At several stages age and sex-adjusted logistic regression analyses were run as checks to ensure that trends emerging from the crude risk ratios and age and sex-adjusted odds ratios were consistent.

4.2.2.1 Selection bias

The census includes nearly all the New Zealand population. (The 1996 Post Enumeration Survey suggested a less than 1% undercount of the 1996 census among 25–64 year olds (Ewing 1997).) However, using the restricted cohort for the univariate (and latter multivariate) analyses may introduce some selection bias. A simple test of selection bias was, therefore, to repeat the univariate analyses where possible for the fullest cohort available for the given socioeconomic factor. For example, nearly the entire full cohort had a non-missing highest qualification.

4.2.2.2 Health selection

Many studies control for health selection effects by discarding the first few years of outcome data – a luxury not available in the NZCMS where deaths were only linked up to three years following the census. Therefore, considerable effort was expended in this report trying to determine how much bias may have arisen due to health selection in the NZCMS. The amount of health selection effect will vary for each socioeconomic measure (and for each health outcome measure). It should be most marked for labour force status, whereby people in poor health move into the non-active labour force. Household income is likely to be affected by health selection effects, but less so than individual income if the unhealthy person in the household is not the only or main income earner. For people older than 25, health selection effects should have no (or very little) effect on gradients in mortality by educational status, as education is by then a (nearly) fixed characteristic. Likewise, one would expect little health selection for gradients of mortality by small area deprivation in the short term.

Several strategies were available to examine the possible impact of health selection.

Strategy 1: Plot mortality risks over time by level of the given socioeconomic factor

If health selection effects exist, they would be expected to diminish over time (Fox and Goldblatt 1982; Fox et al 1985). Thus changes in the observed mortality risks were plotted over time for each level of a given socioeconomic factor. Depending on the socioeconomic factor, health selection would predict a convergence or divergence of the mortality risks over time. For example, if the income mortality gradient was biased by *drift* health selection, then one would expect to see a high mortality risk among low-income people initially, but it would fall over time as those with a low income consequent on their poor health either died or returned to good health. Conversely, the mortality risk among the high-income people would initially be very low as (under the health selection argument) one could only have a high income if in good health. But over time, the mortality risk among those with a high income would increase as some people succumbed to poor health. Thus, under the drift health selection hypothesis, one would expect to see a convergence of mortality risks by strata of income over time among the whole cohort (ie, a funnel plot). Alternatively, if the income mortality gradient was affected by *differential* health selection, one would expect a divergence over time in the mortality risks by income level among the active labour force. That is, as people from lower socioeconomic groups were supposedly more likely to be forced out of the labour force by poor health than people of higher socioeconomic groups, one would expect to see low mortality risk initially among low-income people in the labour force. But over time, the mortality risk for low-income people in the labour force would rise towards its 'background' risk. Conversely, the mortality risk for high-income people would not rise as much over time. Thus, the mortality risk plots by income-level for the labour force

would tend to diverge over time. Given that the NZCMS was a short-duration follow-up study, the plotting of mortality risks over time was only useful for investigating short-term health selection effects.

Strategy 2: Exclude recipients of sickness benefits

If *drift* health selection exists, one might expect the income mortality gradient among those people not receiving a sickness benefit (ie, excluding some of the 'unhealthy') to be reduced in magnitude. However, as sickness benefits are more likely to be received by poorer people for any given level of sickness (receipt of the benefit was means tested), restricting the cohort to non-recipients of sickness benefits may actually overadjust the income–mortality gradient for drift health selection. Assuming that both education and small area deprivation are unaffected by short-duration health selection drift, the baseline changes were established (ie, the 'overadjustment changes') to the education and deprivation mortality gradients after excluding sickness beneficiaries. For evidence that drift health selection was affecting the income mortality gradient, the reduction in the income gradient after excluding sickness beneficiaries had to be substantially larger than that observed for both education and deprivation.

Strategy 3: Exclude death outcomes that were hospitalised between 1988 and census night

It was possible to identify all deaths with a hospitalisation between 1988 and 1991 census night, and exclude them from the analyses. The rationale for this exclusion was that deaths among people with no hospitalisation might represent a subset of more 'acute' deaths, and thereby include deaths that were less prone to health selection effects on census night. There were four limitations to this approach. First, it would have been preferable to exclude all those census respondents who had been hospitalised in the three years prior to census night, thus restricting the cohort-base – not just restricting the outcomes. However, this would have required linking all hospitalisations for the 1991–94 period to the 1991 census – a massive task with small marginal gain to the NZCMS. Second, and like receipt of a sickness benefit, hospitalisation in the three years prior to census night was an imperfect marker of ill health. Not all people with poor health on census night would have been admitted to a (public) hospital in the preceding three years, and not everyone admitted would have been ill in the sense of being prone to health selection (eg, hospitalisation for injury). Third, and also similar to the receipt of a sickness benefit, hospitalisation for a given disease or health-state probably varies by socioeconomic position. For example, an unwell patient in a hospital's emergency department may be more likely to be admitted if they do not have good social support at home, and social support may in turn be correlated with socioeconomic position (Berkman and Glass 2000). Moreover, private hospitalisations were not captured, which are also highly correlated with socioeconomic position. Fourth, the likelihood of hospitalisation prior to census night would vary by cause of death.

Nevertheless, excluding pre-hospitalised deaths was still a useful test of health selection. To maximise the validity of this sensitivity analyses, baseline analyses were again conducted for the socioeconomic factors where short-term health selection was assumed not to be important (small area deprivation, and highest qualification). As with receipt of a sickness benefit, for the exclusion of pre-hospitalised deaths for the income analyses to suggest health selection the change in the gradient had to be substantially more than that for the baseline deprivation and education analyses. Finally, greater attention was paid to cause-specific mortality.

4.2.2.3 Misclassification of the mortality outcome

The results from the linkage bias analyses were used to adjust the results from the cohort analyses as a sensitivity analysis for misclassification bias of the mortality outcome. For example, assume that decedents of lower socioeconomic position were 10% less likely to be linked to a census record than decedents of higher socioeconomic position. That is, the risk ratio from the linkage bias would have been 0.90 for the lower compared to the higher socioeconomic position groups. Further, assume that the risk ratio (approximated by the odds ratio) for the association of lower socioeconomic position compared to higher socioeconomic position was measured as 2.0 in the cohort analysis. The adjusted risk ratio would therefore be 2.0/0.9 = 2.22. Note that the observed risk ratio was an underestimate by 10%, but that the *excess* risk ratio (ie, RR – 1.0) was underestimated by 18% (0.22/1.22).

4.2.3 Univariate analyses excluding the non-active labour force

Multivariate analyses that control for labour force status were problematic to interpret. Therefore, as a prelude to these multivariate analyses, univariate analyses were conducted that simply excluded the non-active labour force. The difference in effect size (1 minus the odds ratio) for the analyses excluding the non-active labour and the analyses including all labour force categories was then determined. A change in the effect size of a given socioeconomic factor with mortality after excluding the non-active labour force may represent one or more of six processes:

- variation (*effect modification*) of the association within the active labour force compared to all labour force categories, due to:
  - differential health selection
  - factors other than differential health selection
- controlling for *confounding* of the association by labour force status, where labour force status is a marker of:
  - health status (ie, drift health selection)
  - factors other than health status
- explaining that component of the association that is *mediated* by labour force status, where labour force status is a proxy for:
  - health status (ie, health status as an intermediary variable between socioeconomic position and mortality)
  - factors other than health status.

Thus, interpretation of the univariate results excluding the non-active labour force required cross-reference to results for sensitivity analyses of health selection, and results for other socioeconomic factors where one would expect a different balance of these processes to be working.

4.3 Multivariate analyses

Unless stated otherwise, all logistic regression results for the univariate analyses actually adjusted for age and ethnicity. Thus, the use of the term univariate was one of convenience. The term 'multivariate' in this report is reserved for those logistic regression analyses that include two or more socioeconomic factors (eg, education and income).

The multivariate models were specified and interpreted as best as possible with reference to the causal model shown in Figure 3, although problems with labour force status limited the full application of this simple framework. All independent variables were specified as dummy categorical variables. Categories were aggregated further compared to those for the univariate analyses to avoid problems with sparse data. The logistic regression models were conducted in SAS version 6.12 and 8.0.

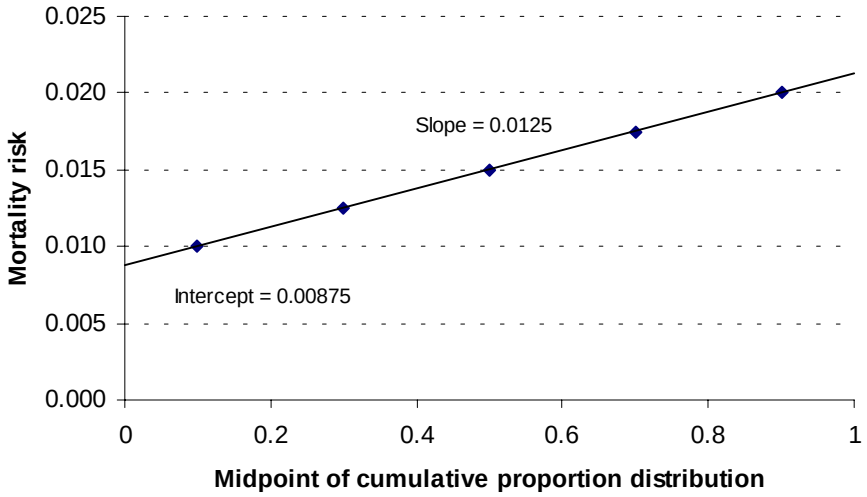
4.3.1 Relative index of inequality

To assist interpretation of the multivariate analyses, relative indices of inequality (RIIs) were calculated for income and education (Mackenbach and Kunst 1997). Conceptually, the RII is the relative risk of mortality for the person with the lowest socioeconomic position compared to the person with the highest socioeconomic position. As such, it assumes that socioeconomic factors simply *rank* individuals within a society. For example, assume that income was measured as a quintile variable, and that the mortality risk in each quintile (from highest to lowest income) was 0.010, 0.0125, 0.015, 0.0175, and 0.020. That is, the relative risk for people living in the poorest income quintile (0.020) compared to those in the richest income quintile (0.010) was 2.0. This relative risk compares two groups, and is thus essentially a comparison of the average mortality risk about the 10th percentile compared to the 90th percentile of the population. For two reasons, it may be useful to consider the mortality risk for the poorest (zero percentile) compared to the richest (100th percentile):

- intrinsically, it is of interest to estimate the gradient right across the socioeconomic hierarchy, rather than just comparing groups
- often the groupings vary in size between studies. For example, in one study half the population may be in the reference category of education, whereas in another study only 10% of the population may be in the reference category. Thus, comparing the effect measures between these two studies is confounded by differences in group sizes.

Figure 8 below demonstrates how the RII is calculated, using the hypothetical mortality risks given above. The income quintiles are ranked from the highest socioeconomic group to the lowest. Each quintile comprises 20% of the population. Thus, the richest quintile is plotted at 0.1 on the cumulative proportionate distribution of the population (x axis), with a mortality risk of 0.01. The next richest income quintile will have an x-axis value of 0.3 (0.2 for the previous quintile, plus half of the current quintile), and a y-axis value of 0.0125, and so on.

Figure 8: Hypothetical example of mortality risk by income to demonstrate the calculation of the RII



Having plotted these x-y points, the slope and intercept can be calculated. In this simple example, the slope is 0.0125 and the intercept is 0.00875. The RII is then (0.0125 + 0.00875)/0.00875 = 2.43. That is, the poorest person has a mortality risk that is 2.43 times that of the richest person, somewhat more than the relative risk of 2.0 derived from simply comparing the lowest and highest income quintiles.

In this report, RIIs are calculated for age and sex adjusted analyses, and for multivariate analyses. Therefore, odds ratios were used as the y-variable. (Given that mortality was a relatively rare outcome, the odds ratios are directly proportional to the adjusted mortality risks.) Weighted linear regression was conducted of the odds ratios on the midpoints for each income and education group on the cumulative proportion distribution. Weights were the inverse of the variance of the crude mortality risk for each income or education group (ie, p(1-p)/n (Kirkwood 1988), where n was the census count and p the mortality risk/proportion). (A more statistically 'correct' way to calculate the RIIs would have been to rerun all the logistic regressions specifying income and education as continuous variables ranging from zero to one.)

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).

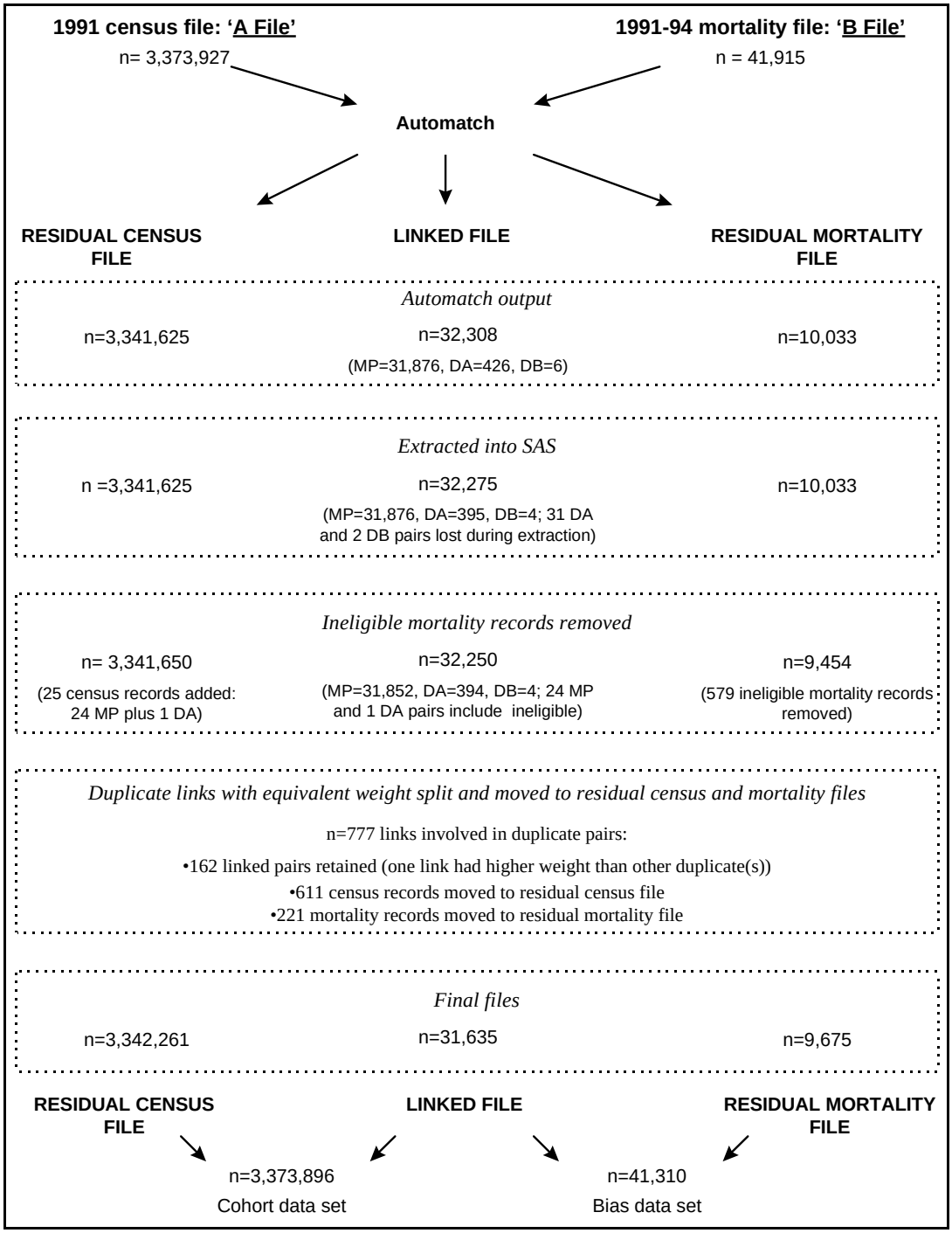
The record linkage involved submitting the mortality and census files to Automatch®, and deriving three output files:

- 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

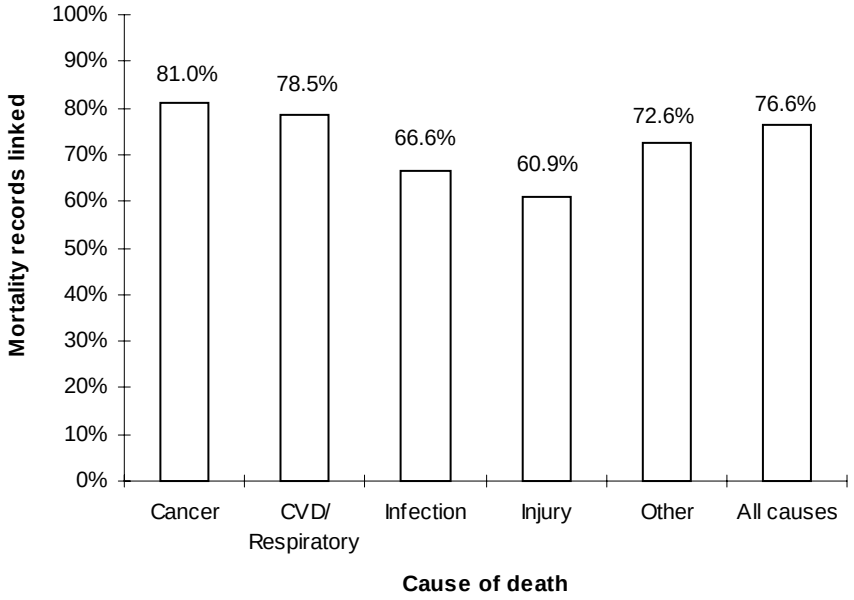
Pass number and blocking variable(s)	Links (% of eligible mortality records) †	
Meshblock	25,311	61.27%
Vitals-CAU, and month of birth	3473	8.41%
Post-CAU, and month of birth	1117	2.70%
Pre-CAU, and month of birth	340	0.82%
NHI-CAU, and month of birth	416	1.01%
Meshblock	429	1.04%
Meshblock	91	0.22%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

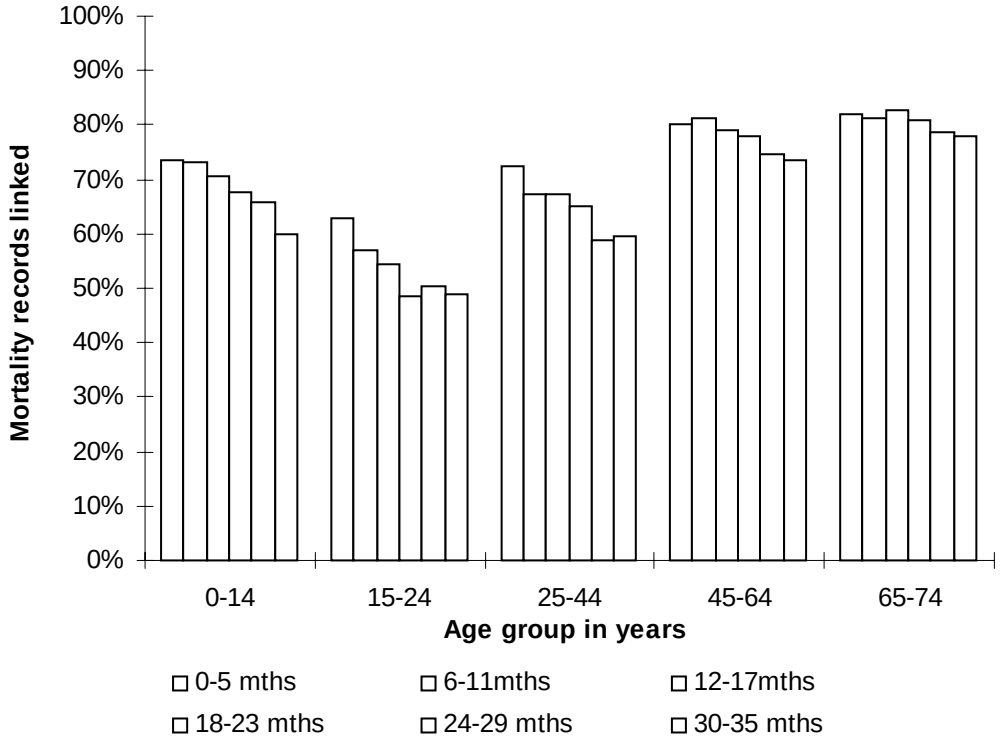
3 Linkage bias

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

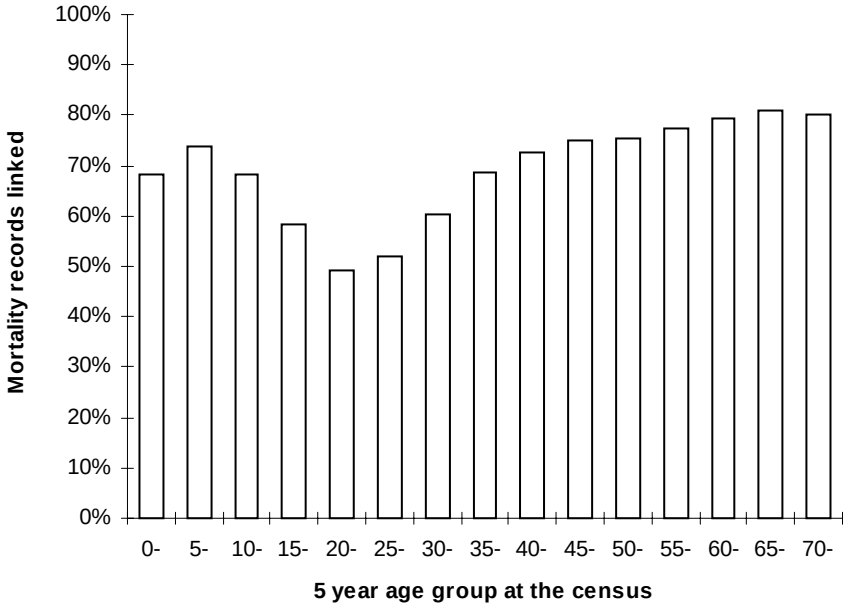
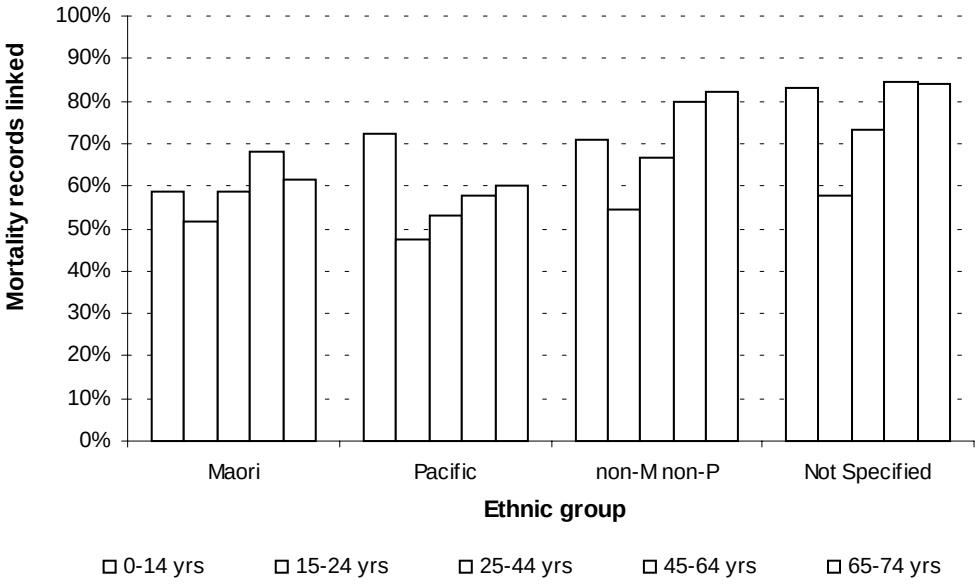


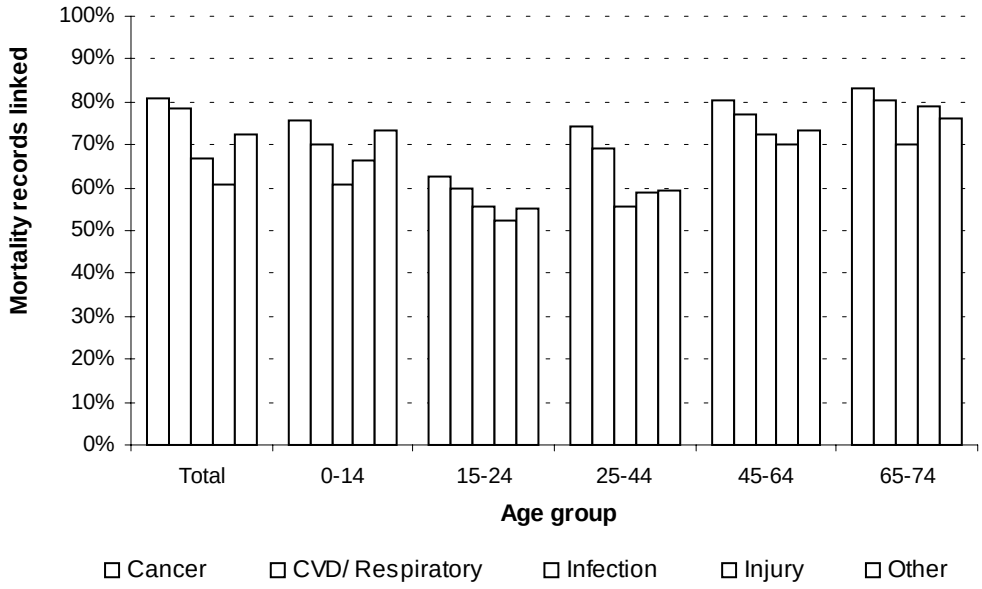
Figure 13: Percentage of mortality records linked to a census record by age group by ethnic group



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.

Chapter 4: Results – Cohort Analyses

Box 4: Overview of Chapter 4

Structure of this chapter

The distribution of the census cohort and deaths by demographic factors is presented in Section 1. Sections 2 to 8 present univariate analyses for each of the socioeconomic factors (deprivation, education, car access, housing tenure, labour force status, occupational class, and household income). Detailed sensitivity analyses of bias about these univariate results are presented in Appendix C, but a summary is provided in this chapter. Multivariate analyses, including labour force status, income, education and car access, conclude the chapter (Section 9).

Univariate analyses

Strong socioeconomic all-cause mortality gradients were found for all ranked socioeconomic factors. Moreover, there were strong associations with each *specific cause of death*, except non-lung cancers. Elevated mortality risks were also found for the unemployed and non-active labour force compared to the employed.

Bias

Some *selection bias* affecting the restricted cohort results was suggested by the NZDep91 sensitivity analyses, but not by the educational sensitivity analyses. The net effect of both *selection and linkage bias* was to probably cause underestimates of the all-cause mortality gradient among females and 25–44 year old males for analyses based on the restricted cohort – but probably had little net effect for 45–64 year old males.

Health selection was without doubt one cause for the elevated mortality among the non-active labour force, but it did not appear to be a cause of the elevated mortality among the unemployed. Unfortunately, a range of sensitivity analyses of the possible effect of health selection on income–mortality gradients did not produce a consistent picture. The most likely conclusion is that there was some health selection bias affecting the income mortality gradients, mostly for males and for cancer deaths (probably also cardiovascular deaths, possibly suicide deaths, but not injury deaths).

Occupational class results based on *current* occupation in the NZCMS (and this report) almost certainly underestimated the gradient had it been possible to measure it using *usual* or *last* occupation. This underestimate was probably due to both confounding by current labour force status and differential health selection. Some important numerator–denominator biases that would bias unlinked occupational class analyses were disclosed.

Multivariate analyses

The two to three fold increased risk of *suicide* death among the *unemployed* (and the non-active labour force) was not due to confounding/mediation by other socioeconomic factors. This was in contrast to other causes of death where unemployment appeared to have little independent association with mortality.

Some of the association of *education* with each cause of death appeared to be due to mediation/confounding by material factors (income and car access), but, generally, a strong independent association of education with each cause of mortality remained.

While the association of *income* with all-cause and cause-specific mortality was very strong in univariate analyses, and largely independent of education and car access, it tended to reduce to the null following control for labour force status for cancer, cardiovascular disease and suicide deaths – but not injury deaths. Some, and perhaps most, of these reductions to the null must be attributed to health selection biasing the income–mortality association. Alternative explanations include confounding by labour force status, and over-control whereby non-labour force status is a proxy for poor health – a precursor to death from cancer, cardiovascular disease and possibly suicide.

Car access was generally a strong independent predictor of mortality.

1 Structure of 1991 census cohort

The 1991 census cohort included 3,373,896 individuals, and 31,635 of these individuals were anonymously and probabilistically linked to a mortality record. As only deaths for people aged 0–74 on census night (as per NZHIS data) were submitted to the record linkage, the cohort was reduced to the 3,220,170 people aged 0–74 on census data. This restriction to 0–74 year olds by census DOB meant that 48 of the linked records were discarded – because the record linkage was probabilistic, it was not necessary to have exactly the same age to be included in the linked data set.

Table 25 presents a breakdown of the 0–74 year old cohort by age, residency status (usual or non-usual residence on census night), and dwelling type (private, hospital or rest home, and other non-private). On census night 3,011,085 (93.5%) of all respondents aged 0–74 were at their usual residence and living in a private dwelling.

Table 25: Numbers of census respondents in 1991 census cohort by age, dwelling type, and residence on census night (column percentages)

Dwelling type	Age group in years						Total
	0–14	15–24	25–44	45–64	65–74	75+	0–74
Individuals at usual residence on census night							
Private	744,711 (95.0%)	498,348 (89.6%)	969,750 (94.3%)	589,548 (94.2%)	208,728 (92.4%)	125,637 (81.7%)	3,011,085 (93.5%)
Hospital or rest home	291 (0.0%)	765 (0.1%)	2220 (0.2%)	2244 (0.4%)	3594 (1.6%)	20,169 (13.1%)	9114 (0.3%)
Other non-private	2472 (0.3%)	12,849 (2.3%)	8829 (0.9%)	4095 (0.7%)	831 (0.4%)	582 (0.4%)	29,076 (0.9%)
Individuals not at usual residence on census night							
Private	18,879 (2.4%)	27,234 (4.9%)	28,422 (2.8%)	16,044 (2.6%)	7602 (3.4%)	3912 (2.5%)	98,181 (3.0%)
Hospital or rest home	1452 (0.2%)	1026 (0.2%)	2319 (0.2%)	1677 (0.3%)	1470 (0.7%)	2355 (1.5%)	7944 (0.2%)
Other non-private	16,023 (2.0%)	15,849 (2.9%)	17,208 (1.7%)	11,958 (1.9%)	3717 (1.6%)	1068 (0.7%)	64,755 (2.0%)
Age total	**783,828**	**556,071**	**1,028,748**	**625,566**	**225,942**	**153,759**	**3,220,170**

Note: Cell values are randomly rounded to a multiple of three as per SNZ protocol.

Table 26 shows the number of census records and linked mortality records by sex and age group, for the 'full' 0–74 year old cohort and sequential restrictions thereafter. As with all count data extracted from SNZ, cell values are randomly rounded to a multiple of three above or below the exact value – therefore cell values may not exactly agree between Table 25 and Table 26 and the sum of cell values may not equal the given totals. The third row in each sex by cohort/death panel in Table 26 excludes all deaths in the first six months after census night – the majority of cohort analyses exclude these observations to mitigate against health selection. The fourth row further excludes census records that have missing data for either household income or car access. The majority of these exclusions were due to individuals in the household aged older than 15 years being absent on census night, making it impossible to calculate the total household income. The final row for each sex by cohort/death panel in Table 26 also excludes individuals with missing values for highest educational qualification obtained, individuals still at school, or individuals with missing values for labour force status. This final row of exclusions applies only to 25–74 year olds, as education is unlikely to be completed for younger people and labour force status is complicated by ongoing education.

Approximately 77% to 80% of the full census cohort, or approximately 85% of the census cohort at their usual and private residence on census night, had complete data for the four main socioeconomic factors of interest in this report – highest qualification, household income, car access and labour force status (Table 26). 3060 of the 31,590 census records from the full cohort with a linked mortality record (9.7%) were either not at their usual residence or not in private dwelling on census night. Of these 3060 records, 1750 (57.2%) were in a hospital or rest home on census night. The largest reduction in linked mortality records occurred when all deaths within the first six months of follow-up were discarded (third row of sex by linked death panels in Table 26).

Table 26: Number of census respondents and linked deaths in cohort by sex and age group, for sequential restrictions of the cohort

Cohort number and linked deaths, by sex and sequential restrictions	Age group in years					Total	% of full census †
	0–14	15–24	25–44	45–64	65–74		
Male cohort							
Full census	400,563	280,983	506,970	313,623	103,584	1,605,723	100.0%
At usual residence, and private dwelling	379,683	248,991	470,661	294,324	95,547	1,489,206	92.7%
Excluding deaths in first six months	379,602	248,874	470,382	293,373	94,215	1,486,446	92.6%
Excluding missing household income or car access	303,357	198,486	396,843	256,077	87,231	1,241,994	77.3%
Excluding missing education and labour force	–	–	395,277	253,473	85,050	733,800	79.4%
Female cohort							
Full census	383,271	275,091	521,778	311,943	122,361	1,614,444	100.0%
At usual residence, and private dwelling	365,028	249,357	499,086	295,224	113,178	1,521,873	94.3%
Excluding deaths in first six months	364,980	249,318	498,909	294,624	112,332	1,520,163	94.2%
Excluding missing household income or car access	291,132	198,681	414,351	259,293	104,829	1,268,286	78.6%
Excluding missing education and labour force	–	–	412,428	254,754	100,647	767,829	80.3%
Male linked deaths							
Full census	342	693	1644	7026	9387	19,092	100.0%
At usual residence, and private dwelling	312	627	1512	6450	8421	17,322	90.7%
Excluding deaths in first six months	231	513	1230	5496	7086	14,556	76.2%
Excluding missing household income or car access	183	402	1014	4827	6594	13,020	68.2%
Excluding missing education and labour force	–	–	1011	4755	6414	12,180	67.5%
Female linked deaths							
Full census	231	255	1101	4542	6369	12,498	100.0%
At usual residence, and private dwelling	213	228	1008	4155	5604	11,208	89.7%
Excluding deaths in first six months	165	189	834	3555	4758	9501	76.0%
Excluding missing household income or car access	117	150	660	3135	4470	8532	68.3%
Excluding missing education and labour force	–	–	651	3048	4287	7986	66.5%

† Percentage of full census cohort. Calculated only for 25–74 year olds for final row 'Excluding missing education and labour force'. Cell values are randomly rounded to multiple of three as per SNZ protocol.

2 Small area deprivation

2.1 Restricted cohort univariate results

The distribution of the restricted cohort by decile of NZDep91, and by age and sex, is shown in Table 27, along with the crude risk ratios and age and ethnicity adjusted odds ratios of mortality. As documented previously (Salmond and Crampton 2000), there is a strong gradient of all-cause mortality by small area deprivation in New Zealand. The strongest gradient was among males aged 45–64 years, with an age and ethnicity adjusted odds ratio of 2.35 (95% confidence interval 2.06 to 2.69) for males in decile 10 compared to decile 1 of small area deprivation. The comparable odds ratios for 25–44 year old males, and 25–44 and 45–64 year old females were 1.86 (1.39 to 2.50), 2.11 (1.51 to 2.94), and 1.80 (1.52 to 2.12), respectively. Note that the results show a linear *gradient* of mortality – there is no evidence of a threshold effect.

Table 27: All-cause mortality age and ethnicity adjusted odds ratio (95% CI) by small area deprivation in the restricted cohort

NZDep91 decile	25–44 years				45–64 years			
	Census number	Linked deaths	Crude RR	Age/ethnicity adjusted OR	Census number	Linked deaths	Crude RR	Age/ethnicity adjusted OR
Males								
1	44,361	87	1.00	1.00	34,692	405	1.00	1.00
2	43,974	87	1.01	1.06 (0.79–1.43)	30,675	420	1.17	1.14 (1.00–1.31)
3	42,714	78	0.93	1.00 (0.73–1.35)	28,542	447	1.34	1.27 (1.11–1.46)
4	42,381	93	1.12	1.17 (0.87–1.57)	27,066	432	1.37	1.29 (1.12–1.47)
5	40,440	96	1.21	1.28 (0.96–1.72)	25,563	435	1.46	1.35 (1.18–1.55)
6	39,714	105	1.35	1.43 (1.07–1.90)	24,294	459	1.62	1.46 (1.28–1.67)
7	38,877	111	1.46	1.54 (1.16–2.05)	22,533	477	1.81	1.62 (1.41–1.85)
8	36,984	117	1.61	1.68 (1.27–2.23)	22,050	495	1.92	1.68 (1.47–1.92)
9	35,382	132	1.90	1.94 (1.47–2.55)	20,805	600	2.47	2.10 (1.85–2.39)
10	30,090	111	1.88	1.86 (1.39–2.50)	17,007	573	2.89	2.35 (2.06–2.69)
Total	394,917	1017			253,227	4743		
Females								
1	47,382	66	1.00	1.00	33,813	273	1.00	1.00
2	45,414	57	0.90	0.98 (0.68–1.39)	30,117	279	1.15	1.12 (0.95–1.33)
3	43,584	51	0.84	0.96 (0.67–1.38)	28,410	267	1.16	1.11 (0.94–1.32)
4	42,519	54	0.91	1.04 (0.72–1.49)	27,060	273	1.25	1.18 (1.00–1.40)
5	40,860	57	1.00	1.15 (0.81–1.65)	25,521	303	1.47	1.37 (1.16–1.61)
6	40,383	69	1.23	1.42 (1.01–1.99)	24,822	297	1.48	1.32 (1.12–1.56)
7	39,639	60	1.09	1.25 (0.88–1.78)	23,235	309	1.65	1.44 (1.22–1.69)
8	38,484	66	1.23	1.39 (0.98–1.97)	22,395	312	1.73	1.47 (1.25–1.73)
9	37,986	72	1.36	1.52 (1.08–2.15)	21,351	384	2.23	1.79 (1.53–2.10)
10	35,856	96	1.92	2.11 (1.51–2.94)	17,877	342	2.37	1.80 (1.52–2.12)
Total	412,107	648			254,601	3039		

Note: The odds ratios are from a logistic regression model with age in five-year age groups and ethnicity specified as Maori, Pacific Island, and the Rest. Raw numbers are random rounded to the nearest multiple of three as per SNZ protocol.

Table 28: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by small area deprivation in the restricted cohort – males

	NZDep91 quintile									
	1 (least deprived)		2		3		4		5 (most deprived)	
	OR	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths
Cancer	1.00	426	1.03 (0.90–1.18)	387	1.16 (1.01–1.33)	402	1.21 (1.06–1.40)	387	1.51 (1.31–1.73)	423
Colorectal	1.00	87	1.07 (0.79–1.45)	81	1.26 (0.93–1.70)	87	1.08 (0.79–1.50)	66	1.10 (0.78–1.56)	54
Lung ‡	1.00	63	1.24 (0.88–1.75)	69	1.69 (1.22–2.33)	30	1.89 (1.37–2.60)	33	2.89 (2.12–3.94)	57
Prostate ‡	1.00	21	1.25 (0.72–2.20)	27	1.08 (0.60–1.96)	21	1.15 (0.63–2.10)	21	1.09 (0.57–2.06)	18
Cardiovascular disease	1.00	378	1.24 (1.08–1.43)	420	1.27 (1.10–1.46)	399	1.72 (1.51–1.97)	504	2.28 (2.00–2.61)	615
IHD	1.00	291	1.21 (1.03–1.42)	309	1.30 (1.11–1.53)	309	1.71 (1.46–1.99)	378	2.20 (1.88–2.56)	438
Cerebrovascular	1.00	30	1.66 (1.06–2.60)	48	1.57 (0.99–2.49)	39	1.82 (1.15–2.88)	45	3.39 (2.21–5.20)	72
Infection and pneumonia	1.00	12	1.71 (0.85–3.44)	21	2.66 (1.38–5.12)	30	2.57 (1.32–5.02)	27	2.37 (1.18–4.79)	24
Respiratory ‡	1.00	24	1.22 (0.68–2.17)	24	1.95 (1.15–3.32)	36	2.37 (1.41–3.99)	42	4.69 (2.88–7.64)	72
COPD ‡	1.00	15	1.25 (0.63–2.51)	18	2.43 (1.31–4.50)	30	2.66 (1.44–4.92)	33	5.15 (2.87–9.22)	57
Unintentional injury	1.00	60	1.16 (0.82–1.65)	63	1.48 (1.06–2.07)	78	1.54 (1.10–2.17)	78	1.67 (1.18–2.38)	78
Road traffic crash	1.00	27	1.42 (0.85–2.36)	33	1.83 (1.12–2.98)	42	1.82 (1.11–2.99)	42	1.63 (0.97–2.77)	36
Other unintentional	1.00	33	0.96 (0.59–1.57)	30	1.21 (0.75–1.94)	36	1.32 (0.82–2.12)	36	1.73 (1.08–2.78)	42
Suicide	1.00	48	1.41 (0.97–2.04)	63	1.01 (0.67–1.53)	42	1.71 (1.18–2.49)	63	2.04 (1.39–2.99)	63
Homicide, intentional injury	1.00	**	3.29 (0.34–31.7)	**	9.30 (1.16–74.6)	**	10.6 (1.33–84.7)	**	5.86 (0.66–52.1)	**
Other	1.00	48	1.47 (1.01–2.14)	63	2.39 (1.69–3.38)	99	2.22 (1.55–3.16)	87	3.54 (2.52–4.99)	138
All causes	1.00	999	1.18 (1.08–1.28)	1047	1.33 (1.22–1.45)	1095	1.57 (1.44–1.70)	1203	2.05 (1.88–2.23)	1416

Table 29: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by small area deprivation in the restricted cohort – females

	NZDep91 quintile									
	1 (least deprived)		2		3		4		5 (most deprived)	
	OR	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths
Cancer	1.00	423	1.01 (0.88–1.16)	375	1.17 (1.02–1.34)	402	1.16 (1.00–1.33)	372	1.28 (1.11–1.49)	369
Colorectal	1.00	93	0.54 (0.38–0.77)	45	0.99 (0.73–1.35)	72	0.90 (0.65–1.24)	57	0.64 (0.43–0.95)	33
Lung ‡	1.00	27	1.75 (1.06–2.89)	42	2.46 (1.53–3.96)	54	2.78 (1.73–4.45)	60	3.38 (2.12–5.41)	72
Breast	1.00	114	1.08 (0.83–1.40)	108	1.01 (0.76–1.33)	93	0.93 (0.70–1.24)	81	1.04 (0.78–1.41)	81
Cardiovascular disease	1.00	120	1.44 (1.13–1.82)	156	1.65 (1.31–2.09)	174	1.86 (1.47–2.34)	192	2.69 (2.15–3.37)	282
IHD	1.00	60	1.64 (1.19–2.28)	90	1.77 (1.28–2.46)	93	2.07 (1.50–2.84)	105	2.85 (2.08–3.89)	144
Cerebrovascular	1.00	33	1.38 (0.87–2.19)	42	1.61 (1.03–2.53)	45	1.63 (1.03–2.57)	42	2.39 (1.54–3.71)	63
Infection and pneumonia	1.00	9	0.96 (0.39–2.37)	9	1.19 (0.50–2.81)	12	0.93 (0.37–2.34)	9	1.19 (0.49–2.88)	15
Respiratory ‡	1.00	24	0.87 (0.48–1.58)	18	1.19 (0.68–2.09)	24	2.35 (1.43–3.85)	48	3.00 (1.83–4.91)	60
COPD ‡	1.00	18	0.77 (0.39–1.53)	15	0.99 (0.52–1.88)	18	1.99 (1.14–3.49)	33	3.09 (1.80–5.30)	48
Unintentional injury	1.00	18	1.43 (0.79–2.60)	24	1.69 (0.94–3.03)	30	1.19 (0.62–2.25)	18	2.26 (1.26–4.06)	36
Road traffic crash	1.00	12	1.93 (0.89–4.18)	18	1.90 (0.87–4.15)	18	1.47 (0.64–3.37)	12	3.12 (1.47–6.64)	30
Other unintentional	1.00	9	0.87 (0.32–2.34)	6	1.49 (0.62–3.60)	9	0.89 (0.31–2.51)	6	1.23 (0.44–3.40)	6
Suicide	1.00	15	1.20 (0.61–2.38)	15	1.48 (0.76–2.89)	21	1.48 (0.75–2.94)	15	2.07 (1.05–4.08)	18
Homicide, intentional injury	1.00	**	**	**	2.03 (0.48–8.57)	**	0.85 (0.14–5.18)	**	1.73 (0.35–8.51)	**
Other	1.00	60	0.84 (0.57–1.23)	45	1.17 (0.82–1.67)	63	1.63 (1.17–2.28)	87	1.91 (1.37–2.67)	108
All causes	1.00	672	1.07 (0.96–1.20)	651	1.28 (1.16–1.43)	732	1.37 (1.24–1.53)	747	1.73 (1.56–1.92)	897

Note: The odds ratios are from a logistic regression model with age in 10-year age groups and ethnicity dichotomised as Maori and Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

‡ Only age-group 45–64 included in analysis.
** Number of deaths all small resulting in release by SNZ as 6 in each cell (ie, the smallest possible number released by SNZ in the NZCMS).

The NZDep91 gradients of mortality by cause of death have not been fully documented before in New Zealand. Table 28 and Table 29 present the association of specific causes of death with quintiles of NZDep91, by sex for 25–64 year olds combined. (Respiratory, prostate and lung cancer death analyses were for 45–64 year olds only.) Among both males and females, the strongest gradients were for cardiovascular disease (ORs of 2.28 and 2.69 for the most deprived quintile of small areas compared to the least deprived, for males and females respectively), respiratory disease (4.69 and 3.00, respectively), and suicide (2.04 and 2.07, respectively). Among the remaining first-level aggregates of cause specific mortality (cancer, infection and pneumonia, unintentional injury, homicide, and 'other'), the odds ratios comparing the most deprived quintile to the least deprived quintile were greater than 2.0 among males for infection and pneumonia, homicide (although with a wide confidence interval), and other. Among females, only unintentional injury had an odds ratio greater than 2.0. At the second-level aggregation of cause of death, the following was notable:

- There were strong gradients for lung cancer among both sexes. There was also a suggestion of higher colorectal cancer mortality among the *least* deprived small areas among females, but not among males. There was no obvious gradient for breast and prostate cancer. Thus the modest overall association of cancer with deprivation was mostly due to lung cancer deaths.
- Within cardiovascular disease, both ischaemic heart disease and cerebrovascular disease had strong gradients by small area deprivation.
- Among unintentional injury deaths, on balance, the gradient was stronger for road traffic crashes than other unintentional injuries for both males and females – except for the most deprived quintile the trend among males was towards a stronger association with road traffic crashes. (However, there were notable effects on the injury odds ratios from selection bias and misclassification of the mortality outcome, as demonstrated in subsequent sections.)
- While based on few deaths, there was an apparent gradient of homicide deaths by small area deprivation among males.

Perhaps the most notable finding overall from Table 28 and Table 29 is that, *except for cancers other than lung cancer, **each** cause of death was strongly associated with small area deprivation.* Previous analyses among males by occupational class in New Zealand found a weak association of occupational class with IHD mortality in the 1970s, but strengthening in the 1980s (Kawachi et al 1991; Pearce et al 1983b). This transition of the IHD gradient is confirmed by the NZDep (and other socioeconomic factor) analyses presented in this report demonstrating strong IHD mortality gradients during the period 1991–94.

2.2 Possible impact of bias

Without doubt there was a strong association of small area deprivation with mortality. Extensive sensitivity analyses about the impact of selection and linkage biases, health selection and exclusion of the non-active labour force are presented in Appendix C. While of interest in themselves, these analyses were particularly helpful for estimating the impact of the same biases on the income results.

Considering the net impact of selection and linkage biases on *all-cause* mortality, they were relatively modest for the mid-decile comparisons, but were more notable comparing the least and most deprived deciles. Among 45–64 year olds linkage and selection bias tended to off-set each. However, among females aged 25–44 and 45–64 years and males aged 25–44 years the selection and linkage biases compounded each other such that analyses on the restricted cohort *underestimated* the excess odds ratio comparisons of the most and least deprived deciles by about 25%.

Considering the net impact of selection and linkage biases on *cause-specific male* mortality for 25–64 year olds combined, linkage and selection biases off-set each other for *injury* deaths in the restricted cohort. For other causes of death the analyses on the restricted cohort *underestimated* the 'true' gradient by:

- approximately 10% for *cancer* and *cardiovascular disease* deaths
- approximately 25% for *suicide* deaths.

For *cause-specific female* mortality for 25–64 year olds combined, the 'true' gradient for analyses on the restricted cohort for *cancer* deaths was underestimated by approximately 30% – but the cancer gradient was modest to start with. The net effect of linkage and selection biases for *cardiovascular disease* was negligible. The net effect for female *suicide* and *injury* deaths was unable to be robustly determined, but presumably the gradients by non-cancer and non-cardiovascular disease on average tended to be *underestimated* given the approximately 25% underestimate of the all-cause mortality.

Drift health selection was assumed to not affect mortality gradients by small area deprivation. However, deprivation mortality gradients (as for gradients by any measure of socioeconomic position) were susceptible to *differential health selection* when the non-active labour force was excluded. The magnitude of any such differential health selection was important for two reasons. First, it provides an estimate of the likely bias for occupational class mortality gradients where data was only available for those currently in employment. Second, it might disclose potential problems for the inclusion of labour force status in multivariate analyses.

The sensitivity analyses about possible differential health selection in Appendix C were difficult to interpret. At best, *the results **suggested** the possibility of differential health selection for cancer (and cardiovascular disease) – but no stronger conclusion could be made.* If there was differential health selection for cancer deaths (and cardiovascular disease death), then adjusting for labour force status in multivariate analyses may result in an underestimate of the independent association of small area deprivation (and by extrapolation other socioeconomic factors) with mortality. But it was not possible to specify with precision the relative contributions of differential health selection, effect modification and confounding/mediation by labour force status to the change in the deprivation-mortality gradients following exclusion of the non-active labour force.

3 Highest qualification

3.1 Restricted cohort univariate results

The distribution of the restricted cohort by category of highest qualification by age groups and sex is shown in Table 30, along with the crude risk ratios and age and ethnicity adjusted odds ratios of mortality. There was a strong association of education with all-cause mortality in each sex by age group. When assessing the gradient, the 'other school qualification' category should be put aside as being of indeterminate rank compared to the other school categories. Doing so, the following trends emerge:

- Among 25–44 year olds (both sexes), there was little to separate the four middle categories of '10 years of school' (School Certificate), '11–12 years ...' (University Entrance and Bursary), 'trade ...', and 'undergraduate ...'.
- Among 45–64 year olds (both sexes), the age and ethnicity adjusted odds ratio of mortality for those with 'graduate' compared to those with 'nil' education for males was 0.55 (95% confidence interval 0.47 to 0.64) and for females was 0.75 (0.59 to 0.94). The same odds ratios for 'undergraduate' compared to 'nil' education were 0.72 (0.64 to 0.81) for males and 0.69 (0.62 to 0.78) for females. Otherwise the 'trade', '11–12 years of school', and '10 years of school' categories were inseparable for males (odds ratios of 0.80 or 0.81) and fairly similar for females (odds ratios 0.79, 0.81, and 0.92)
- For all sex by age groups, those with nil education consistently had a greater mortality odds than the middle education groups, and those with 'graduate' or 'postgraduate' education consistently had a lower mortality odds than the middle education groups.

The highest and lowest educational categories stood out from the remaining categories, giving the impression of a stepped gradient of mortality rather than a monotonic gradient.

Table 30: All-cause mortality age and ethnicity adjusted odds ratio (95% CI) by highest qualification in the restricted cohort

Category of highest qualification	25–44 years				45–64 years			
	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR
Males								
Graduate, postgraduate	47,187	72	0.42	0.47 (0.36–0.61)	18,486	186	0.43	0.55 (0.47–0.64)
Undergraduate, technical, teaching	43,218	96	0.62	0.67 (0.53–0.84)	24,723	354	0.61	0.72 (0.64–0.81)
Trade certificate, other tertiary	115,677	276	0.66	0.74 (0.63–0.87)	72,069	1203	0.71	0.80 (0.74–0.85)
11–12 years of school	39,090	84	0.60	0.73 (0.58–0.93)	12,177	210	0.74	0.81 (0.70–0.93)
10 years of school	41,256	99	0.67	0.75 (0.60–0.94)	19,116	324	0.72	0.80 (0.71–0.90)
Other school qualification	5,301	12	0.63	0.57 (0.30–1.07)	7,704	156	0.87	0.79 (0.67–0.94)
Nil (reference category)	103,551	372	1.00	1.00	99,192	2319	1.00	1.00
Total	395,277	1011			253,473	4755		
Females								
Graduate, postgraduate	35,079	36	0.41	0.50 (0.35–0.72)	9726	78	0.56	0.75 (0.59–0.94)
Undergraduate, technical, teaching	71,481	93	0.52	0.59 (0.46–0.74)	40,695	345	0.60	0.69 (0.62–0.78)
Trade certificate, other tertiary	68,340	87	0.51	0.63 (0.49–0.80)	27,654	261	0.66	0.79 (0.69–0.90)
11–12 years of school	45,213	48	0.43	0.64 (0.47–0.87)	11,940	132	0.78	0.81 (0.68–0.97)
10 years of school	68,472	78	0.46	0.55 (0.43–0.71)	27,846	312	0.79	0.92 (0.81–1.04)
Other school qualification	8,463	18	0.85	0.89 (0.55–1.44)	14,979	189	0.89	0.81 (0.70–0.95)
Nil (reference category)	115,374	288	1.00	1.00	121,917	1731	1.00	1.00
Total	412,428	651			254,754	3048		

Note: The odds ratios are from a logistic regression model with age in five-year age groups and ethnicity specified as Maori, Pacific Island, and the Rest. Raw numbers are random rounded to the nearest multiple of three as per SNZ protocol.

Table 31: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by highest qualification in the restricted cohort – males

	Category of highest qualification									
	Postgraduate, graduate, technical, teaching, nursing			Trade certificate, other tertiary			Scholarship, bursary, school certificate, other school			Nil
	OR (95% CI)		Deaths	OR (95% CI)		Deaths	OR (95% CI)		Deaths	OR	Deaths
Cancer	0.74	(0.65–0.85)	279	0.79	(0.71–0.88)	510	0.84	(0.74–0.96)	315	1.00	918
Colorectal	1.00	(0.75–1.32)	72	0.75	(0.58–0.97)	90	0.68	(0.49–0.95)	48	1.00	165
Lung ‡	0.42	(0.30–0.60)	36	0.60	(0.47–0.76)	96	0.70	(0.53–0.92)	63	1.00	255
Prostate ‡	0.71	(0.40–1.26)	15	0.43	(0.25–0.73)	18	0.75	(0.44–1.29)	15	1.00	60
Cardiovascular disease	0.57	(0.49–0.65)	249	0.80	(0.72–0.88)	609	0.70	(0.62–0.80)	315	1.00	1140
IHD	0.53	(0.45–0.63)	180	0.79	(0.71–0.89)	465	0.65	(0.56–0.76)	219	1.00	861
Cerebrovascular	0.52	(0.33–0.82)	24	0.87	(0.64–1.18)	66	0.86	(0.59–1.25)	36	1.00	111
Infection and pneumonia	1.47	(0.85–2.52)	24	1.09	(0.65–1.82)	27	1.79	(1.08–2.98)	30	1.00	33
Respiratory ‡	0.48	(0.29–0.80)	18	0.70	(0.50–1.00)	45	0.68	(0.44–1.04)	24	1.00	108
COPD ‡	0.45	(0.24–0.82)	12	0.68	(0.45–1.01)	36	0.70	(0.43–1.14)	21	1.00	84
Unintentional injury	0.51	(0.36–0.71)	48	0.75	(0.58–0.97)	96	0.79	(0.59–1.06)	69	1.00	147
Road traffic crash	0.59	(0.38–0.93)	27	0.77	(0.53–1.11)	51	0.71	(0.46–1.08)	33	1.00	75
Other unintentional	0.42	(0.25–0.70)	21	0.72	(0.50–1.05)	48	0.88	(0.59–1.31)	36	1.00	75
Suicide	0.56	(0.38–0.82)	39	0.89	(0.66–1.20)	84	1.06	(0.77–1.46)	63	1.00	99
Homicide, intentional injury	0.80	(0.24–2.65)	**	0.94	(0.35–2.52)	**	1.03	(0.35–3.08)	**	1.00	**
Other	0.53	(0.38–0.73)	48	0.62	(0.48–0.79)	93	0.68	(0.51–0.90)	63	1.00	234
All causes	0.62	(0.57–0.68)	705	0.78	(0.73–0.83)	1482	0.78	(0.72–0.85)	885	1.00	2691

Note: The odds ratios are from a logistic regression model with age in 10-year age groups and ethnicity dichotomised as Maori and Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

‡ Only age-group 45–64 included in analysis.

** Number of deaths all small resulting in release by SNZ as 6 in each cell (ie, the smallest possible number released by SNZ in the NZCMS).

Table 32: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by highest qualification in the restricted cohort – females

	Category of highest qualification							
	Postgraduate, graduate, technical, teaching, nursing		Trade certificate, other tertiary		Scholarship, bursary, school certificate, other school		Nil	
	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR	Deaths
Cancer	0.81 (0.72–0.92)	342	0.88 (0.75–1.02)	207	0.85 (0.75–0.95)	408	1.00	987
Colorectal	1.14 (0.85–1.52)	66	0.97 (0.66–1.43)	30	0.88 (0.65–1.19)	60	1.00	141
Lung ‡	0.24 (0.14–0.42)	12	0.44 (0.26–0.77)	15	0.75 (0.55–1.02)	51	1.00	174
Breast	0.77 (0.60–1.00)	84	0.79 (0.58–1.09)	48	0.87 (0.69–1.10)	102	1.00	240
Cardiovascular disease	0.49 (0.40–0.61)	102	0.54 (0.41–0.70)	63	0.74 (0.62–0.87)	183	1.00	573
IHD	0.35 (0.25–0.49)	39	0.47 (0.32–0.69)	30	0.75 (0.60–0.93)	102	1.00	324
Cerebrovascular	0.77 (0.53–1.11)	39	0.56 (0.33–0.95)	18	0.75 (0.53–1.06)	42	1.00	126
Infection and pneumonia	1.17 (0.55–2.48)	12	0.40 (0.09–1.70)	**	1.54 (0.81–2.94)	**	1.00	24
Respiratory ‡	0.39 (0.23–0.67)	15	0.52 (0.28–0.97)	12	0.67 (0.45–0.99)	33	1.00	120
COPD ‡	0.35 (0.19–0.66)	12	0.41 (0.19–0.88)	9	0.51 (0.31–0.83)	21	1.00	93
Unintentional injury	0.68 (0.41–1.13)	24	0.92 (0.54–1.58)	18	0.83 (0.53–1.30)	30	1.00	51
Road traffic crash	0.56 (0.29–1.08)	12	0.79 (0.41–1.54)	12	0.85 (0.50–1.45)	24	1.00	39
Other unintentional	0.96 (0.41–2.21)	9	1.26 (0.51–3.13)	6	0.75 (0.32–1.78)	9	1.00	18
Suicide	1.18 (0.68–2.06)	24	0.87 (0.42–1.80)	9	1.17 (0.68–2.01)	24	1.00	30
Homicide, intentional injury	1.94 (0.56–6.70)	**	0.53 (0.06–4.64)	**	0.59 (0.11–3.14)	**	1.00	**
Other	0.35 (0.24–0.52)	30	0.67 (0.46–0.97)	33	0.80 (0.62–1.04)	81	1.00	222
All causes	0.66 (0.60–0.73)	552	0.74 (0.66–0.83)	348	0.81 (0.74–0.88)	783	1.00	2019

Note: The odds ratios are from a logistic regression model with age in 10-year age groups and ethnicity dichotomised as Maori and Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

‡ Only age-group 45–64 included in analysis.

** Number of deaths all small resulting in release by SNZ as 6 in each cell (ie, the smallest possible number released by SNZ in the NZCMS).

At face value, the education all-cause mortality gradients appear weaker than those for small area deprivation. However, the reference category (nil education) is large – 39.1% and 47.9% of 45–64 year old males and females, respectively. Thus the highest qualification comparisons were for a less 'extreme' variation of socioeconomic position.

Cause-specific mortality gradients by highest qualification are shown in Table 31 for males and Table 32 for females. The gradients closely resemble those observed for small area deprivation shown in Tables 28 and 29. The odds of mortality among those with nil qualification was at least twice that of those with 'postgraduate, graduate, technical, teaching, or nursing training' for lung cancer, and respiratory disease for both sexes, and additionally for ischaemic heart disease and 'other' diseases among females. These causes of disease all share tobacco smoking and other lifestyle risk factors that may be influenced by awareness and receptivity to health education measures.

3.2 Possible impact of bias

It was not possible to directly determine the impact of linkage bias as education is not recorded on mortality data. Regarding *selection bias*, analyses of education–mortality gradients conducted on the restricted cohort were little different from those conducted on the full cohort (ie, no compelling evidence of selection bias) except for male injury deaths (see Appendix C). Given the lack of a consistent pattern between the NZDep91 and education selection bias analyses, it seems difficult to reliably predict what the magnitude and direction of any selection bias might be for other socioeconomic factors such as income.

As with small area deprivation, *drift health selection* was assumed to not affect mortality gradients by highest qualification. There was some suggestion of *differential health selection* by educational status for cancer mortality among 45–64 year old males, but it was not compelling. But overall, the results were consistent with the results by NZDep91 failing to strongly suggest differential health selection, although the possibility remained for cancer (and cardiovascular disease).

4 Car access

4.1 Restricted cohort univariate results

Only all-cause mortality results are presented for car access. The age and ethnicity adjusted odds ratios of all-cause mortality were stronger for males than females, but comparable between 25–44 and 45–64 year olds within sexes (Table 33 below). For each of the four sex by age groups, there was a consistent incremental association such that access to two cars in the household was associated with lower mortality than access to one car which in turn was better than access to none.

4.2 Possible impact of bias

Formal sensitivity analyses of the likely impact of selection bias, linkage bias and health selection on the association of car access with all-cause mortality are not presented. However, one might expect that the odds ratios of mortality for those with nil car access (proxy for low-asset wealth) compared to those with access to two or more cars (proxy for high-asset wealth) shown in Table 33 are:

- little, if any, affected by selection bias, based on sensitivity analyses for NZDep91 and highest qualification
- modestly underestimated due to linkage bias, based on previous analyses suggesting that linkage bias caused a modest underestimate of the odds ratios of mortality comparing low occupational classes or people living in more deprived areas with high occupational classes or people living in less deprived areas respectively
- probably unaffected by health selection acting over the short term, due to car ownership probably not being affected in the short term by health status.

Table 33: All-cause mortality age and ethnicity adjusted odds ratio (95% CI) by car access in the restricted cohort

Car access	25–44 years				45–64 years			
	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR
Males								
≥2 cars	213,522	459	1.00	1.00	149,349	2163	1.00	1.00
1 car	162,021	456	1.31	1.31 (1.15–1.50)	92,766	2127	1.58	1.31 (1.24–1.40)
Nil	19,734	96	2.26	2.21 (1.76–2.78)	11,358	462	2.81	2.21 (1.99–2.45)
Total	395,277	1011			253,473	4752		
Females								
≥2 cars	211,716	312	1.00	1.00	133,452	1191	1.00	1.00
1 car	173,232	270	1.06	1.15 (0.98–1.36)	104,985	1518	1.62	1.30 (1.20–1.40)
Nil	27,477	69	1.70	1.76 (1.34–2.31)	16,320	342	2.35	1.60 (1.41–1.81)
Total	412,425	651			254,757	3051		

Note: The odds ratios are from a logistic regression model with age in five-year age groups and ethnicity specified as Maori, Pacific Island, and the Rest. Raw numbers are random rounded to the nearest multiple of three as per SNZ protocol.

Table 34: All-cause mortality age and ethnicity adjusted odds ratio (95% CI) by housing tenure in the restricted cohort

Housing tenure	25–44 years				45–64 years			
	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR
Males								
Owned with a mortgage	228,036	585	1.00	1.00	100,926	1410	1.00	1.00
Owned freehold	59,331	171	1.12	1.08 (0.91–1.29)	117,819	2472	1.50	0.99 (0.92–1.06)
Private tenancy	66,606	93	0.54	0.63 (0.50–0.78)	15,927	288	1.29	1.14 (1.00–1.30)
Public tenancy	21,546	108	1.95	1.88 (1.51–2.33)	10,089	402	2.85	1.88 (1.68–2.12)
Tenancy not specified	3,039	6	0.77	0.76 (0.34–1.71)	1,362	42	2.21	1.63 (1.19–2.24)
Total	378,561	966			246,120	4611		
Females								
Owned with a mortgage	245,271	363	1.00	1.00	85,500	801	1.00	1.00
Owned freehold	61,953	123	1.34	1.15 (0.94–1.41)	136,077	1758	1.38	0.97 (0.89–1.06)
Private tenancy	58,827	63	0.72	0.87 (0.66–1.14)	12,108	96	0.85	0.74 (0.60–0.92)
Public tenancy	29,961	69	1.56	1.59 (1.21–2.08)	13,125	273	2.22	1.41 (1.22–1.63)
Tenancy not specified	2,598	6	1.56	1.85 (0.87–3.93)	1,161	24	2.21	1.53 (1.02–2.30)
Total	398,613	627			247,974	2952		

5 Household tenure

Only the all-cause mortality results are presented for housing tenure – a proxy for asset wealth.

5.1 Restricted cohort univariate results

Some of the odds ratios of all-cause mortality by household tenure shown in Table 34 above were contrary to expectation. If owning your own home (ie, owned with or without a mortgage) is taken as a proxy for asset wealth, then one would expect such people to have a lower mortality risk than people renting (ie, public and private tenants). However, this pattern is not evident among the results in Table 34. Rather, there was tendency for people living in private tenancy dwellings to have a *lower* odds of mortality than those owning a home with a mortgage, and there was a tendency for people living in public tenancy dwellings to have a *higher* odds of mortality. If private tenancy were a marker of lower socioeconomic position, then one would have expected them to have a higher risk of mortality than home-owners. Differential residential mobility by tenure is the likely explanation, and is considered further in the next section on bias.

It is interesting to note that the crude risk ratios all suggest a higher mortality for freehold homeowners compared to mortgaged homeowners. However, the age and ethnicity adjusted odds ratios are all near 1.0. This confounding is due to freehold owners tending to be older.

5.2 Possible impact of bias

Formal sensitivity analyses of the likely impact of selection bias, linkage bias and health selection on the association of housing tenure with all-cause mortality are not presented. The results in Table 34 are probably not substantially affected by selection bias or short-term health selection. There are, however, strong reasons to be concerned about linkage bias. Of all the socioeconomic factors considered in this report, housing tenure is probably the one most strongly associated with residential mobility. For all 25–64 year olds combined on the 1991 census, the percentage of people having lived less than two years at their usual residence on census night was:

- 34.9% for mortgaged owners
- 20.6% for freehold owners
- 76.4% for private tenants
- 45.6% for public tenants.

Private tenants are by far the most mobile population, and therefore probably the least likely to be linked upon death to their census record, all else being equal. Thus, there was strong reason to believe that the odds ratios of mortality for private tenants are *underestimated* in Table 34. Likewise, public tenants were reasonably mobile, and the already elevated odds ratios of mortality for public tenants would probably have been greater again if all deaths had been linked to their census record.

An alternative method to assess the likely linkage bias by housing tenure is to plot the mortality risks for each category of housing tenure over time, ie, the same type of plot as that for assessing the possible impact of health selection used in Appendix C. If greater residential mobility was a particular problem for, say, private tenants, then one would expect the observed mortality risk to diminish steadily over time for that group of people as an increasing proportion of deaths were unable to be liked back to their census record. An inspection of these plots (not shown) suggested that this was *probably* occurring for 25–44 year old male private tenants, and *possibly* for both 25–44 and 45–64 year old male public housing tenants.

In conclusion, there was strong evidence of an elevated mortality risk for public tenants compared to mortgaged owners – although it was probably still underestimated. The generally lower mortality risk for private tenants was probably a result of linkage bias.

6 Labour force status

6.1 Restricted cohort univariate results

The odds ratios of all-cause mortality by labour force status are shown in Table 35 below. (The derivation of the five labour force status groups is shown in Figure 4 of this report.) As expected, there was a marked excess mortality risk among the non-labour force ('seeking or available', and 'retired, homemaker, permanently sick, student, etc'). The majority of this elevated mortality risk is undoubtedly due to those in poor health being among the non-labour force (ie, drift health selection).

Perhaps the two findings of note are:

- the unemployed had an approximately 50% increased odds of mortality compared to the full-time employed for each sex by age group, except 45–64 year old females
- the part-time employed had an increased odds of mortality compared to the full-time employed for each sex by age group, except 25–44 year old females.

The association of labour force status with broad causes of death is shown in Table 36 below. *A priori*, the result of greatest interest was that for unemployment and suicide. For both males and females, unemployment was associated with a notably elevated odds ratio of suicide death compared to full-time employed, with odds ratios of 2.86 (95% CI 1.94 to 4.20) and 2.51 (1.03 to 6.14) for males and females, respectively. Interestingly, part-time employed males also had an elevated odds ratio of suicide death for males (2.20, 1.21 to 1.56), but not part-time employed females (0.62, 0.28 to 1.38). Also, there were substantially elevated odds ratios (with 95% CIs excluding 1.00) of suicide death among both non-labour force categories, for both males and females. Thus, compared to the full-time employed, and excepting part-time employed females, there were elevated odds of suicide death for all non-referent labour force status categories.

Table 35: All-cause mortality age and ethnicity adjusted odds ratio (95% CI) by labour force status in the restricted cohort

Labour force status	25–44 years				45–64 years			
	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR
Males								
Full-time employed	322,701	681	1.00	1.00	169,566	1884	1.00	1.00
Part-time employed	13,644	51	1.77	1.75 (1.31–2.32)	13,284	234	1.59	1.23 (1.07–1.42)
Unemployed (seeking and available)	28,749	87	1.43	1.45 (1.15–1.82)	10,563	183	1.56	1.47 (1.26–1.71)
Seeking or available †	6,957	30	2.04	2.00 (1.38–2.91)	3,684	84	2.05	1.64 (1.31–2.05)
Retired, homemaker, permanently sick, student, etc ‡	23,229	165	3.37	3.24 (2.72–3.86)	56,373	2367	3.78	2.44 (2.26–2.62)
Total	395,280	1014			253,470	4752		
Females								
Full-time employed	182,637	231	1.00	1.00	88,395	534	1.00	1.00
Part-time employed	89,802	120	1.06	1.04 (0.83–1.30)	45,276	345	1.26	1.19 (1.04–1.36)
Unemployed (seeking and available)	23,163	39	1.33	1.50 (1.07–2.10)	6,597	45	1.13	1.05 (0.77–1.43)
Seeking or available †	10,791	30	2.20	2.25 (1.51–3.33)	3,615	45	2.06	1.73 (1.26–2.37)
Retired, homemaker, permanently sick, student, etc ‡	106,035	234	1.74	2.00 (1.66–2.41)	110,874	2082	3.11	2.10 (1.89–2.33)
Total	412,428	654			254,757	3051		

† The 'seeking or available' include those people that were either available for work, or seeking work, but not both. Both these requirements must be meet to be labelled as unemployed (and hence part of the active labour force).

‡ The 'Retired, homemaker, permanently sick, students, etc' are all those people that were not employed, and neither seeking nor available for work.

Table 36: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by labour force status in the restricted cohort

	Category of labour force status									
	Full-time employed		Part-time employed		Unemployed		Seeking or available		Retired, homemaker, etc	
	OR	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths
Males										
Cancer	1.00	951	1.27 (1.04–1.56)	108	1.27 (1.01–1.59)	81	1.34 (0.94–1.91)	33	2.13 (1.92–2.36)	852
Lung cancer	1.00	168	1.33 (0.86–2.07)	24	2.27 (1.51–3.41)	27	2.43 (1.35–4.38)	12	2.48 (2.00–3.09)	228
Cardiovascular disease	1.00	972	1.40 (1.15–1.70)	120	1.46 (1.19–1.79)	102	1.59 (1.16–2.18)	42	2.62 (2.38–2.89)	1083
Unintentional injury	1.00	252	0.94 (0.54–1.65)	15	1.24 (0.84–1.85)	27	1.30 (0.64–2.64)	6	1.44 (1.05–1.97)	60
Suicide	1.00	150	2.20 (1.33–3.64)	18	2.86 (1.94–4.20)	33	4.13 (2.28–7.46)	12	3.43 (2.48–4.73)	72
All causes	1.00	2565	1.37 (1.21–1.56)	285	1.44 (1.27–1.63)	270	1.71 (1.41–2.07)	111	2.74 (2.58–2.92)	2535
Females										
Cancer	1.00	480	1.20 (1.04–1.39)	309	0.94 (0.67–1.31)	39	1.62 (1.16–2.28)	36	1.63 (1.45–1.83)	1083
Lung cancer	1.00	60	1.00 (0.64–1.54)	30	1.68 (0.83–3.39)	9	2.27 (1.03–4.98)	9	1.74 (1.27–2.39)	177
Cardiovascular disease	1.00	144	1.11 (0.85–1.45)	84	1.20 (0.71–2.01)	18	1.59 (0.88–2.87)	12	2.84 (2.34–3.44)	663
Unintentional injury	1.00	39	1.05 (0.61–1.79)	18	0.63 (0.20–2.06)	6	1.75 (0.62–4.91)	6	1.76 (1.15–2.69)	60
Suicide	1.00	27	0.62 (0.28–1.38)	9	2.51 (1.03–6.14)	6	5.14 (2.10–12.6)	6	2.21 (1.31–3.72)	45
All causes	1.00	765	1.16 (1.03–1.30)	462	1.21 (0.96–1.51)	84	1.88 (1.47–2.40)	69	2.20 (2.01–2.40)	2316

† The odds ratios are from a logistic regression model with age in 10-year age groups and ethnicity dichotomised as Maori and Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

There was an elevated risk of cardiovascular disease mortality among all non-referent groups (Table 36). Some of the elevation among the non-active labour force may be due to health selection. Also, some of the elevation among all non-referent groups may be due to confounding/mediation by other socioeconomic factors – this will be investigated further in the multivariate analyses presented in Section 9 of this chapter. It is tantalising to speculate that some of the elevated cardiovascular disease mortality risk may be attributable to psychosocial pathways – this too will be pursued in Section 9.

A general finding of note in Table 36 was an elevated risk of mortality among the two non-labour force categories for *all* causes of death. However, the elevation was least for injury deaths, consistent with much, but not all, of the elevated risk of mortality among the non-labour force being due to health selection.

6.2 Possible impact of bias

There was no evidence of selection bias affecting analyses of the association of labour force status with mortality on the restricted cohort. It was not possible to conduct a direct assessment of the likely impact of linkage bias. However, it was likely that any linkage bias would have caused an *underestimate* of the odds ratios for the unemployed, part-time employed, and non-labour force compared to the full-time employed. The reason is that the increased residential mobility and lower socioeconomic position of these labour force groups are likely to be associated with a reduced probability of being linked to a census record.

The elevated mortality among the *non-labour force* is, in large part at least, due to health selection (see Appendix C, Section 4.1 for a detailed sensitivity analysis). Both the *unemployed* and the *part-time employed* had elevated mortality risks compared to the full-time employed. While this elevation may be the result of confounding (and will be assessed later in the multivariate analyses), the analyses here strongly suggest that it was *not* due to health selection.

7 Occupational class

Occupational class analyses presented a challenge due to only current occupation being available on the census. The unlinked occupational class analyses by Pearce and colleagues', however, provide a useful point of comparison (Davis et al 1999a; Kawachi et al 1991; Marshall et al 1993; Pearce and Bethwaite 1997; Pearce et al 1983a; Pearce et al 1983b; Pearce et al 1984; Pearce et al 1985; Pearce and Howard 1986; Pearce et al 1991; Pearce et al 1993). The objectives of this section, therefore, are to measure the association of occupational class (based on *current* occupation) with mortality in the NZCMS.

Extensive sensitivity analyses are presented in Appendix C (Section 5) with the objectives of:

- assessing the likely impact of health selection on the occupational class mortality gradients observed in the NZCMS
- comparing and contrasting the results from unlinked analyses (ie, as done by Pearce and colleagues) and results using the NZCMS, with particular attention to identifying sources of numerator–denominator bias that may be affecting unlinked analyses.

One objective of the NZCMS is to include occupational class analyses in multivariate analyses, but it is a limited exercise due to many people having no occupation in the restricted cohort and the availability of *current* occupation only on census data. Thus, to maximise numbers the univariate results by occupational class in this section are presented for *all* those census respondents with an occupational class – not just those in the restricted cohort.

For the occupational class analyses, only deaths in the second and third year of follow-up were included. Three reasons contributed to this decision. First, when plotting the mortality risks over time for 45–64 year old males there was evidence of differential health selection acting on occupational class 6 in both the first and second six months of follow-up (presented in Appendix C, Section 5.4.1). Second, occupation on the census was that in the last four weeks. Discarding deaths in the first year for occupational class analyses meant a time-lag between exposure ascertainment and outcome of at least one year. By comparison, income on the census was that for the year prior to census night, making the exposure ascertainment period at least 6 to 18 months prior to the death event when deaths in the first six months were excluded. Thus, excluding deaths in the first year for occupational class analyses gave a roughly comparable 'average' minimum time-lag between exposure assessment and outcome as that for income analyses excluding deaths in the first six months of follow-up. Third, there were problems with the coding of occupation on the mortality data for all of 1991. These problems caused a large overestimate of occupational class 1 deaths and a large underestimate of occupational class 5 deaths. While the same coding problems did not affect census data, it was parsimonious to include the same period of follow-up for occupational class as that available for the analysis of linkage bias by occupational class.

7.1 Full cohort univariate results

Occupational class mortality gradients for all-cause mortality, for males and females aged 25–44 and 45–64 years, are shown in Table 37 below. Confidence intervals are presented for the age and ethnicity adjusted odds ratios only, using NZSEI class 4 (the largest) as the reference group. Both the age and ethnicity and age-only adjusted odds ratios are presented with NZSEI class 1 as the reference group to aid interpretation and comparison with the results of Pearce and colleagues.

The results for females should be treated with considerable caution. 39.6% and 50.6% of the 25–44 and 45–64 year old female cohort, respectively, had no current occupational class. Moreover, 54.2% and 73.1%, respectively, of the deaths had no current occupation on census night. These caveats issued, there was some suggestion of a gradient among females with lower occupational classes having higher mortality.

The results for males were more robust. The majority of the cohort had a current occupational class (81.0% and 69.4% of 25–44 and 45–64 year olds, respectively), and 68.9% and 43.3% of 25–44 and 45–64 year old deaths, respectively, had a current occupation on census night. There was an approximately two thirds greater mortality risk among occupational class 6 compared to class 1 for both age groups when adjusting for age only. This excess mortality risk fell to approximately 50% for both age-groups after adjusting additionally for ethnicity.

Table 37: All-cause mortality by NZSEI occupational class in the *full* cohort, crude risk ratios and age/ethnicity and age only adjusted odds ratio

NZSEI occupational class	25–44 years						45–64 years					
	Census number	Linked deaths	Crude risk ratio (ref = 1)	Age/ethnicity adjusted odds ratio (ref = class 4)	Age/ethnicity adjusted OR (ref = 1)	Age only adjusted OR (ref = 1)	Census number	Linked deaths	Crude risk ratio (ref = 1)	Age/ethnicity adjusted odds ratio (ref = class 4)	Age/ethnicity adjusted OR (ref = 1)	Age only adjusted OR (ref = 1)
Males												
1	42,435	66	1.00	0.88 (0.65–1.17)	1.00	1.00	22,557	174	1.00	0.86 (0.71–1.03)	1.00	1.00
2	50,619	90	1.12	0.94 (0.72–1.23)	1.08	1.10	28,305	213	0.98	0.87 (0.74–1.03)	1.02	1.03
3	87,087	123	0.91	0.80 (0.63–1.01)	0.91	0.92	44,001	375	1.12	1.00 (0.87–1.16)	1.17	1.16
4	86,682	147	1.10	1.00	1.14	1.18	41,829	381	1.19	1.00	1.17	1.19
5	74,835	171	1.47	1.26 (1.01–1.57)	1.44	1.58	36,855	456	1.62	1.28 (1.11–1.47)	1.49	1.64
6	27,801	63	1.43	1.28 (0.95–1.73)	1.47	1.59	14,526	192	1.73	1.27 (1.06–1.51)	1.48	1.65
Farmers	40,809	69	1.10	0.97 (0.73–1.29)	1.11	1.13	27,924	234	1.11	0.82 (0.70–0.97)	0.96	0.97
Total	**410,268**	**729**					**215,997**	**2025**				
No occupation	96,114	330	2.19				95,388	2763	3.80			
Females												
1	22,974	21	1.00	1.03 (0.65–1.64)	1.00	1.00	8388	45	1.00	0.95 (0.68–1.33)	1.00	1.00
2	43,905	42	0.96	0.91 (0.62–1.32)	0.88	0.88	20,844	99	0.93	0.88 (0.69–1.12)	0.92	0.91
3	58,713	54	0.94	0.93 (0.66–1.31)	0.91	0.91	24,759	120	0.95	0.94 (0.75–1.18)	0.98	0.95
4	77,304	78	1.02	1.00	0.97	0.98	35,892	186	1.02	1.00	1.05	1.00
5	46,584	60	1.33	1.26 (0.90–1.76)	1.23	1.26	24,936	159	1.24	1.17 (0.95–1.45)	1.23	1.21
6	43,137	51	1.20	1.09 (0.77–1.56)	1.06	1.12	25,692	159	1.21	1.06 (0.86–1.32)	1.12	1.16
Farmers	20,808	18	0.91	0.86 (0.52–1.43)	0.84	0.84	12,807	63	0.93	0.83 (0.62–1.11)	0.87	0.84
Total	**313,425**	**324**					**153,318**	**831**				
No occupation	207,972	384	1.84				157,167	2253	2.80			

Note: The odds ratios are from a logistic regression model with age in five-year age groups and ethnicity trichotomised as Maori, Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

Table 38: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by current occupational class in the full cohort – males

	Classes 1 and 2		Classes 3 and 4		Classes 5 and 6		Deaths among:	
	OR	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	Farmers	No occupation (% total)
Cancer	1.00	210	1.03 (0.87–1.22)	375	1.32 (1.10–1.58)	306	120	978 (49%)
Lung cancer	1.00	30	1.34 (0.87–2.05)	72	1.71 (1.10–2.66)	63	21	276 (60%)
Cardiovascular disease	1.00	198	1.19 (1.00–1.40)	411	1.50 (1.26–1.79)	354	111	1290 (55%)
Injury	1.00	48	1.04 (0.73–1.48)	90	1.53 (1.06–2.20)	84	33	108 (30%)
Suicide	1.00	27	1.12 (0.71–1.77)	54	1.65 (1.03–2.66)	45	15	120 (46%)
All	**1.00**	**537**	**1.09 (0.98–1.21)**	**1026**	**1.43 (1.28–1.60)**	**879**	**306**	**3093 (53%)**

Note: The odds ratios are from a logistic regression model with age in 10-year age groups and ethnicity dichotomised as Maori and Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

The associations of occupational class with broad groupings of cause-specific mortality are shown in Table 38 above, for 25–64 year old males combined. Note first, though, the percentage of deaths that had no current occupational class on census night (last column of Table 38). 53% of all deaths had no occupation (and hence no occupational class) on census night, but this percentage was only 30% for injury deaths. This reduction for injury deaths is consistent with less health selection out of the labour force prior to death when the cause of death is sudden. Consequently, there would have been less opportunity for differential health selection to bias the injury gradient than, say, the cancer gradient. (Some of the percentage difference for injury deaths would have also been due to age at death.) However, the male cause of death mortality gradients by occupational class in Table 38 are similar to those for small area deprivation, highest qualification, and household income (Tables 27, 31, and 40 respectively). If differential health selection was having a large impact on the occupational mortality gradients in the NZCMS, one would expect a notable flattening of the cancer and cardiovascular disease death gradients, yet there were still notable gradients by these causes of death as shown in Table 38. However, the exclusion of the non-labour force reduced the deprivation mortality gradients for male cancer and cardiovascular disease deaths more so than for injury and suicide deaths, arguing the possibility of some differential health selection taking place (Appendix C) – although the evidence was not convincing.

7.2 Possible impact of bias and conclusion

The analyses of the association of occupational class with mortality in the NZCMS were limited by the availability of current occupation only. However, based on the extensive sensitivity analyses presented in Appendix C it was reasonable to conclude that:

- there was an occupational class mortality gradient in the expected direction for both males and females aged 25–64 years, although only weakly for females
- compared to male occupational mortality gradients by *usual* occupation, the gradients observed in the NZCMS by *current* occupation were probably an underestimate – particularly for 45–64 year old males
- it was unclear whether the shallower gradient for *current* occupational class compared to *usual* occupational class was due to health selection or confounding/mediation by the range of variables that labour force status may be a proxy for. The plots of mortality risk over time for current occupational class suggested that differential health selection had largely worn-off after the first year of follow-up
- the occupational class mortality gradient by *current* occupation observed in the NZCMS was steeper among 25–44 year old males compared to 45–64 year old males. However, if *usual* occupation data were available there would probably have been little difference in the gradient between the two age groups.

If we wish to use the occupational class results above based on current occupation as estimates of the gradient by *usual/last* occupational class, then perhaps the best conclusion is that the male occupational class mortality gradients for cancer and cardiovascular disease are probably somewhat underestimated, as a result of both:

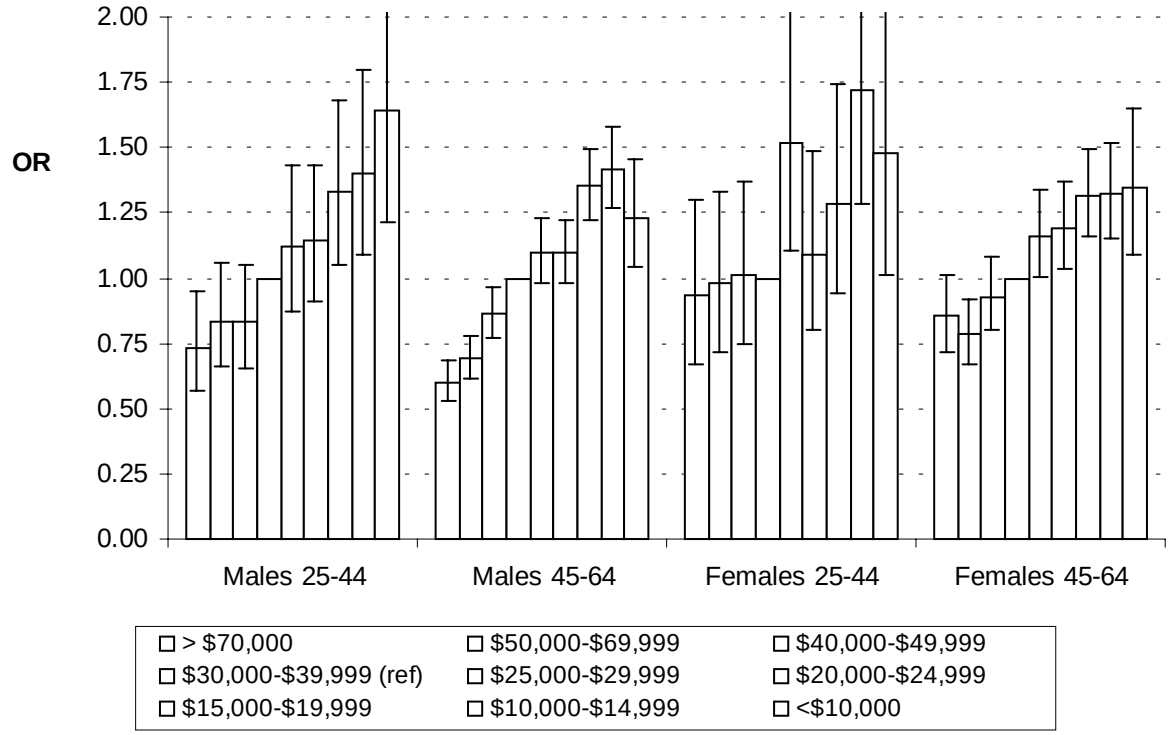
- differential health selection, (to a small extent)
- confounding/mediation by (current) labour force status.

8 Equivalised household income

8.1 Restricted cohort univariate results

Table 32 presents the distribution of the 25–64 year old restricted cohort by equivalised household income, and the crude risk ratios and age and ethnicity adjusted odds ratios of all-cause mortality by income. The associations of equivalised household income with all-cause mortality were indisputably strong and (usually) monotonic gradient (Figure 15). The gradient was strongest among males, being an approximate doubling in mortality risk for low income compared to high income.

Figure 15: Age and ethnicity adjusted odds ratios of all-cause mortality by equivalised household income



Note: Error bars are 95% confidence intervals.

Table 39: All-cause mortality age and ethnicity adjusted odds ratio (95% CI) by equivalised household income in the restricted cohort

Equivalised household income	25–44 years				45–64 years			
	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR	Number	Linked deaths	Crude risk ratio	Age/ethnicity adjusted OR
Males								
≥ $70,000	49,026	87	0.70	0.73 (0.56–0.95)	32,814	306	0.51	0.60 (0.53–0.69)
$50,000–$69,999	60,186	117	0.76	0.83 (0.66–1.06)	42,378	462	0.60	0.69 (0.62–0.78)
$40,000–$49,999	58,062	117	0.79	0.83 (0.65–1.05)	36,792	528	0.79	0.86 (0.77–0.96)
$30,000–$39,999	66,033	168	1.00	1.00	42,906	783	1.00	1.00
$25,000–$29,999	35,994	99	1.08	1.12 (0.87–1.44)	24,132	531	1.21	1.10 (0.98–1.23)
$20,000–$24,999	47,556	138	1.14	1.14 (0.91–1.43)	23,256	534	1.26	1.09 (0.98–1.22)
$15,000–$19,999	36,849	126	1.34	1.33 (1.05–1.68)	25,794	858	1.82	1.35 (1.22–1.50)
$10,000–$14,999	28,299	102	1.42	1.40 (1.09–1.79)	18,000	576	1.75	1.42 (1.27–1.58)
<$10,000	13,272	57	1.69	1.64 (1.21–2.22)	7,398	174	1.29	1.23 (1.04–1.45)
	395,277	1011			253,470	4752		
Females								
≥ $70,000	46,002	60	0.93	0.93 (0.67–1.30)	26,718	195	0.70	0.85 (0.72–1.01)
$50,000–$69,999	54,585	75	0.98	0.98 (0.72–1.33)	36,687	258	0.67	0.79 (0.67–0.92)
$40,000–$49,999	55,581	78	1.00	1.01 (0.75–1.37)	33,294	291	0.84	0.93 (0.80–1.08)
$30,000–$39,999	64,392	90	1.00	1.00	41,634	435	1.00	1.00
$25,000–$29,999	32,784	66	1.44	1.52 (1.10–2.08)	25,653	342	1.28	1.16 (1.01–1.34)
$20,000–$24,999	50,439	72	1.02	1.09 (0.80–1.49)	25,239	369	1.40	1.19 (1.03–1.37)
$15,000–$19,999	44,298	72	1.16	1.28 (0.94–1.74)	34,725	639	1.76	1.32 (1.16–1.49)
$10,000–$14,999	43,431	96	1.58	1.72 (1.29–2.30)	23,250	399	1.64	1.32 (1.15–1.52)
<$10,000	20,913	42	1.44	1.48 (1.01–2.16)	7,554	120	1.52	1.34 (1.09–1.65)
	412,425	651			254,754	3048		

Note: The odds ratios are from a logistic regression model with age in five-year age groups and ethnicity specified as Maori, Pacific Island, and the Rest. Raw numbers are random rounded to the nearest multiple of three as per SNZ protocol.

Table 40: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by equivalised household income in the restricted cohort – males

	Category of equivalised household income							
	≥$50,000		$30,000–$49,999		$20,000–$29,999		<$20,000	
	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR	Deaths
Cancer	0.61 (0.53–0.69)	390	0.81 (0.72–0.91)	603	0.89 (0.78–1.00)	444	1.00	582
Colorectal	0.85 (0.63–1.15)	90	1.04 (0.79–1.37)	123	1.00 (0.73–1.36)	78	1.00	87
Lung ‡	0.39 (0.29–0.52)	66	0.59 (0.46–0.75)	117	0.69 (0.54–0.89)	99	1.00	174
Prostate ‡	0.43 (0.24–0.77)	15	0.67 (0.42–1.07)	30	0.63 (0.37–1.07)	21	1.00	42
Cardiovascular disease	0.43 (0.38–0.49)	363	0.63 (0.56–0.70)	627	0.78 (0.70–0.87)	531	1.00	798
IHD	0.44 (0.38–0.51)	276	0.64 (0.57–0.73)	471	0.80 (0.70–0.91)	399	1.00	582
Cerebrovascular	0.36 (0.24–0.55)	33	0.64 (0.46–0.88)	69	0.70 (0.49–0.99)	51	1.00	87
Infection and pneumonia	0.90 (0.52–1.54)	27	0.99 (0.60–1.64)	39	0.80 (0.45–1.44)	18	1.00	27
Respiratory ‡	0.23 (0.14–0.37)	18	0.34 (0.23–0.50)	39	0.59 (0.41–0.84)	45	1.00	96
COPD ‡	0.18 (0.09–0.34)	12	0.37 (0.24–0.58)	30	0.70 (0.48–1.04)	39	1.00	72
Unintentional injury	0.43 (0.31–0.60)	57	0.72 (0.54–0.94)	108	0.91 (0.69–1.21)	90	1.00	102
Road traffic crash	0.50 (0.32–0.80)	30	0.81 (0.55–1.19)	57	0.95 (0.63–1.43)	45	1.00	51
Other unintentional	0.37 (0.23–0.60)	27	0.63 (0.43–0.94)	51	0.88 (0.59–1.31)	48	1.00	54
Suicide	0.44 (0.32–0.62)	57	0.57 (0.42–0.78)	81	0.66 (0.48–0.93)	57	1.00	87
Homicide, intentional injury	0.31 (0.11–0.92)	**	0.33 (0.12–0.91)	**	0.26 (0.07–0.93)	**	1.00	**
Other	0.25 (0.18–0.34)	48	0.42 (0.32–0.54)	96	0.68 (0.54–0.86)	108	1.00	186
All causes	0.46 (0.43–0.50)	975	0.65 (0.61–0.70)	1596	0.79 (0.74–0.85)	1305	1.00	1890

† The odds ratios are from a logistic regression model with age in 10-year age groups and ethnicity dichotomised as Maori and Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

‡ Only age-group 45–64 included in analysis.

** Number of deaths suppressed due to small cell size.

Table 41: Cause-specific mortality age and ethnicity adjusted odds ratios (95% CI) by equivalised household income in the restricted cohort – females

	Category of equivalised household income							
	≥$50,000		$30,000–$49,999		$20,000–$29,999		<$20,000	
	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR (95% CI)	Deaths	OR	Deaths
Cancer	0.86 (0.75–0.98)	390	0.91 (0.81–1.02)	531	0.99 (0.88–1.13)	432	1.00	594
Colorectal	1.23 (0.90–1.67)	81	0.81 (0.59–1.12)	72	0.98 (0.71–1.36)	63	1.00	87
Lung ‡	0.40 (0.26–0.62)	27	0.65 (0.47–0.89)	60	0.78 (0.57–1.07)	60	1.00	108
Breast	1.08 (0.82–1.41)	108	1.00 (0.78–1.29)	126	1.31 (1.02–1.69)	120	1.00	123
Cardiovascular disease	0.38 (0.30–0.47)	96	0.55 (0.46–0.65)	189	0.81 (0.69–0.95)	225	1.00	405
IHD	0.27 (0.19–0.38)	42	0.44 (0.34–0.56)	90	0.70 (0.56–0.87)	120	1.00	246
Cerebrovascular	0.71 (0.48–1.07)	39	0.85 (0.61–1.21)	60	0.97 (0.68–1.38)	54	1.00	75
Infection and pneumonia	0.71 (0.28–1.77)	6	1.27 (0.63–2.56)	18	1.11 (0.53–2.36)	12	1.00	18
Respiratory ‡	0.20 (0.11–0.38)	12	0.31 (0.20–0.48)	27	0.61 (0.43–0.89)	42	1.00	99
COPD ‡	0.18 (0.09–0.37)	9	0.30 (0.18–0.51)	18	0.56 (0.36–0.86)	30	1.00	78
Unintentional injury	0.38 (0.22–0.67)	18	0.56 (0.35–0.88)	30	0.67 (0.42–1.07)	27	1.00	54
Road traffic crash	0.54 (0.29–1.01)	15	0.47 (0.26–0.86)	15	0.80 (0.46–1.39)	21	1.00	33
Other unintentional	0.12 (0.03–0.53)	6	0.70 (0.34–1.43)	12	0.43 (0.17–1.08)	6	1.00	18
Suicide	0.50 (0.27–0.91)	18	0.63 (0.37–1.07)	24	0.49 (0.26–0.93)	15	1.00	33
Homicide, intentional injury	0.12 (0.01–0.97)	**	0.30 (0.08–1.15)	**	0.31 (0.07–1.47)	**	1.00	**
Other	0.43 (0.31–0.61)	45	0.57 (0.43–0.75)	75	0.94 (0.73–1.21)	99	1.00	150
All causes	0.60 (0.54–0.66)	585	0.70 (0.64–0.76)	897	0.87 (0.80–0.95)	852	1.00	1368

Noe: The odds ratios are from a logistic regression model with age in 10-year age groups and ethnicity dichotomised as Maori and Pacific Island, and the Rest. Numbers of deaths are random rounded to the nearest multiple of three as per SNZ protocol, but odds ratios are calculated with exact data.

‡ Only age-group 45–64 included in analysis.
** Number of deaths all small resulting in release by SNZ as 6 in each cell (ie, the smallest possible number released by SNZ in the NZCMS).

Cause-specific mortality odds ratios are presented in Tables 40 and 41. The income gradients by cause of death are remarkably similar to those for education (Tables 31 and 32) and small area deprivation (Tables 28 and 29) – particularly small area deprivation.

However, one general observation is noteworthy – the similarity of the income mortality gradient by cause of death comparing males and females. For example, the income mortality gradients for lung cancer, cardiovascular disease, respiratory, unintentional injury, and suicide are very similar for males and females. This finding supports the conclusion of other researchers (Koskinen and Martelin 1994) that the weaker socioeconomic gradient usually observed for female all-cause mortality is a function of the different causes of death for males and females. That is, females are more likely to die of cancer which has a weaker gradient for both sexes.

8.2 Possible impact of bias and conclusion

There was a strong univariate association of equivalised household income with all-cause mortality, and most specific causes of mortality. These associations were probably not greatly affected by selection bias, and somewhat underestimated due to linkage bias (see Appendix C).

The two major difficulties for the income analyses were determining: a) whether health selection affected the income–mortality gradients, and b) what was causing the income–mortality gradients to decrease dramatically following the exclusion of the non-active labour force.

Theoretically, there are two possible types of health selection: *drift* health selection and *differential* health selection. Further, we would only expect health selection to operate for causes of death where a period of poor health is common before death (eg, cancer and cardiovascular disease). Finally, health selection has been framed as being a bias over the short-term (ie, a couple of years). The mortality risk plots over time in Appendix C found some evidence of *differential* health selection for the association of income with cancer among males, but not for the association of income with other causes of death. Regarding *drift* health selection, there was some occasional evidence from the mortality risk plots and exclusions of sickness beneficiaries and pre-hospitalised deaths, but it was patchy, inconsistent, and not usually notably different from the 'baseline' NZDep91 and education analyses. Thus, the tests of health selection did not strongly suggest health selection *over the short term*.

How can this conclusion that the income–mortality association was little affected by health selection be reconciled with the larger reductions in the income–mortality gradient following exclusion of the non-active labour force than for NZDep91 and highest qualification?

There are five possible ways that excluding the non-active labour force might decrease the income–mortality association:

1 *Drift health selection, whereby labour force status is a proxy for health status.* While the above sensitivity analyses suggested little drift health selection by income in the short-term, it may be that:

- the 'tests' for short-term drift health selection used were too crude
- drift health selection was acting over a longer period than the three years observable in the NZCMS
- while short-term drift health selection of the income–mortality association was modest, it was enough in combination with the reasons listed below to drive the income–mortality association further to the null that the associations of other socioeconomic factors with mortality.

2 *Differential health selection, whereby labour force status is a proxy for health status.* This may also have been one reason, but certainly not the major reason as discussed above. Also, differential health selection, if present, should apply to all socioeconomic factors – not just income.

3 *Confounding by labour force status* (by means other than short-term drift health selection). Such confounding was undoubtedly occurring, and for good reason given the strong correlation between labour force status and income. But if confounding was the major reason for the reduction of the income-gradient, more notable reductions would have been expected for the injury (and suicide) gradients – not just for the cancer and cardiovascular disease gradients.

4 *As a proxy for health status – an intermediary variable between income and mortality.* This mechanism is different from drift health selection, where health status actually influences the socioeconomic factor of interest (ie, reverse causation). Rather, this possible mechanism is common to all socioeconomic factors, and involves over-controlling the association of a socioeconomic factor with mortality by the inclusion of a proxy for an intermediary variable (ie, health status. It is not clear why this mechanism would reduce the income gradient more than, say, the education gradient).

5 *Effect modification of the income–mortality gradient by labour force status* by mechanisms other than differential health selection, such that the gradient is weaker among the active labour force. It is not clear why this mechanism should be more important for income than other socioeconomic factors.

On balance, and putting aside the fifth reason, it seems likely that a *combination* of the first four reasons explains why the income gradient diminishes dramatically when excluding the non-active labour force. There was a similar pattern of varying reduction in the cause-specific mortality gradients by NZDep91 (Appendix C) – although not as marked as that for income. Thus, a moderate variation in the mix of the first four reasons between NZDep91 and income may have been enough to make the difference. It is interesting to speculate that a modest amount of drift health selection for income, either over a period longer than three years or simply not reliably detected by the sensitivity analyses in this report, may be enough to make the difference between the NZDep91 and income analyses. However, it is impossible to be more precise in drawing conclusions.

9 Multivariate results

Education, income, labour force status and car access were included as independent variables in the multivariate analyses. Housing tenure was not included due to the likely substantive linkage bias, and occupational class was not included due to both health selection effects and many missing observations. The multivariate results for all-cause mortality are better understood after considering the cause-specific results, because different patterns were observed with different causes of death.

An initial framework or causal diagram of the association of education, income, labour force status and car access with mortality was presented earlier in this report (Figure 3). However, for the reasons discussed in the previous section, controlling for labour force status in multivariate analyses makes interpretation difficult – particularly for interpreting the association of income with mortality. There are some advantages controlling for labour force status in the assessment of the socioeconomic mortality gradients, including controlling for confounding by labour force status and (for income) adjusting for drift health selection. Regarding the disadvantages, controlling for labour force status may be over-control due to labour force status being a proxy for health status (but this should be true for all socioeconomic factors), and differential health selection may cause a underestimate of the income–mortality gradient. Accordingly, labour force status has been included in final 'exploratory' models. Otherwise, the three remaining socioeconomic factors were included in the order suggested by the framework in Figure 3: education, car access, and household income.

Linkage and selection biases have been quantified as best as possible for the univariate results in this report (Appendix). How such biases will affect the multivariate results is uncertain. It might be expected that as more socioeconomic factors are introduced, the biases for each socioeconomic factor would become less substantial (as the bias is distributed across a greater number of socioeconomic factors). Perhaps more importantly the relative changes in the association of each socioeconomic factor with mortality following the inclusion of other socioeconomic factors would not be expected to be substantially affected by linkage and selection biases. Thus, it is assumed that it is possible to make unbiased (ie, free of linkage and selection bias) comparisons of the relative association of each socioeconomic factor with mortality following stepwise inclusion of other socioeconomic factors.

Misclassification bias of socioeconomic factors would have caused underestimation of the univariate results presented in previous sections. In the multivariate analyses, misclassification bias also reduces the ability to control for confounding/mediation by variables. Quantitative sensitivity analyses have not been conducted into such effects of misclassification bias in the multivariate analyses in the NZCMS, but qualitative interpretations have been made regarding the impact of misclassification bias on the results reported below.

9.1 Cause-specific mortality

The multivariate results for the four aggregated causes of death (cancer, cardiovascular disease, unintentional injury and suicide) are presented in Tables 42–45 below. Analyses were conducted separately by sex, but 25–44 and 45–64 year olds were combined.

The first column in each table merely reproduces the univariate results reported previously in this report for education and household income. Cause-specific results for labour force status were previously presented using a five-category variable; here they are reported for a three-category variable (employed, unemployed and non-labour force). Cause-specific mortality results for car access were not previously reported. In addition to the odds ratios by category, relative indices of inequality (RIIs) for income and education are also presented. The rationale, method, and interpretation of the RII was described in Section 4.3.1 of the Methods chapter. Briefly, the RII is an estimate of the relative risk of mortality for the hypothetical person of lowest socioeconomic position compared to the hypothetical person of highest socioeconomic position. Using RIIs assumes socioeconomic factors simply rank individuals in society, and that this rank position is linearly associated with mortality risk. This latter assumption was well satisfied for income, but not strictly correct for education where the association was more of a stepped function: low mortality risk for tertiary educated, high mortality risk for nil education, and intermediary for the rest. Thus, the RII for the univariate association of education must be treated with caution, however it does assist comparison between models of how much of the association of education with mortality is 'explained' by controlling for other socioeconomic factors. It is inappropriate to calculate RIIs for labour force status as it is not an ordinal categorical variable. Regarding car access, RIIs were not calculated as the association was not linear against ranking in society.

The second column in each of Tables 42–45 presents the odds ratios and RIIs for income and education, controlled for each other. The third column presents the results for income, education and car access controlled for each other. Finally, the fourth column presents results from models also including labour force status.

9.1.1 Cancer

The multivariate results for cancer mortality are presented in Table 42 below.

9.1.1.1 Education as exposure of interest

There was a modest univariate association of education with cancer mortality, with an RII of 1.49 for males and 1.32 for females. Previous analyses in this report suggested that much of this association was due to lung cancer (Tables 31 and 32). The RII among males was reduced by 43% following control for the possible confounder/intermediary income, but did not change substantially following further control for car access and labour force status. As income and car access (and material factors as an underlying construct) were undoubtedly measured with error, the percentage attribution of the education-mortality association to mediation/confounding by these two variables would have actually been greater than 43%. The RII for females reduced somewhat following control for each of the other socioeconomic factors – but the association was modest to start with.

9.1.1.2 Household income as exposure of interest

There was a strong and monotonic univariate association of income with male cancer mortality, with an RII of 1.94. Previous analyses in this report found that the variables education and small area deprivation were only notably associated with lung cancer mortality, whereas income was also notably associated with male prostate cancer (Table 40). The association of income with male cancer mortality is only modestly reduced following control for education (RII 1.82) and then car access (RII 1.71), but weakens substantially following the control for labour force status (RII 1.19). This latter RII is unreliable due to a non-linear pattern of the odds ratios, making it more sensible to conclude that there is no residual association of income with cancer mortality. For females, the univariate association is weaker (RII 1.25), and is not dramatically changed following control for labour force status. As labour force status is a proxy for health status and males would often have been the major income earners in the household, the reduction for the male income–mortality association (but less so for females) was suggestive of health selection.

Considering both the sensitivity analyses and the univariate results presented in Appendix C of this report together with the multivariate results, it seems reasonable to conclude that:

- the strong univariate association of income with male cancer mortality is in part due to health selection
- any residual association of income with cancer after allowing for health selection:
  - is probably mostly due to the association with lung cancer
  - is in part confounded by education (a prior variable in the causal model).

Table 42: Cancer mortality univariate and multivariate odds ratios (95% CI) for 25–64 year olds in the restricted cohort

	Socioeconomic variables included in logistic regression model							
	Univariate only (eg, income only)		Income and education		Income, education, car		All four socioeconomic factors	
Males								
Tertiary	0.74	(0.65–0.85)	0.83	(0.72–0.96)	0.84	(0.73–0.96)	0.84	(0.73–0.97)
Trade, other tertiary	0.79	(0.71–0.88)	0.82	(0.73–0.91)	0.82	(0.74–0.92)	0.83	(0.75–0.93)
School	0.84	(0.74–0.96)	0.88	(0.77–1.00)	0.88	(0.77–1.00)	0.88	(0.77–1.00)
Nil	1.00		1.00		1.00		1.00	
RII for education	*1.49*		*1.28*		*1.27*		*1.26*	
≥ $50,000	0.61	(0.53–0.69)	0.64	(0.55–0.73)	0.67	(0.58–0.77)	0.91	(0.78–1.06)
$30,000–$49,999	0.81	(0.72–0.91)	0.82	(0.73–0.92)	0.86	(0.76–0.97)	1.10	(0.96–1.25)
$20,000–$29,999	0.89	(0.78–1.00)	0.89	(0.79–1.01)	0.92	(0.81–1.04)	1.05	(0.92–1.19)
<$20,000	1.00		1.00		1.00		1.00	
RII for income	*1.94*		*1.82*		*1.71*		*1.19*	
>2 cars	1.00				1.00		1.00	
1 car	1.19	(1.09–1.31)			1.10	(1.00–1.21)	1.06	(0.96–1.16)
Nil cars	1.48	(1.23–1.78)			1.27	(1.05–1.54)	1.15	(0.95–1.39)
Employed	1.00						1.00	
Unemployed	1.24	(0.99–1.56)					1.20	(0.95–1.51)
Non-labour force	2.01	(1.82–2.23)					1.92	(1.72–2.15)
Females								
Tertiary	0.81	(0.72–0.92)	0.83	(0.73–0.95)	0.84	(0.73–0.95)	0.88	(0.77–1.00)
Trade, other tertiary	0.88	(0.75–1.02)	0.89	(0.76–1.04)	0.89	(0.76–1.04)	0.93	(0.80–1.08)
School	0.85	(0.75–0.95)	0.86	(0.76–0.96)	0.86	(0.76–0.96)	0.87	(0.77–0.98)
Nil	1.00		1.00		1.00		1.00	
RII for education	*1.32*		*1.27*		*1.26*		*1.17*	
≥ $50,000	0.86	(0.75–0.98)	0.90	(0.78–1.03)	0.94	(0.81–1.08)	1.08	(0.93–1.26)
$30,000-$49,999	0.91	(0.81–1.02)	0.93	(0.82–1.04)	0.96	(0.85–1.09)	1.08	(0.95–1.23)
$20,000-$29,999	0.99	(0.88–1.13)	1.00	(0.88–1.13)	1.03	(0.90–1.16)	1.09	(0.96–1.24)
<$20,000	1.00		1.00		1.00		1.00	
RII for income	*1.25*		*1.18*		*1.11*		*0.91*	
>2 cars	1.00				1.00		1.00	
1 car	1.10	(1.00–1.21)			1.08	(0.97–1.19)	1.07	(0.97–1.18)
Nil cars	1.22	(1.03–1.45)			1.16	(0.97–1.39)	1.13	(0.94–1.36)
Employed	1.00						1.00	
Unemployed	0.88	(0.63–1.22)					0.88	(0.63–1.22)
Non-labour force	1.52	(1.37–1.68)						

Note: All odds ratios are adjusted for ethnicity and 10-year age group.

9.1.1.3 Car access as exposure of interest

The association of car access, as a proxy for asset wealth, with cancer was modest. The univariate odds ratio for nil car access compared to two or more cars for males is 1.48 (1.23 to 1.78), and reduces to 1.15 (0.95 to 1.39) controlling for other socioeconomic factors. For females, the odds ratio reduced from 1.22 (1.03 to 1.45) to 1.13 (0.94 to 1.36). Presuming that car access, and underlying asset wealth, are influenced by education and income (rather than the other way around), then the unconfounded association of car access with cancer mortality appears negligible.

9.1.1.4 Labour force status as exposure of interest

Unemployed males had a 24% greater univariate odds of cancer mortality than employed males, although the 95% confidence interval about the odds ratio just included 1.0 (0.99 to 1.56). Previous analyses in this PhD found that this elevated odds of cancer mortality among unemployed males was due to lung cancer mortality (Table 36). Relatively little of this association was explained by confounding/mediation by income, education and car access (OR 1.20, 95% CI 0.95 to 1.51). There was no excess univariate odds of cancer mortality among unemployed females (OR 0.88, 0.63 to 1.22), and the odds ratio did not alter following control for the other socioeconomic factors.

Males among the non-labour force had twice the univariate odds of cancer mortality than employed males (OR 2.01, 1.82 to 2.23), and females among the non-labour force had an odds ratio of 1.52 (1.37 to 1.68). These odds ratios changed little following control for other socioeconomic factors, and almost certainly represent drift health selection; people with terminal cancer tend to move out of the labour force prior to death.

9.1.2 Cardiovascular disease

The results for cardiovascular disease mortality are presented in Table 43 below.

9.1.2.1 Education as exposure of interest

The univariate association of highest qualification with cardiovascular disease mortality was stronger among females than males, with RIIs of 2.94 and 1.90 respectively. Mediation by income and car access explained a third to half of this association, and would have presumably been a greater fraction if not for inevitable measurement error of income and car access. Control for labour force status (a proxy for the intermediary variable health status and other pathway mechanisms in this instance) did not alter the association for males and only reduced the gradient modestly for females.

Table 43: Cardiovascular disease mortality univariate and multivariate odds ratios (95% CI) for +25–64 year olds in the restricted cohort

	Socioeconomic variables included in logistic regression model							
	Univariate only (eg, income only)		Income and education		Income, education, car		All four socioeconomic factors	
Males								
Tertiary	0.57	(0.49–0.65)	0.68	(0.59–0.79)	0.70	(0.60–0.80)	0.70	(0.61–0.81)
Trade, other tertiary	0.80	(0.72–0.88)	0.85	(0.77–0.94)	0.87	(0.79–0.96)	0.88	(0.80–0.98)
School	0.70	(0.62–0.80)	0.75	(0.66–0.85)	0.75	(0.66–0.86)	0.76	(0.67–0.86)
Nil	1.00		1.0		1.0		1.0	
RII for education	*1.90*		*1.51*		*1.46*		*1.44*	
≥ $50,000	0.43	(0.38–0.49)	0.47	(0.41–0.54)	0.54	(0.47–0.61)	0.75	(0.64–0.87)
$30,000–$49,999	0.63	(0.56–0.70)	0.65	(0.58–0.72)	0.72	(0.64–0.80)	0.94	(0.83–1.06)
$20,000–$29,999	0.78	(0.70–0.87)	0.79	(0.71–0.88)	0.85	(0.76–0.95)	0.98	(0.88–1.10)
<$20,000	1.0		1.0		1.0		1.0	
RII for income	*3.11*		*2.72*		*2.32*		*1.54*	
>2 cars	1.0				1.0		1.0	
1 car	1.35	(1.24–1.47)			1.18	(1.08–1.29)	1.13	(1.03–1.24)
Nil cars	2.42	(2.09–2.79)			1.90	(1.64–2.21)	1.70	(1.46–1.98)
Employed	1.0						1.0	
Unemployed	1.42	(1.15–1.74)					1.19	(0.97–1.48)
Non-labour force	2.45	(2.23–2.68)					2.05	(1.85–2.27)
Females								
Tertiary	0.49	(0.40–0.61)	0.59	(0.48–0.74)	0.60	(0.48–0.75)	0.66	(0.53–0.82)
Trade, other tertiary	0.54	(0.41–0.70)	0.59	(0.45–0.77)	0.59	(0.45–0.77)	0.64	(0.49–0.84)
School	0.74	(0.62–0.87)	0.78	(0.66–0.93)	0.79	(0.66–0.93)	0.81	(0.68–0.96)
Nil	1.00		1.0		1.0		1.0	
RII for education	*2.94*		*2.20*		*2.17*		*1.87*	
≥ $50,000	0.38	(0.30–0.47)	0.43	(0.34–0.54)	0.51	(0.40–0.64)	0.67	(0.52–0.86)
$30,000-$49,999	0.55	(0.46–0.65)	0.58	(0.49–0.69)	0.65	(0.54–0.78)	0.81	(0.67–0.98)
$20,000-$29,999	0.81	(0.69–0.95)	0.82	(0.70–0.97)	0.89	(0.75–1.05)	0.98	(0.83–1.16)
<$20,000	1.0		1.0		1.0		1.0	
RII for income	*4.16*		*3.33*		*2.65*		*1.81*	
>2 cars	1.0				1.0		1.0	
1 car	1.63	(1.41–1.88)			1.37	(1.17–1.59)	1.34	(1.15–1.56)
Nil cars	2.05	(1.65–2.55)			1.51	(1.20–1.90)	1.43	(1.14–1.81)
Employed	1.0						1.0	
Unemployed	1.16	(0.70–1.93)					0.94	(0.56–1.57)
Non-labour force	2.68	(2.28–3.15)					2.12	(1.78–2.52)

Note: All odds ratios are adjusted for ethnicity and 10-year age group.

9.1.2.2 Household income as exposure of interest

The univariate association of household income with cardiovascular disease was one of the strongest associations in this report, with RIIs of 3.11 and 4.16 for males and females, respectively. Controlling for education accounted for about 20–25% of this univariate association. Control for confounding/mediation by both education and car access explained about a third of the income gradient. Additional control for labour force status (a proxy for pathway mechanisms and any health selection) reduced the RII to 1.54 for males and 1.81 for females. Assuming that some of the reduction following the control for labour force status was over-control (eg, any differential health selection component and as a proxy for health status), then it appears that household income among adults has a strong residual association with cardiovascular disease – either controlling for genuine confounding by education and labour force status or after additional control for confounding/mediation by car access.

9.1.2.3 Car access as exposure of interest

The association of car access with cardiovascular disease was also notably strong, confirming the role of material factors as strong predictors of cardiovascular disease. Odds ratios for nil car access compared to two or more were 2.42 (2.09 to 2.79) for males and 2.05 (1.65 to 2.55) for females. Control for potential confounding by education and income reduced this association by up to 50%, but it was not reduced much further following additional control for labour force status.

9.1.2.4 Labour force status as exposure of interest

The association of non-labour force status with cardiovascular disease was also strong, with univariate odds ratios of 2.45 (2.23 to 2.68) for males and 2.68 (2.28 to 3.15) for females. Controlling for possible confounding by education and possible confounding/ mediation by income and car access, these associations only reduced by a quarter to a third. While there will be residual confounding by the previous socioeconomic factors and other variables, it seems likely that much of the remaining association was due to health selection.

A 42% excess odds of cardiovascular disease mortality among unemployed males compared to employed males appeared to be largely due to confounding/mediation by other socioeconomic factors. Both the univariate and multivariate odds ratios of female cardiovascular disease mortality were not significantly different from 1.0. It is possible that the association of unemployment with cardiovascular disease was underestimated by linkage bias. Also, the results apply only to those unemployed in the four weeks preceding census night, causing a likely underestimate of the association for the long-term unemployed. Despite these caveats, a larger multivariate association of unemployment with cardiovascular mortality would have been expected if psychosocial mechanisms (rather than material mechanisms and confounding by socioeconomic factors) were a major mediating mechanism between unemployment and cardiovascular disease mortality.

9.1.3 Unintentional injury

The results for unintentional injury mortality are presented in Table 44 below.

Table 44: Unintentional injury mortality univariate and multivariate odds ratios (95% CI) for 25–64 year olds in the restricted cohort

	Socioeconomic variables included in logistic regression model			
	Univariate only (eg, income only)	**Income and education**	**Income, education, car**	**All four socioeconomic factors**
Males				
Tertiary	0.51 (0.36–0.71)	0.61 (0.43–0.87)	0.61 (0.43–0.87)	0.61 (0.43–0.87)
Trade, other tertiary	0.75 (0.58–0.97)	0.80 (0.62–1.04)	0.80 (0.61–1.04)	0.80 (0.62–1.04)
School	0.79 (0.59–1.06)	0.86 (0.64–1.15)	0.86 (0.64–1.15)	0.86 (0.64–1.15)
Nil	1.00	1.0	1.0	1.0
RII for education	*2.37*	*1.86*	*1.88*	*1.87*
≥ $50,000	0.43 (0.31–0.60)	0.49 (0.35–0.68)	0.48 (0.34–0.67)	0.49 (0.34–0.71)
$30,000–$49,999	0.72 (0.54–0.94)	0.75 (0.57–0.99)	0.74 (0.56–0.98)	0.76 (0.56–1.02)
$20,000–$29,999	0.91 (0.69–1.21)	0.93 (0.70–1.24)	0.92 (0.69–1.22)	0.93 (0.70–1.25)
<$20,000	1.0	1.0	1.0	1.0
RII for income	*3.35*	*2.83*	*2.90*	*2.78*
>2 cars	1.0		1.0	1.0
1 car	1.14 (0.92–1.42)		0.99 (0.79–1.24)	0.99 (0.79–1.23)
Nil cars	1.07 (0.67–1.71)		0.82 (0.50–1.32)	0.80 (0.49–1.30)
Employed	1.0			1.0
Unemployed	1.25 (0.84–1.85)			1.01 (0.67–1.52)
Non-labour force	1.42 (1.06–1.91)			1.14 (0.83–1.56)
Females				
Tertiary	0.68 (0.41–1.13)	0.84 (0.50–1.42)	0.83 (0.49–1.40)	0.86 (0.51–1.46)
Trade, other tertiary	0.92 (0.54–1.58)	1.03 (0.60–1.77)	1.02 (0.60–1.76)	1.07 (0.62–1.84)
School	0.83 (0.53–1.30)	0.90 (0.57–1.42)	0.89 (0.57–1.40)	0.91 (0.58–1.43)
Nil	1.00	1.0	1.0	1.0
RII for education	*1.64*	*1.22*	*1.24*	*1.17*
≥ $50,000	0.38 (0.22–0.67)	0.40 (0.23–0.71)	0.40 (0.22–0.72)	0.43 (0.23–0.81)
$30,000–$49,999	0.56 (0.35–0.88)	0.57 (0.36–0.90)	0.56 (0.35–0.90)	0.60 (0.36–0.99)
$20,000–$29,999	0.67 (0.42–1.07)	0.67 (0.42–1.08)	0.66 (0.41–1.07)	0.69 (0.42–1.12)
<$20,000	1.0	1.0	1.0	1.0
RII for income	*3.71*	*3.47*	*3.52*	*3.05*
>2 cars	1.0		1.0	1.0
1 car	1.30 (0.90–1.89)		1.05 (0.71–1.55)	1.05 (0.71–1.55)
Nil cars	1.19 (0.59–2.38)		0.80 (0.39–1.66)	0.78 (0.38–1.63)
Employed	1.0			1.0
Unemployed	0.62 (0.19–2.00)			0.48 (0.15–1.56)
Non-labour force	1.73 (1.19–2.53)			1.36 (0.90–2.05)

Note: All odds ratios are adjusted for ethnicity and 10-year age group.

9.1.3.1 Education as exposure of interest

The univariate association of highest qualification with unintentional injury deaths was stronger for males (RII 2.37) than for females (RII 1.64). Among males, the RII reduced to 1.86 controlling for income, but did not reduce any further following control for car access and labour force status. Among females, the RII reduced to 1.22 after controlling for income, and did not alter much thereafter. Thus, income and material factors appear to mediate/confound much of the association of education with female injury mortality, but among males education still appears to have a strong and persistent association with injury mortality via mechanisms other than material factors and labour force status.

9.1.3.2 Household income as exposure of interest

The univariate association of income with injury mortality was strong for both males and females (RIIs of 3.35 and 3.71, respectively). Up to 25% of this association was due to confounding by education. Additional control for car access and labour force status did not further reduce the association, except for a 20% marginal reduction for labour force status among females. Two important conclusions are suggested. First, income (and by extension material factors) are strongly associated with injury deaths, independent of any confounding by education. This is consistent with the time lag between socioeconomic position and injury being shorter than for other causes of death and (consequently) the association of a socioeconomic factors measured recently (ie, income in the last year) being more strongly association with the outcome than, say, education.

Second, for each other cause of death the association of income with mortality reduced notably following control for labour force status – but not for injury deaths. This 'exception to the rule' is suggestive of health selection (either drift or differential) affecting income for the three other causes of death, but not for injury deaths.

9.1.3.3 Car access as exposure of interest

There was no significant univariate or multivariate association of car access with injury mortality. If car access was only a proxy for asset wealth, then these results seem inconsistent with the strong association with income. One possible explanation is that road traffic crash fatalities are positively associated with the number of cars in the household, negating the inverse association due to car access being a measure of asset wealth.

9.1.3.4 Labour force status as exposure of interest

There were elevated univariate odds of injury for the non-active labour force compared to the active labour force for both males and females (odds ratios of 1.42 (1.06 to 1.91) and 1.73 (1.19 to 2.53), respectively) and for unemployed males compared to employed males (1.25 (0.84 to 1.85)). These odds ratios all substantially reduced following control for the three other potential mediating/confounding socioeconomic factors, with none of the 95% confidence intervals excluding 1.0. Of the four broad causes of disease, injury was the only one where the multivariate odds ratios for the non-active labour force were close to 1.0 with 95% confidence intervals including 1.0. This finding further reinforces the role of health selection in explaining the generally elevated mortality risk among the non-active labour force for causes of death with a likelihood of prior ill health.

9.1.4 Suicide

The results for suicide are presented in Table 45 below. The results for suicide are based on relatively few deaths, particularly for females (87 for females, and 282 for males). Thus, the results were somewhat imprecise with wide 95% confidence intervals.

Table 45: Suicide mortality univariate and multivariate odds ratios (95% CI) for 25–64 year olds in the restricted cohort

	Socioeconomic variables included in logistic regression model			
	Univariate only (eg, income only)	**Income and education**	**Income, education, car**	**All four socioeconomic factors**
Males				
Tertiary	0.56 (0.38–0.82)	0.66 (0.45–0.98)	0.67 (0.45–1.00)	0.71 (0.48–1.05)
Trade, other tertiary	0.89 (0.66–1.20)	0.96 (0.72–1.30)	0.99 (0.73–1.34)	1.05 (0.77–1.41)
School	1.06 (0.77–1.46)	1.15 (0.83–1.58)	1.16 (0.84–1.60)	1.21 (0.88–1.67)
Nil	1.00	1.0	1.0	1.0
RII for education	*2.04*	*1.68*	*1.63*	*1.53*
≥ $50,000	0.44 (0.32–0.62)	0.48 (0.34–0.68)	0.56 (0.39–0.81)	0.86 (0.58–1.28)
$30,000–$49,999	0.57 (0.42–0.78)	0.59 (0.43–0.80)	0.66 (0.48–0.90)	0.95 (0.68–1.34)
$20,000–$29,999	0.66 (0.48–0.93)	0.67 (0.48–0.94)	0.72 (0.52–1.01)	0.91 (0.65–1.29)
<$20,000	1.0	1.0	1.0	1.0
RII for income	*2.79*	*2.44*	*1.97*	*1.18*
>2 cars	1.0		1.0	1.0
1 car	1.56 (1.22–1.99)		1.39 (1.08–1.80)	1.33 (1.03–1.72)
Nil cars	2.37 (1.50–3.72)		1.88 (1.17–3.00)	1.49 (0.93–2.39)
Employed	1.0			1.0
Unemployed	2.70 (1.84–3.95)			2.35 (1.57–3.53)
Non-labour force	3.29 (2.45–4.43)			2.86 (2.06–3.97)
Females				
Tertiary	1.18 (0.68–2.06)	1.37 (0.77–2.43)	1.46 (0.82–2.59)	1.61 (0.90–2.87)
Trade, other tertiary	0.87 (0.42–1.80)	0.94 (0.46–1.96)	0.99 (0.48–2.06)	1.06 (0.51–2.20)
School	1.17 (0.68–2.01)	1.24 (0.72–2.14)	1.31 (0.76–2.26)	1.36 (0.79–2.36)
Nil	1.00	1.0	1.0	1.0
RII for education	*0.89*	*0.72*	*0.66*	*0.57*
≥ $50,000	0.50 (0.27–0.91)	0.46 (0.25–0.86)	0.60 (0.31–1.16)	0.92 (0.46–1.86)
$30,000–$49,999	0.63 (0.37–1.07)	0.61 (0.36–1.03)	0.76 (0.43–1.33)	1.08 (0.59–1.95)
$20,000–$29,999	0.49 (0.26–0.93)	0.48 (0.25–0.92)	0.58 (0.30–1.11)	0.69 (0.35–1.34)
<$20,000	1.0	1.0	1.0	1.0
RII for income	*2.16*	*2.42*	*1.65*	*0.90*
>2 cars	1.0		1.0	1.0
1 car	1.31 (0.82–2.07)		1.20 (0.74–1.94)	1.18 (0.73–1.92)
Nil cars	3.69 (1.96–6.95)		3.16 (1.58–6.30)	2.87 (1.43–5.79)
Employed	1.0			1.0
Unemployed	2.86 (1.19–6.85)			2.58 (1.04–6.38)
Non-labour force	2.77 (1.74–4.43)			2.66 (1.60–4.44)

Note: All odds ratios are adjusted for ethnicity and 10-year age group.

9.1.4.1 Education as exposure of interest

There was an approximately twofold excess univariate odds of suicide death by education for males in the expected direction (RII 2.04), but a null association among females. Controlling for the potential mediators/confounders of income and car access reduced the association by approximately a third for males, and actually introduced a reverse (albeit 'non-statistically significant') gradient among females. Among males, additional control for labour force status essentially reduced the association to the null. Thus, education does not appear to have a particularly strong association with suicide.

9.1.4.2 Household income as exposure of interest

The univariate association of income with suicide was strong for both males (RII 2.79) and females (RII 2.16). Control for the potential confounder education did not notably change the association, but control for the intermediary/confounder car access reduced the gradient (RIIs 1.97 and 1.65 for males and females, respectively). Further control for the labour force status reduced the associations to the null for both sexes. Thus, the suspicion of health selection as one reason for the strong association of income with suicide is raised. However, the multivariate association of labour force status with suicide is so strong and persistent (discussion to come) that particularly strong confounding of the income association by labour force status is also a distinct possibility.

9.1.4.3 Car access as the exposure of interest

Car access was strongly associated with suicide death in univariate and multivariate (particularly females) analyses. Given that controlling for labour force status did not greatly reduce the car access association, it does not appear that nil car access is simply a proxy for chronic mental illness that would also be associated with labour force status. In addition to being a proxy for material factors, it may be that car access here is acting as a proxy for social isolation as an independent risk factor for suicide.

9.1.4.4 Labour force status as the exposure of interest

Unemployment

Being unemployed was strongly associated with an excess suicide risk compared to employed people in univariate analyses among both males (odds ratio 2.70, 1.84 to 3.95) and females (2.86, 1.19 to 6.85). Controlling for the potential confounders/mediators of income, car access and education did not substantially reduce these associations (odds ratios 2.35 and 2.58 for males and females respectively). Plots of mortality risk over time in Appendix C of this report found evidence of a 'healthy worker' effect among the unemployed, just as for the employed (Figure 36). Also, the definition of unemployment in the 1991 census requires the person to be both actively looking for work and available for work. Thus, it seems unlikely that the elevated suicide risk among the unemployed is a consequence of health selection. Of the four broad causes of death, suicide was the only one to be strongly associated with unemployment following the multivariate analyses. Given this pattern, and the lack of reduction of the association controlling for other socioeconomic factors, *these results suggest a strong causal association of unemployment with suicide via, perhaps, psychosocial mechanisms.*

Assuming that unemployment is strongly and causally associated with suicide, how important is this in population terms? The numbers of suicide deaths by labour force category shown in Table 36 show that about 33 male and six female suicide deaths occurred among the unemployed. (These numbers are rounded to a multiple of three to meet SNZ privacy requirements. For the NZCMS, there was also a further requirement that the minimum cell value released six – thus one death in a cell would be rounded to six. However, exact data was used in the logistic regression analyses.) In contrast, 168 male and 36 female suicide deaths occurred among the employed.

Non-active labour force

The non-active labour force had a very similar pattern to the unemployed of high univariate and multivariate odds ratios of suicide death. Some of this association was most likely due to health selection. However, some of the association may have also been due to the same mechanism as that among the unemployed – perhaps psychosocial mechanisms. In population terms, this would be important as 84 male and 51 female suicide deaths occurred among the non-active labour force.

9.2 All-cause mortality

The all-cause results are basically a weighted average of the cause-specific analyses. The cause-specific analyses are more useful for understanding likely causal mechanisms and bias. The added value of presenting the all-cause multivariate analyses is:

- to obtain a summary picture
- to determine any differences between 25–44 and 45–64 year olds, as there were enough deaths for all-causes to conduct multivariate analyses for both age-groups
- to examine socioeconomic mortality gradients at a finer level, eg, nine levels of household income as opposed to four.

The all-cause analyses are presented below in Tables 46 and 47.

Table 46: All-cause mortality univariate and multivariate odds ratios (95% CI) in the restricted cohort – males

	Socioeconomic variables included in logistic regression model			
	Univariate only (eg, income only)	**Income and education**	**Income, education, car**	**All four socioeconomic factors**
25–44 year olds				
Full-time employed	1.00			1.00
Part-time employed	1.75 (1.31–2.32)			1.65 (1.24–2.20)
Unemployed	1.45 (1.15–1.82)			1.26 (0.99–1.61)
Seeking or available †	2.00 (1.38–2.91)			1.70 (1.16–2.50)
Retired, etc	3.24 (2.72–3.86)			2.76 (2.27–3.35)
Graduate, postgrad	0.47 (0.36–0.61)	0.55 (0.42–0.72)	0.56 (0.43–0.73)	0.59 (0.45–0.77)
Undergraduate, etc	0.67 (0.53–0.84)	0.74 (0.59–0.93)	0.76 (0.60–0.96)	0.82 (0.65–1.03)
Trade, other tertiary	0.74 (0.63–0.87)	0.79 (0.67–0.92)	0.81 (0.69–0.95)	0.87 (0.74–1.02)
11–12 years of school	0.73 (0.58–0.93)	0.80 (0.63–1.02)	0.82 (0.64–1.04)	0.86 (0.68–1.10)
10 years of school	0.75 (0.60–0.94)	0.79 (0.63–0.99)	0.81 (0.65–1.02)	0.87 (0.69–1.09)
Other school qualification	0.57 (0.30–1.07)	0.59 (0.31–1.11)	0.59 (0.31–1.10)	0.60 (0.32–1.12)
Nil	1.00	1.00	1.00	1.00
RII for education ‡	*2.08*	*1.78*	*1.74*	*1.63*
≥ $70,000	0.73 (0.56–0.95)	0.82 (0.63–1.07)	0.85 (0.65–1.11)	0.88 (0.68–1.15)
$50,000–$69,999	0.83 (0.66–1.06)	0.88 (0.70–1.12)	0.90 (0.71–1.15)	0.93 (0.73–1.18)
$40,000–$49,999	0.83 (0.65–1.05)	0.85 (0.67–1.07)	0.86 (0.68–1.09)	0.88 (0.69–1.11)
$30,000–$39,999	1.00	1.00	1.00	1.00
$25,000–$29,999	1.12 (0.87–1.44)	1.10 (0.85–1.41)	1.07 (0.84–1.38)	1.03 (0.80–1.32)
$20,000–$24,999	1.14 (0.91–1.43)	1.11 (0.89–1.39)	1.09 (0.87–1.37)	1.02 (0.81–1.28)
$15,000–$19,999	1.33 (1.05–1.68)	1.27 (1.01–1.61)	1.21 (0.96–1.53)	1.05 (0.82–1.33)
$10,000–$14,999	1.40 (1.09–1.79)	1.33 (1.04–1.71)	1.21 (0.94–1.56)	0.93 (0.71–1.21)
<$10,000	1.64 (1.21–2.22)	1.58 (1.17–2.14)	1.40 (1.03–1.91)	1.05 (0.76–1.45)
RII for income	*2.17*	*1.80*	*1.57*	*1.20*
>2 cars	1.00		1.00	1.00
1 car	1.31 (1.15–1.50)		1.21 (1.06–1.39)	1.19 (1.04–1.36)
Nil cars	2.21 (1.76–2.78)		1.83 (1.44–2.32)	1.53 (1.20–1.95)
45–64 year olds				
Full-time employed	1.00			1.00
Part-time employed	1.23 (1.07–1.42)			1.18 (1.02–1.35)
Unemployed	1.47 (1.26–1.71)			1.28 (1.10–1.51)
Seeking or available †	1.64 (1.31–2.05)			1.43 (1.13–1.79)
Retired, etc	2.44 (2.26–2.62)			2.10 (1.93–2.28)
Graduate, postgrad	0.55 (0.47–0.64)	0.72 (0.61–0.84)	0.72 (0.62–0.85)	0.73 (0.63–0.86)
Undergraduate, etc	0.72 (0.64–0.81)	0.81 (0.72–0.91)	0.82 (0.73–0.92)	0.82 (0.74–0.93)
Trade, other tertiary	0.80 (0.74–0.85)	0.84 (0.78–0.90)	0.86 (0.80–0.92)	0.86 (0.80–0.93)
11–12 years of school	0.81 (0.70–0.93)	0.88 (0.76–1.01)	0.88 (0.76–1.02)	0.88 (0.76–1.02)
10 years of school	0.80 (0.71–0.90)	0.84 (0.75–0.95)	0.85 (0.75–0.96)	0.86 (0.76–0.97)
Other school qualification	0.79 (0.67–0.94)	0.81 (0.69–0.96)	0.81 (0.68–0.95)	0.80 (0.68–0.95)
Nil	1.00	1.00	1.00	1.00
RII for education ‡	*1.81*	*1.44*	*1.41*	*1.40*

	Socioeconomic variables included in logistic regression model			
	Univariate only (eg, income only)	**Income and education**	**Income, education, car**	**All four socioeconomic factors**
≥ $70,000	0.60 (0.53–0.69)	0.64 (0.56–0.74)	0.67 (0.58–0.77)	0.74 (0.64–0.85)
$50,000–$69,999	0.69 (0.62–0.78)	0.71 (0.63–0.80)	0.73 (0.65–0.82)	0.79 (0.70–0.89)
$40,000–$49,999	0.86 (0.77–0.96)	0.87 (0.78–0.97)	0.88 (0.79–0.99)	0.93 (0.83–1.04)
$30,000–$39,999	1.00	1.00	1.00	1.00
$25,000–$29,999	1.10 (0.98–1.23)	1.09 (0.97–1.22)	1.07 (0.95–1.19)	1.00 (0.90–1.12)
$20,000–$24,999	1.09 (0.98–1.22)	1.08 (0.97–1.21)	1.06 (0.95–1.18)	0.93 (0.83–1.04)
$15,000–$19,999	1.35 (1.22–1.50)	1.33 (1.20–1.47)	1.24 (1.12–1.37)	1.00 (0.90–1.11)
$10,000–$14,999	1.42 (1.27–1.58)	1.39 (1.24–1.55)	1.25 (1.11–1.40)	0.97 (0.86–1.09)
<$10,000	1.23 (1.04–1.45)	1.21 (1.02–1.43)	1.08 (0.91–1.28)	0.88 (0.74–1.04)
RII for income	*2.62*	*2.37*	*2.05*	*1.43*
>2 cars	1.00		1.00	1.00
1 car	1.31 (1.24–1.40)		1.17 (1.10–1.25)	1.14 (1.07–1.21)
Nil cars	2.21 (1.99–2.45)		1.81 (1.62–2.02)	1.65 (1.47–1.84)

Note: All odds ratios are adjusted for ethnicity and five-year age group.

† The 'seeking or available' include those people that were either available for work, or seeking work, but not both. Both these requirements must be meet to be labelled as unemployed (and hence part of the active labour force).

‡ RII calculations exclude 'other school qualifications' as it is of indeterminate rank.

Table 47: All-cause mortality univariate and multivariate odds ratios (95% CI) in the restricted cohort – females

	Socioeconomic variables included in logistic regression model			
	Univariate only (eg, income only)	**Income and education**	**Income, education, car**	**All four socioeconomic factors**
25–44 year olds				
Full-time employed	1.00			1.00
Part-time employed	1.04 (0.83–1.30)			1.03 (0.82–1.29)
Unemployed	1.50 (1.07–2.10)			1.38 (0.97–1.97)
Seeking or available	2.25 (1.51–3.33)			2.03 (1.35–3.06)
Retired, etc ‡	2.00 (1.66–2.41)			1.81 (1.47–2.23)
Graduate, postgrad	0.50 (0.35–0.72)	0.55 (0.38–0.80)	0.56 (0.39–0.80)	0.58 (0.40–0.83)
Undergraduate, etc	0.59 (0.46–0.74)	0.62 (0.49–0.79)	0.63 (0.50–0.81)	0.67 (0.52–0.85)
Trade, other tertiary	0.63 (0.49–0.80)	0.65 (0.51–0.83)	0.66 (0.52–0.84)	0.69 (0.54–0.88)
11–12 years of school	0.64 (0.47–0.87)	0.67 (0.49–0.91)	0.68 (0.50–0.93)	0.70 (0.51–0.95)
10 years of school	0.55 (0.43–0.71)	0.57 (0.44–0.73)	0.58 (0.45–0.75)	0.60 (0.46–0.77)
Other school qualification	0.89 (0.55–1.44)	0.92 (0.57–1.49)	0.93 (0.57–1.50)	0.93 (0.57–1.50)
Nil	1.00	1.00	1.00	1.00
RII for education ‡	*1.94*	*1.75*	*1.72*	*1.62*
≥ $70,000	0.93 (0.67–1.30)	1.03 (0.73–1.43)	1.04 (0.74–1.45)	1.08 (0.77–1.51)
$50,000–$69,999	0.98 (0.72–1.33)	1.03 (0.76–1.40)	1.04 (0.76–1.42)	1.09 (0.80–1.48)
$40,000–$49,999	1.01 (0.75–1.37)	1.03 (0.76–1.40)	1.04 (0.77–1.41)	1.08 (0.80–1.46)
$30,000–$39,999	1.00	1.00	1.00	1.00
$25,000–$29,999	1.52 (1.10–2.08)	1.48 (1.08–2.03)	1.47 (1.07–2.02)	1.39 (1.01–1.91)
$20,000–$24,999	1.09 (0.80–1.49)	1.06 (0.77–1.44)	1.05 (0.77–1.43)	0.96 (0.70–1.31)
$15,000–$19,999	1.28 (0.94–1.74)	1.22 (0.89–1.66)	1.18 (0.86–1.61)	1.01 (0.73–1.38)
$10,000–$14,999	1.72 (1.29–2.30)	1.61 (1.20–2.15)	1.51 (1.12–2.04)	1.18 (0.86–1.61)
<$10,000	1.48 (1.01–2.16)	1.37 (0.94–1.99)	1.26 (0.86–1.86)	0.98 (0.66–1.46)
RII for income	*1.80*	*1.48*	*1.36*	*0.97*
>2 cars	1.00		1.00	1.00
1 car	1.15 (0.98–1.36)		1.05 (0.88–1.25)	1.03 (0.87–1.23)
Nil cars	1.76 (1.34–2.31)		1.36 (1.01–1.83)	1.26 (0.94–1.70)
45–64 year olds				
Full-time employed	1.00			1.00
Part-time employed	1.19 (1.04–1.36)			1.17 (1.02–1.34)
Unemployed	1.05 (0.77–1.43)			0.95 (0.70–1.30)
Seeking or available	1.73 (1.26–2.37)			1.57 (1.14–2.15)
Retired, etc ‡	2.10 (1.89–2.33)			1.91 (1.71–2.14)
Graduate, postgrad	0.75 (0.59–0.94)	0.86 (0.68–1.08)	0.85 (0.67–1.07)	0.94 (0.74–1.19)
Undergraduate, etc	0.69 (0.62–0.78)	0.75 (0.67–0.85)	0.76 (0.67–0.86)	0.81 (0.72–0.92)
Trade, other tertiary	0.79 (0.69–0.90)	0.83 (0.73–0.95)	0.83 (0.73–0.95)	0.88 (0.77–1.01)
11–12 years of school	0.81 (0.68–0.97)	0.84 (0.71–1.01)	0.85 (0.71–1.01)	0.87 (0.73–1.04)
10 years of school	0.92 (0.81–1.04)	0.95 (0.84–1.07)	0.96 (0.85–1.08)	0.99 (0.87–1.12)
Other school qualification	0.81 (0.70–0.95)	0.83 (0.71–0.97)	0.83 (0.71–0.96)	0.84 (0.72–0.97)
Nil	1.00	1.00	1.00	1.00
RII for education ‡	*1.68*	*1.46*	*1.45*	*1.29*

	Socioeconomic variables included in logistic regression model			
	Univariate only (eg, income only)	Income and education	Income, education, car	All four socioeconomic factors
≥ $70,000	0.85 (0.72–1.01)	0.89 (0.75–1.06)	0.94 (0.79–1.11)	1.00 (0.84–1.19)
$50,000–$69,999	0.79 (0.67–0.92)	0.81 (0.70–0.95)	0.83 (0.71–0.97)	0.90 (0.77–1.05)
$40,000–$49,999	0.93 (0.80–1.08)	0.94 (0.81–1.09)	0.95 (0.82–1.11)	1.01 (0.87–1.17)
$30,000–$39,999	1.00	1.00	1.00	1.00
$25,000–$29,999	1.16 (1.01–1.34)	1.15 (1.00–1.33)	1.13 (0.98–1.31)	1.09 (0.94–1.25)
$20,000–$24,999	1.19 (1.03–1.37)	1.17 (1.02–1.35)	1.15 (1.00–1.32)	1.05 (0.91–1.21)
$15,000–$19,999	1.32 (1.16–1.49)	1.29 (1.14–1.46)	1.21 (1.07–1.38)	1.07 (0.94–1.21)
$10,000–$14,999	1.32 (1.15–1.52)	1.29 (1.12–1.48)	1.18 (1.02–1.36)	1.01 (0.88–1.17)
<$10,000	1.34 (1.09–1.65)	1.32 (1.07–1.62)	1.20 (0.98–1.48)	1.06 (0.86–1.31)
RII for income	*1.89*	*1.73*	*1.51*	*1.15*
>2 cars	1.00		1.00	1.00
1 car	1.30 (1.20–1.40)		1.20 (1.10–1.30)	1.20 (1.11–1.31)
Nil cars	1.60 (1.41–1.81)		1.38 (1.21–1.58)	1.37 (1.20–1.57)

Note: All odds ratios are adjusted for ethnicity and five-year age group.

† The 'seeking or available' include those people that were either available for work, or seeking work, but not both. Both these requirements must be meet to be labelled as unemployed (and hence part of the active labour force).

‡ RII calculations exclude 'other school qualifications' as it is of indeterminate rank.

9.2.1 Highest qualification as the exposure of interest

The univariate association of education with all-cause mortality tended to be stronger among 25–44 year olds than 45–64 year olds, with RIIs (excluding the 'other school qualification' category) of 2.08 compared to 1.81 for males, and 1.94 compared to 1.68 for females.

Controlling for the potential intermediaries/confounders of income and car access moderately reduced the education gradients, and additional control for labour force status had little impact other than among 45–64 year old females. This general pattern is consistent with that for the four causes of death. Of note, though, was that the association of education with all-cause mortality among 25–44 year olds was reduced less following multivariate control than among 45–64 year olds.

Thus, education appeared to be both a stronger and more persistent predictor of mortality among 25–44 year olds. Possible explanations for this difference by age include:

- an age effect, whereby education is a stronger determinant of early adulthood mortality due to proximity in the life-course
- a cohort effect, whereby education is a stronger determinant of life-chances (eg, income, status, and health) among more recent birth cohorts.

9.2.2 Equivalised household income as the exposure of interest

As already noted several times in this report, income was a strong univariate predictor of mortality – more so among 45–64 year olds than 25–44 year olds (Tables 46 and 47 above). Controlling for the potential confounder education reduced the all-cause mortality gradients by up to third, and controlling for both education and car access reduced the income gradients by a third to half for each of the four sex by age groups. Additional control for labour force status reduced to the null the association of adult household income with adult all-cause mortality, the only exception of note being some remaining protective effect of high income among 45–64 year old males. (Note that the residual association of income with mortality varied by cause of death as described in previous sections.) As discussed previously in this report, this reduction was probably function of all of: confounding by labour force status and the other socioeconomic factors, mediation by asset accumulation and health status (for which labour force status is a proxy) and health selection.

9.2.3 Car access as the exposure of interest

Car access was a stronger predictor of all-cause mortality among males, but there was no difference in the strength of the association between 25–44 and 45–64 year olds. About half of the association was attributable to potential confounding by the other socioeconomic factors.

9.2.4 Labour force status as the exposure of interest

Unemployment

There was an approximately 50% excess mortality risk among the unemployed compared to the full-time employed for each of the sex by age groups, except 45–64 year old females. Sensitivity analyses presented previously in this report have suggested that this association was neither due to selection bias linkage bias, nor health selection. Controlling for these socioeconomic factors, the odds ratio of all-cause mortality for the unemployed compared to the full-time employed remained elevated at 1.28 (95% CI 1.10 to 1.51) for males aged 45–64 years (Table 46). For males and females aged 25–44 years, the odds ratios remained elevated (1.26 and 1.38, respectively), but the 95% confidence intervals just included 1.0 (Tables 46 and 47).

Non-active labour force status

There were strong univariate and multivariate associations of the two non-active labour force categories (seeking or available, and retired etc) with all-cause mortality. The cause-specific analyses presented previously are more useful for interpretation than the all-cause analyses.

Part-time employment

The part-time employed were not differentiated from the part-time employed in cause-specific mortality analyses. Of interest, the elevated univariate odds ratios of mortality among the part-time versus full-time employed (except 25–44 year old females) remained largely unchanged following multivariate control (Tables 46 and 47). This was surprising, as it was expected that the univariate elevations would be due to confounding by socioeconomic position. A possible explanation was drift health selection whereby people in poor health cut down their hours rather than exit the labour force. (While an underpowered test, there was, however, no direct evidence of drift health selection among the part-time employed in Figure 36 in Appendix C.) Alternatively, it may be that part-time jobs are intrinsically more health damaging than full-time jobs, independent of confounding by socioeconomic or other factors.

Chapter 5: Discussion

1 Causal inference

A commonly used set of criteria for assessing causality are the Bradford Hill criteria (Hill 1965): strength, consistency, specificity, temporality, dose-response, biological plausibility, coherence, experimental evidence and analogy. Except temporality, none of the criteria are necessary (Rothman and Greenland 1998), and neither single criteria nor all criteria met simultaneously are sufficient for causal inference. However, they do provide some guidance. As to the association of socioeconomic position with mortality:

- the association is strong for a range of socioeconomic factors and most causes of death
- socioeconomic mortality gradients are consistently found in different settings and time periods
- socioeconomic position is associated with nearly all diseases – thus there is a lack of specificity. However, many exposures (eg, tobacco) have been found to be associated to many diseases
- many studies have shown that socioeconomic position in childhood or years prior to death are associated with mortality
- the association of socioeconomic position with mortality usually has a notable dose-response – the so-called gradient
- models of how socioeconomic position affects health are plausible, with a range of behaviours, psychosocial and materialistic factors likely to mediate the association of socioeconomic position with health
- there is an emerging literature on the neuroendocrine and physiological responses to experimental manipulation of social rank among primates, which is probably applicable to humans.

Thus, socioeconomic position in the general sense is causally associated with health. The more challenging issues for causal inference include:

- identifying which periods of exposure to socioeconomic position are critical (eg, is the critical exposure period intergenerational, early childhood, youth or adult socioeconomic position?)
- given the correlation of socioeconomic factors at one point in time, what are the socioeconomic factors that if changed would result in a change in health status or mortality risk (eg, is it movement out of unemployment into a satisfying job that inherently improves health, or the associated change in income?)
- determining the appropriate level of analysis and intervention (eg, this report has focused on the individual-level and leads to interpretations such as 'if we prevented this man from becoming unemployed, then his suicide risk would not be doubled'). However, the observed association of unemployment with suicide may vary by population-level factors such as the inclusiveness of society and the background unemployment rate.

These issues essentially reduce to unravelling causal relationships and confounding. The results from the NZCMS can make some progress in this regard, but are limited due to having measurement of socioeconomic position at only one point in time (although a variable like education arguably captures socioeconomic position from earlier life) and no measures of intermediary variables (other than smoking for the 1981 and 1996 census). The object of the following sections is to consider the causal implications for the association of each socioeconomic factor with mortality, drawing on:

- a general understanding of how sources of error may bias the epidemiological investigation of socioeconomic mortality gradients, presented in Appendix B of this report
- the international and national body of knowledge on socioeconomic mortality gradients and, more generally, socioeconomic determinants of health, reviewed in Appendix A of this report
- the cohort analyses presented in Chapter 4
- the sensitivity analyses presented in Appendix C.

The above issues are considered under subheadings of the four socioeconomic factors that were included in the multivariate analyses: education, labour force status, income and car access. The interpretation of mortality gradients by other socioeconomic factors included in this report – particularly small area deprivation – will be addressed more fully in subsequent work. Further discussions have not been presented specifically by cause of death: this was the framework used in the last chapter, although some issues will need further investigation in subsequent work. For example, the association of prostate and breast cancer with education and deprivation reported in this report was not in the expected direction of lower mortality among lower socioeconomic groups as reported in other studies (Kogevinas et al 1997). This may be due to more recent age-cohorts no longer having socioeconomic variation in the incidence of disease, and may also result from differential survival by socioeconomic position post-diagnosis (Harding et al 1999; Martikainen 1990).

1.1 Education

Education was a strong and independent predictor of all-cause mortality and cause-specific mortality among 25–64 year olds in this report, and the association was stronger among 25–44 year olds than 45–64 year olds. Other studies have also found a stronger association of education with mortality among younger adults (Kitagawa and Hauser 1973; Sorlie et al 1995).

The 'middle' categories of highest qualification (trade or technical qualifications, any school qualification) varied little in mortality risk in this report. Compared to these middle categories though, people with a higher tertiary qualification had a notably lower mortality risk, and people with nil qualification had a notable higher mortality risk. Such a 'trichotomy' or 'three-step' association of education with mortality (rather than a smooth monotonic gradient) has also been described previously in the US NLMS (Backlund et al 1999).

The results for highest qualification in this report demonstrated a similar gradient to that in the US NLMS for years of education (Sorlie et al 1995), and a stronger gradient that that for highest education in the US 1960 Matched Records Study (Kitagawa and Hauser 1973). Unlike the NLMS, controlling for possible intermediary factors of income and labour force status did not remove the NZCMS education gradient. Other than the difference in time (1980s) and country, one possible reason was that income was measured with much greater accuracy in the US NLMS compared to a single tick-box question in the New Zealand census. An analysis by Schrijvers et al in the Netherlands suggested that about a half of the mortality gradient by education was explained by material factors, and of this materialistic pathway half was a 'direct' effect on mortality and half mediated by identifiable health related behaviours (Schrijvers et al 1999) (see Figure 19 in Appendix A) and associated discussion for more details). Thus, somewhat less of the association of highest qualification with mortality in the NZCMS was explained by material factors than would have been expected on the basis of current international studies.

Further work is required in the NZCMS to make a direct and accurate comparison with the cross-national comparison study of Mackenbach and colleagues (Mackenbach et al 1999), but it appears that the mortality gradients in New Zealand are as strong, if not stronger, than those in western European countries. The further work required in the NZCMS includes incorporating the results for 65–74 year olds, and dichotomising the educational variable to match as closely as possible that used by Mackenbach and colleagues.

Returning to the NZCMS results and potential biases, as highest qualification is usually fixed by adulthood the association of highest qualification and mortality observed in the NZCMS should not be biased by health selection. Selection bias for analyses conducted on the restricted cohort were found to be minor. Linkage bias, while not measured directly by education, probably caused a modest underestimate of the association of highest qualification with mortality.

The association of highest qualification with mortality almost certainly represents a truly causal association via mechanisms such as knowledge, behaviour, material factors and so on, but much of the association may still be due to confounding. According to one life-course explanation, education may simply be a marker of socioeconomic position in childhood and youth, and it is the constellation of factors making up childhood socioeconomic position that influence most strongly later adult mortality risk. If it does act principally as a marker of childhood socioeconomic position, change in educational qualification may not have any effect on health status. Rather, those things for which highest qualification is a proxy (eg, parental income) may be the causal factors. Put another way, highest qualification may simply be a marker of a credentialising process whereby young people are ranked according to their aptitude, willingness to invest in the future, family environment and expectations, and intelligence. Rather than highest qualification, it may be these latter factors that affect future health. Testing this possibility will require longitudinal studies beginning at birth with thorough measurement of family circumstances, repeated measures of personal characteristics, and long follow-up for the occurrence of health events. In New Zealand, two such longitudinal studies are in process: the Dunedin Multidisciplinary Health, and Development Study and the Christchurch Health and Development Study. The participants in both these studies are currently aged in their 20s.

Assuming that the causal component of the education-mortality association is sizeable, improving an individual's educational status would take years, probably decades, to effect a change in their expected health status due to the time lag incurred on the various possible causal pathways. At a community or country-level, it seems likely that increasing the average educational status of the population would have health benefits over and above the individual-level effects. For example, a more highly educated population would probably perform better economically on an international scale, attracting more health enhancing resources to the population as a whole. Thus, higher educational attainment is likely to have health (and other social and economic) benefits via both individual-level and population-level mechanisms.

1.2 Labour force status

The patterns of adult mortality among the unemployed and non-active labour force compared to the employed in the 1991 NZCMS cohort were consistent with findings in the OPCS LS (Bethune 1996; Lewis and Sloggett 1998; Moser et al 1984; Moser et al 1986; Moser et al 1987) and linked census–mortality studies in other countries (Iversen et al 1987; Martikainen 1990; Sorlie et al 1995; Sorlie and Rogot 1990). In particular, the up to 50% excess mortality risk among the unemployed compared to employed in this report, and the twofold to threefold increased risk among the non-active labour force, were consistent with international studies. Likewise, controlling for other socioeconomic factors reduced the all-cause mortality gradient by about a third, and this was consistent with international studies.

Assuming that much of the association of non-active labour force status with mortality was due to health selection, perhaps the most interesting labour force results in this report were for the association of unemployment with cause-specific mortality. The univariate risk ratio of suicide death among unemployed compared to employed males was 2.70 (1.84 to 3.95) for males and 2.86 (1.19 to 6.85) for females. Furthermore, controlling for income, education and car access only reduced the risk ratios to 2.35 and 2.58, respectively. This minor reduction in the suicide risk ratio following multivariate control was in contrast to the larger reductions for cardiovascular disease and unintentional injury. International studies have also found relatively little confounding/mediation of the association by other socioeconomic factors (Iversen 1974; Lewis and Sloggett 1998; Martikainen 1990; Moser et al 1984). It seems likely, therefore, that the suicide association is not mediated to a large extent by material consequences of unemployment (eg, poverty) nor is it due to residual confounding.

Neither does it seem likely that the association of unemployment with suicide is due to health selection. First, the all-cause mortality risk plots over time for the unemployed did not show decreasing mortality as would be expected by health selection (Appendix C). Second, international studies with longer follow-up than the NZCMS have found that the association of suicide is still strong after five, or even ten years (Iversen 1974; Lewis and Sloggett 1998). (Note, however, that if psychosocial mechanisms mediate the association of unemployment with suicide, we might expect a short time lag and hence stronger effects soon after the onset of unemployment.) Third, while it is possible to argue that mental ill-health might be a cause of unemployment, research using repeated measures has demonstrated the reverse – unemployment among young men pre-dates symptoms of depression and anxiety (Montgomery et al 1999).

Regarding cardiovascular disease and unintentional injury, there was a modest association with unemployment in the 1991 NZCMS cohort, and that association was mostly explained by the potential mediators/confounders of education, income and car access. Whether the association for these other causes of disease was entirely due to confounding is impossible to state. One also needs to consider that there was probably linkage bias such that the association of unemployment with mortality was underestimated in the NZCMS.

The relationship of unemployment with mortality is plausible – particularly for suicide. Unemployment is stressful and socially isolating in addition to having financial consequences (Bartley 1994). Stress and loss of social support have both been shown to be related to mortality (Berkman and Glass 2000). Particularly for non-suicide deaths, the best way to conceptualise the association of unemployment with mortality might be in a life-course framework as advanced by Montgomery and colleagues (Montgomery et al 1996; Montgomery et al 1998; Montgomery et al 1999). They have shown that unemployment itself is predicted by childhood health and social indicators that will also be associated with adult mortality (ie, technically confounders), but also that unemployment may help establish life-long patterns of hazardous behaviour and symptoms of depression and anxiety.

Finally, it would be imprudent to dismiss the increased mortality risk among the non-active labour force as spurious or inconsequential. First, about half the deaths among 25–64 year olds in the 1991 NZCMS cohort were in the non-active labour force. Second, while much of their excess mortality risk may be due health selection, probably not all of it is. It seems likely that some of the putative causal mechanisms associating unemployment and mortality might apply for the non-active labour force. For example, while not meeting the definition for unemployment, there will be many people among the non-active labour force who are suffering from chronic unemployment in the more general sense, but may have ceased actively looking for work.

1.3 Household income

Interpreting the association of household income with mortality is complex and difficult. In this report there were strong univariate associations of household income with all-cause mortality (particularly among males and 45–64 year olds) and all specific causes of death other than non-lung cancers.

Exhaustive testing was undertaken for possible health selection effects, the result being that there were some suggestions of both drift and differential health selection – but not much. For example, the mortality risk plots over time (not excluding the non-active labour force) were mostly parallel by income-level, including for cancer deaths, failing to suggesting that drift health selection was not evident within three-years of follow-up (Appendix C). Excluding sickness beneficiaries and pre-hospitalised deaths did not dramatically reduce the cancer and cardiovascular disease mortality gradients by income – certainly no more than for the baseline analyses by NZDep91 that were assumed to be immune from drift health selection effects. However, controlling for labour force status dramatically reduced the income–mortality gradients for cancer (and somewhat for cardiovascular disease and suicides) but not for unintentional injury deaths – exactly the pattern we would expect if health selection was inflating the univariate associations of income with cancer mortality.

Several reasons may explain why the income–mortality gradients in this report dramatically reduced following the control for labour force status. These included confounding by labour force status and health selection. Short-term health selection was not directly detected during the three years of follow-up, but both a small amount of undetected short-term health selection and some longer-term health selection may have been present and affected the income–mortality gradients. For example, long-standing illness among the non-active labour force on census night may have caused death at a fairly constant rate over three years.

However, controlling for labour force status may have overcontrolled the income–mortality association for two reasons. First, labour force status is a proxy for health status, and health status is on the causal pathway between socioeconomic factors and chronic diseases. However, if this overcontrol due to including an intermediary variable in the analyses was a substantive issue, then the education gradients should also have reduced substantially following additional control for labour force status – yet they did not. Second, health selection out of the active labour force that was differential by household income-level would have caused the income–mortality association to be underestimated. But again, this phenomenon should also have applied to the education gradients.

What have other researchers found regarding health selection and income–mortality gradients? In short, little is known, and this is due to the lack of international studies with income data. Of interest though, income–mortality gradients were strong in the US NLMS, and reduced following control for labour force status, but not quite as much as in the NZCMS (Backlund et al 1996; Backlund et al 1999; Sorlie et al 1995). Also, all-cause mortality gradients reduced in a similar manner following control for labour force status in a linked census–mortality study using 1990 Finnish census and taxation data and five years of subsequent follow-up (personal communication, Pekka Martikainen, July 2000). (The recent addition of income data to the Finnish-linked census mortality data will provide important comparative data for the NZCMS.)

Considering morbidity, controlling for labour force status has been found to dramatically reduce the association of income with chronic conditions and perceived general health (Stronks et al 1997). Stronks et al interpreted this reduction as being due to health selection. However, longitudinal income dynamics studies find that while there is health selection, there remains an underlying association of income and health status – not just reverse causation (Benzeval et al 2000).

Having concluded that health selection probably causes an overestimate of the association of cancer (and to a lesser degrees cardiovascular disease and suicide) with income in the NZCMS, was there still a residual association of income with mortality? For cardiovascular disease and unintentional injury deaths, almost certainly yes – as there were still strong associations of income with mortality in the multivariate models. For cancer and suicide deaths, possibly not.

International studies have shown that measures of income averaged over time are better predictors of mortality than a one-off cross-sectional measures such as a census (Lynch et al 1997b; McDonough et al 1997). However, there is also some measure of stability over time in household income. The New Zealand 1991 census household income variable is also not measured particularly accurately, being based on answers given by individuals within each household to a single tick-box question on gross income in the last year. Therefore, along with opposing health selection effects, the associations of income and mortality in this report are probably underestimated due to measurement error of income – particularly if the association of 'usual' or 'average' income with mortality is being inferred.

1.4 Car access

The association of car access with adult all-cause mortality, cardiovascular disease and suicide deaths was strong in the 1991 NZCMS cohort. These findings agree well with those previously documented in the UK (Davey Smith et al 1990; Filakti and Fox 1995). Assuming that car access was a (rather poor) proxy for asset wealth, this finding suggests that the association of mortality with wealth would be stronger again, were assets more accurately measured.

2 The NZCMS – a precedent for New Zealand

The NZCMS is a precedent in New Zealand. Not only have census and mortality records never been linked before, but no other data set has been linked with census data. Of the submitted mortality records for 0–74 year olds, 76.6% were anonymously and probabilistically linked to a 1991 census record – the study base for this report. This linkage success is comparable with other linked census–mortality studies that, unlike the Scandinavian countries, have not had the luxury of unique personal identifiers. The classic 1960 Matched Records Study in the USA by Kitagawa and Hauser (Kitagawa and Hauser 1973) achieved a 77% linkage rate using matching variables not dissimilar to the NZCMS. Moreover, the linkage rate by age was similar to that in the NZCMS (Table 18). The 1960 Matched Records Study only linked deaths in the four months after census night, compared to three years for the NZCMS. In Italy, 75–78% of deaths in the six months following the 1981 census in Italy were linked back to that census (Faggiano et al 1995; Kunst et al 1996).

The 1991 census–mortality study was prone to linkage bias consequent on the less than complete ascertainment of the mortality outcome. Within demographic strata, there was residual linkage bias by socioeconomic position. By small area deprivation, decedents from the most deprived decile were 13%, 11%, 9%, and 4% less likely to be linked to a mortality record than decedents from the least deprived decile for 25–44 and 45–64 year old males, and 25–44 and 45–64 year old females, respectively (Table 20). However, for deciles 1 to 9 (particularly 2 to 8), there was very little difference in the probability of being linked to a census record. Likewise, by NZSEI occupational class for males there was little substantive difference in the probability of being linked to a census record for male decedents from occupational classes 1 to 5 (Table 23). Male decedents from occupational class 6, however, were 25% and 6% less likely to be linked than decedents from class 1, for 25–44 and 45–64 year olds respectively. (There was little difference between classes 2 to 5 for 25–44 year old males, rather the comparison between classes 1 and 6 appeared to be extreme results.)

This systematic difference in linkage success led to a misclassification bias of the mortality outcome (ie, more deaths missed in the lowest socioeconomic groups compared to the highest). This linkage bias caused an underestimate of the risk ratios comparing the lowest and highest socioeconomic groups, most notably for males and 25–44 year olds. However, for the majority of socioeconomic groups there was little difference in the proportion of mortality records linked to a census record, and hence little expected bias of the risk ratios.

Comparisons of the full and restricted data sets identified little selection bias for analyses based on the restricted cohort (approximately 80% of the census respondents) for whom there was complete data for education, income, labour force status and car access. Linkage and selection biases tended to cancel each other out for all-cause mortality among 45–64 year old males, and unintentional injury mortality among 25–64 year old males. Otherwise, the combined impact of selection and linkage bias on the univariate results for the 1991 census cohort reported in this report was to cause socioeconomic mortality gradients to be underestimated by up to 25% – although this net bias varied by demographic group and cause of death.

3 Future directions and policy implications of the NZCMS

This report, and the NZCMS more generally, may be viewed as 'infrastructure research'. That is, the NZCMS has demonstrated that anonymously and probabilistically linking census and mortality records in New Zealand provides robust data that may be used to increase the understanding of health inequalities in New Zealand. The NZCMS has obvious limitations: census data do not capture socioeconomic factors as completely as we would like; intermediary variables (eg, blood pressure, diet) are not recorded on census data; there is some linkage bias – although it is modest and quantified; and the follow-up is short, making some analyses (eg, income) prone to health selection. However, there are many strengths including: cohort studies of the entire population are created at relatively little marginal cost (n = 3 to 4 million); socioeconomic mortality gradients can be measured for both sexes, for ages 0–74 years, and for a range of socioeconomic factors other than occupational class and small area deprivation for which we already have data in New Zealand; by linking other censuses (ie, 1981, 1986, and 1996) to mortality data, we can compare socioeconomic mortality gradients over time in New Zealand; and the NZCMS results are comparable to results from similar studies overseas, so allowing cross-national comparisons.

The future directions for the 1991 census cohort of the NZCMS are:

1 To measure socioeconomic mortality gradients among children and 65–74 year olds in the cohort.

2 To conduct multilevel analyses to measure the association of contextual variables (eg, income inequality and small area deprivation) with mortality having controlled for individual-level socioeconomic factors.

3 To determine the numerator–denominator bias between ethnicity data collected on mortality data and ethnicity data collected on census data. There are large differences in New Zealand between mortality and census data as to who is assigned as Maori, Pacific, or non-Maori non-Pacific (Pomare et al 1995), causing routinely published (ie, unlinked analyses) ethnic-specific mortality rates to be grossly underestimated. As the NZCMS links census and mortality records, it is possible to directly compare ethnicity between census and mortality data for what is highly likely to be the same person.

4 To measure socioeconomic mortality gradients within ethnic groups, *and* to determine the overlap between socioeconomic and ethnic mortality inequalities.

5 To compare socioeconomic mortality gradients in New Zealand during 1991–94 with those in other countries.

Perhaps the most exciting future direction in the NZCMS is the anonymous and probabilistic record linkage of the 1981, 1986 and 1996 censuses each to three subsequent years of mortality data. Comparisons of socioeconomic mortality gradients over a 20-year period of radical socioeconomic change in New Zealand will then become possible – essentially a unique natural experiment. Further into the future, the linkage of the 2001 and subsequent censuses with mortality could become a routine function of Statistics New Zealand and the Ministry of Health.

The policy implications of this report, and the NZCMS more directly, might be considered twofold. First, despite its limitations, the NZCMS does throw light on causal mechanisms of relevance to policy makers. For example, results in this report strongly suggest that:

- unemployment is causally associated with suicide
- education is a strong underlying determinant of premature mortality – increased education attainment of today's children and youth should yield lower adult mortality in the future
- redistribution of material factors (income and wealth) in our society should lower mortality rates among poorer people and (because of the non-linear association of income with mortality) the population as a whole.

The results in this report also provide further evidence of health inequalities in New Zealand. This evidence should be considered during resource allocation, for example, when constructing population-based funding formulas for distribution of health-care resources.

The second policy implication of the NZCMS is more diffuse. While we may not fully understand the causal mechanisms driving socioeconomic inequalities in health at the individual-level, the NZCMS does provide a tool to monitor socioeconomic mortality gradients over time and between countries. As such, we will be able to determine whether mortality gradients increase or decrease coincident with macro-social and economic changes in New Zealand. Determining which features of the macro environment are causally related to health inequalities in the population is a complex task, but NZCMS results compared both over time and with other countries have the potential to identify changes in the social structure that are associated with smaller or greater health inequalities.

References

Adler N, Ostrove J. 1999. Socioeconomic status and health: what we know and what we don't. In N Adler, M Marmot, B McEwen, et al (eds) *Socioeconomic Status and Health in Industrial Nations: Social, Psychological, and Biological Pathways* (pp.3–15). New York: The New York Academy of Sciences.

Anderson R, Sorlie P, Backlund E, et al. 1997. Mortality effects of community socioeconomic status. *Epidemiol* 8: 42–7.

Arber S. 1987. Social class, non-employment, and chronic illness: Continuing the heath inequalities debate. *BMJ* 294: 1069–73.

Armitage P, Colton T. 1998. *Encyclopedia of Biostatistics*. New York: John Wiley & Sons.

Atkinson A., Rainwater L, Smeeding T. 1995. *Income Distribution in OECD Countries: Evidence for the Luxembourg Income Study*. Paris: OECD.

Backlund E, Sorlie PD, Johnson NJ. 1996. The shape of the relationship between income and mortality in the United States: Evidence from the National Longitudinal Mortality Study. *Ann Epidemiol* 6: 12–20.

Backlund E, Sorlie PD, Johnson NJ. 1999. A comparison of the relationship of education and income with mortality: the National Longitudinal Mortality Study. *Soc Sci Med* 49: 1373–84.

Baldwin J, Acheson E, Graham W. 1987. *Textbook of Medical Record Linkage*. Oxford: Oxford University Press.

Barker D, Martyn C. 1992. The maternal and fetal origins of cardiovascular disease. *J Epidemiol Community Health* 46: 8–11.

Bartley M. 1994. Unemployment and ill health: understanding the relationship. *J Epidemiol Community Health* 48: 333–7.

Bartley M. 1996. Unemployment and health selection [commentary]. *Lancet* 348: 904.

Bartley M, Ferrie J, Montgomery S. 1999. Living in a high-unemployment economy: understanding the health consequences. In M Marmot, R Wilkinson (eds) *Social Determinants of Health* (pp.81–104). Oxford: Oxford University Press.

Bartley M, Owen C. 1996. Relation between socioeconomic status, employment, and heath during economic change, 1973–93. *BMJ* 313: 445–9.

Benzeval M, Dilnot A, Judge K, et al. 2000. Income and health: the time dimension. *Soc Sci Med*, in press.

Benzeval M, Judge K, Whitehead M. 1995. *Tackling inequalities in health: An agenda for action*. London: Kings Fund.

Beral V, Chilvers C, Fraser P. 1979. On the estimation of relative risk from vital statistical data. *J Epidemiol Community Health* 33: 159–62.

Berkman L, Glass T. 2000. Social integration, social networks, social support, and health. In L Berkman, I Kawachi (eds) *Social Epidemiology* (pp.137–73). New York: Oxford University Press.

Berkman L, Macintyre S. 1997. The measurement of social class in health studies: old measures and new formulations. In M Kogevinas, N Pearce, M Susser, et al (eds) *Social Inequalities and Cancer* (pp.51–64). Lyon: IARC Scientific Publications.

Bethune A. 1996. Economic activity and mortality of the 1981 census cohort in the OPCS longitudinal study. *Population Trends* 83: 37–42.

Blakely T. 1996. *Epidemiology and Socioeconomic Mortality Gradients: The role of ecological and multilevel methods in research*. Unpublished Masters of Public Health [dissertation], University of Otago.

Blakely T. 2001. *Socioeconomic Factors and Mortality Among 25–64 year olds: The New Zealand Census–Mortality Study*. (Also at http://www.wnmeds.ac.nz/nzcms-info.htm.) Unpublished Doctorate, University of Otago.

Blakely T, Kawachi I. 2001. What is the difference between controlling for mean versus median income in analyses of income inequality? *J Epidemiol Community Health* 55: 352–3.

Blakely T, Kennedy B, Glass R, et al. 2000a. What is the lag time between income inequality and heath status? *J Epidemiol Community Health* 54: 318–19.

Blakely T, Kennedy B, Kawachi I. 2001. Socioeconomic inequality in voting participation and self-rated health. *Am J Public Health* 91: 99–104.

Blakely T, Lochner K, Kawachi I. 2002. Metropolitan area income inequality and self-rated health: a multilevel study. *Soc Sci Med* 54: 65–77.

Blakely T, Robson B, Atkinson J, et al. 2002. Unlocking the numerator–denominator bias. I: Adjustment ratios by ethnicity for 1991–94 mortality data. *NZ Med J* 115: 39–43.

Blakely T, Salmond C, Woodward A. 1999. *Anonymous record linkage of 1991 census records and 1991–94 mortality records: The New Zealand Census–Mortality Study*. (Also at http://www.wnmeds.ac.nz/nzcms-info.htm.) Wellington: Department of Public Health, Wellington School of Medicine, University of Otago.

Blakely T, Salmond C, Woodward A. 2000b. Anonymous linkage of New Zealand mortality and Census data. *Aust NZ J Public Health* 24: 92–5.

Blakely T, Woodward A. 2000a. Counterfactual challenges to social epidemiology. *Australasian Epidemiologist* 7: 28–31.

Blakely T, Woodward A. 2000b. Ecological effects in multilevel studies. *J Epidemiol Community Health* 54: 367–74.

Blane D. 1999. The life-course, the social gradient, and health. In M Marmot, R Wilkinson (eds) *Social Determinants of Health* (pp.64–80). Oxford: Oxford University Press.

Blane D, Bartley M, Davey Smith G. 1997. Disease aetiology and materialistic explanations of socioeconomic mortality differentials. *Eur J Public Health* 7: 385–91.

Blane D, Davey Smith G, Bartley M. 1993. Social selection: what does it contribute to social class differences in health? *Soc Health Illness* 15: 1–15.

Blane D, Hart C, Davey Smith G, et al. 1996. Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. *BMJ* 313: 1434–8.

Bobak M, Hertzman C, Skodova Z, et al. 1998. Association between psychological factors at work and nonfatal myocardial infarction in a population-based case-control study in Czech men. *Epidemiol* 9: 43–7.

Bobak M, Hertzman C, Skodova Z, et al. 1999. Socioeconomic status and cardiovascular risk factors in the Czech Republic. *Int J Epidemiol* 28: 46–52.

Bobak M, Hertzman C, Skodova Z, et al. 2000. Own education, current conditions, parental material circumstances, and risk of myocardial infarction in a former communist country. *J Epidemiol Community Health* 54: 91–6.

Borrell C, Regidor E, Arias L-C, et al. 1999. Inequalities in mortality according to education level in two large Southern European cities. *Int J Epidemiol* 28: 58–63.

Bosma H, Marmot M, Hemingway H, et al. 1997. Low job control and risk of coronary health disease in Whitehall II (prospective cohort) study. *BMJ* 314: 558–65.

Bosma H, Peter R, Siegrist J, et al. 1998. Two alternative job stress models and the risk of coronary heart disease. *Am J Public Health* 88: 68–74.

Bosma H, Schrijvers C, Mackenbach J. 1999. Socioeconomic inequalities in mortality and importance of perceived control: cohort study. *BMJ* 319: 1469–70.

Brenner H, Gefeller O. 1993. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *Am J Epidemiol* 138: 1007–15.

Brunner E. 1997. Stress and the biology of inequality. *BMJ* 314: 1472–6.

Brunner E, Marmot M. 1999. Social organization, stress, and health. In M Marmot, R Wilkinson (eds) *Social Determinants of Health* (pp.17–43). Oxford: Oxford University Press.

Bryk A, Radenbush S. 1992. *Hierarchical Linear Models: Applications and data analysis methods*. Newbury Park: Sage.

Bucher H, Raglan D. 1995. Socioeconomic indicators and mortality from coronary heart disease and cancer: a 22-year follow-up of middle age men. *Am J Public Health* 85: 1231–6.

Burton P, Gurrin L, Sly P. 1998. Extending the simple linear regression model for correlated responses: an introduction to generalized estimating equations and multilevel mixed modelling. *Statist Med* 17: 1261–91.

Calle E, Terrell D. 1993. Utility of the National Death Index for ascertainment of mortality among Cancer Prevention Study II Participants. *Am J Epidemiol* 137: 235–41.

Cardano M., Costa G, Demaria M, et al. 1999. Inequalities in mortality in the Italian longitudinal studies. [Italian]. *Epidemiologia e Prevenzione* 23: 141–52.

Cavelaars A. 1998. *Cross-national Comparisons of Socioeconomic Differences in Health Indicators [thesis]*. Rotterdam: Erasmus University.

Cavelaars A, Kunst A, Geurts J, et al. 2000. Educational differences in smoking: international comparisons. *BMJ* 320: 1102–7.

Cavelaars A, Kunst A, Mackenbach J. 1997. Socioeconomic differences in risk factors for morbidity and mortality in the European Community. *J Health Psychol* 2: 353–72.

Cheng Y, Kawachi I, Coakley E, et al. 2000. Association between psychosocial work characteristics and health functioning in American women: prospective study. *BMJ* 320: 1432–6.

Clemance P. 1985. An application of correspondence analysis as a multi-dimensional scaling technique. *The New Zealand Statistician* 20: 26–34.

Copeland K, Checkoway H, McMichael A, et al. 1977. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 105: 488–95.

Costa G, Segnen N. 1987. Unemployment and mortality. *BMJ* 294: 1550–1.

Costantini A, Pirastu R, Lagorio S, et al. 1994. Studying cancer among female workers: methods and preliminary results from a record linkage system in Italy. *J Occup Med* 36: 1180–6.

Crampton P, Salmond C, Sutton F. 1997. NZDep91: A new index of deprivation. *Social Policy Journal of New Zealand* 9: 186–93.

Dahl E. 1993. High mortality in lower salaried Norwegian men: the healthy worker effect. *J Epidemiol Community Health* 47: 192–4.

Dahlgren G, Whitehead M. 1991. *Policies and Strategies to Promote Social Equity in Health*. Stockholm: Institute for Future Studies.

Daly M, Duncan G, Kaplan G, et al. 1998. Macro-to-micro links in the relation between income inequality and mortality. *The Milbank Quarterly* 76: 315–39.

Daly M, Duncan G, McDonough P, et al. Optimal indicators of socioeconomic status for health research. *Am J Public Health*, in press.

Davey Smith G, Blane D, Bartley M. 1994. Explanations for socioeconomic differentials in mortality: evidence from Britain and elsewhere. *Eur J Public Health* 4: 131–44.

Davey Smith G, Harding S. 1997. Is control at work the key to socioeconomic gradients in mortality? *Lancet* 350: 1369–70.

Davey Smith G, Hart C, Blane C, et al. 1997. Lifetime socioeconomic position and mortality: prospective observational epidemiology. *BMJ* 314: 547–52.

Davey Smith G, Hart C, Blane D, et al. 1998a. Adverse socioeconomic conditions in childhood and cause specific adult mortality: prospective observational study. *BMJ* 316: 1631–5.

Davey Smith G, Hart C, Hole D, et al. 1998b. Education and occupational social class: which is the more important indicator of mortality risk? *J Epidemiol Community Health* 52: 153–60.

Davey Smith G, Neaton J, Wentworth D, et al. 1996a. Socioeconomic differentials in mortality risk among men screened for the multiple risk factor intervention trial: I. White Men. *Am J Public Health* 86: 486–96.

Davey Smith G, Neaton J, Wentworth D, et al. 1998c. Mortality differences between black and white men in the USA: contribution of income and other risk factors among men screen for the MRFIT. *Lancet* 351: 934–9.

Davey Smith G, Phillips A. 1992. Confounding in epidemiological studies: why 'independent' effects may not be all they seem. *BMJ* 305: 757–9.

Davey Smith G, Shipley M, Rose G. 1990. Magnitude and causes of socioeconomic differentials in mortality: further evidence from the Whitehall Study. *J Epidemiol Community Health* 44: 265–70.

Davey Smith G, Wentworth D, Neaton J, et al. 1996b. Socioeconomic differentials in mortality risk among men screened for the multiple risk factor intervention trial: II. Black Men. *Am J Public Health* 86: 497–504.

Davis P, Graham P, Pearce N. 1999a. Health expectancy in New Zealand, 1981–1991: social variation and trends in a period of rapid social and economic change. *J Epidemiol Community Health* 53: 519–27.

Davis P, McLeod K, Ransom M, et al. 1997. *The New Zealand Socioeconomic Index of Occupational Status (NZSEI).* Wellington: Statistics New Zealand.

Davis P, McLeod K, Ransom M, et al. 1999b. The New Zealand Socioeconomic Index: Developing and validating an occupationally-derived indicator of socioeconomic status. *Aust NZ J Public Health* 23: 27–33.

Dayal H, Goldberg-Alberts R, Ramos J, et al. 1986. Patterns of mortality from selected causes in an urban population. *J Chron Dis* 39: 877–88.

Department of Statistics. 1992a. *1991 Census of Population and Dwellings: Concepts, definitions and classifications.* Wellington, New Zealand: Department of Statistics.

Department of Statistics. 1992b. *1991 Census of Population and Dwellings: Variable dictionary.* Christchurch, New Zealand: Department of Statistics.

Desplanques G. 1984. Social inequality of mortality (1975–1980): a longitudinal study. [French]. *Social- und Praventivmedizin* 29: 268–72.

Diderichsen F, Hallqvist J. 1997. Trends in Occupational Mortality among Middle-aged Men in Sweden 1961–1990. *Int J Epidemiol* 26: 782–7.

Diez-Roux A. 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 88: 216–22.

Diez-Roux A, Nieto F, Muntaner C, et al. 1997. Neighbourhood environment and coronary heart disease: a multilevel analysis. *Am J Epidemiol* 146: 48–63.

Doornbos G, Kromhout D. 1990. Educational level and mortality in a 32-year follow-up study of 18-year-old men in the Netherlands. *Int J Epidemiol* 19: 374–479.

Duncan C, Jones K, Moon G. 1993. Do places matter? A multilevel analysis of regional variations in health-related behaviour in Britain. *Soc Sci Med* 37: 725–33.

Duncan C, Jones K, Moon G. 1996. Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med* 42: 817–30.

Duncan C, Jones K, Moon G. 1998. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 46: 97–117.

Duncan C, Jones K, Moon G. 1999. Smoking and deprivation: are there neighbourhood effects? *Soc Sci Med* 48: 497–505.

Duncan G. 1996. Income dynamics and health. *Int J Health Services* 26: 419–44.

Duncan G, Brooks-Gun J, Klebanov P. 1994. Economic deprivation and early childhood development. *Child Dev* 65: 296–318.

Eames M, Ben-Shlomo Y, Marmot M. 1993. Social deprivation and premature mortality; regional comparison across England. *BMJ* 307: 1097–1102.

Edgell S. 1993. *Class*. London: Routledge.

Elley W, Irving J. 1976. Revised socioeconomic index for New Zealand. *New Zealand Journal of Educational Studies* 11: 25–30.

Elo I, Preston S. 1996. Educational differential in mortality: United States, 1979–85. *Soc Sci Med* 42: 47–57.

Emmons K. 2000. Health behaviours in a social context. In L Berkman, I Kawachi (eds) *Social Epidemiology* (pp.267–305). New York: Oxford University Press.

Erikson E, Goldthorpe J. 1992. The Constant Flux. [Cited in Kunst et al. 1998. *Soc Sci Med*.] Oxford: Clarendon Press.

Everson S, Lynch J, Chesney M, et al. 1997. Interaction of workplace demands and cardiovascular reactivity in progression of carotid atherosclerosis: population based study. *BMJ* 314: 553–8.

Ewing I. 1997. *Hot Off the Press: Post Enumeration Survey (PES) 1996*. Wellington: Statistics New Zealand.

Faggiano F, Lemma P, Costa G, et al. 1995. Cancer mortality by educational level in Italy. *Cancer Causes and Control* 6: 311–20.

Feldman J, Makuc D, Kleinman J, et al. 1989. National trends in educational differentials in mortality. *Am J Epidemiol* 129: 919–33.

Ferrie J, Shipley M, Marmot M, et al. 1995. Health effects of anticipation of job change and non-employment: longitudinal data from Whitehall II study. *BMJ* 311: 1264–9.

Filakti H, Fox A. 1995. Differences in mortality by housing tenure and by car access from the OPCS Longitudinal Study. *Population Trends* 81: 27–30.

Fox A, Adelstein A. 1978. Occupational mortality: work or way of life? *J Epidemiol Community Health* 32: 73–8.

Fox A, Goldblatt P. 1982. *Longitudinal Study 1971–1975: Sociodemographic mortality differences*. London: Her Majesty's Stationery Office: Office of Population Censuses and Surveys.

Fox A, Goldblatt P, Adelstein A. 1987. Selection and mortality differentials. In J Baldwin, E Acheson, W Graham (eds) *Textbook of Medical Record Linkage* (pp.79–99). Oxford: Oxford University Press.

Fox A, Goldblatt P, Jones D. 1985. Social class mortality differentials: artefact, selection or life circumstances? *J Epidemiol Community Health* 39: 1–8.

Fuchs V. 1979. Economics, health, and post-industrial society. *Milbank Memorial Fund Quarterly* 57: 153–82.

Ganzeboom H, De Graaf P, Treiman D, et al. 1992. A standard international socioeconomic index of occupational status. *Social Science Research* 21: 1–56.

Ganzeboom H, Luijkx R, Treiman D. 1989. Intergenerational class mobility in comparative perspective [cited in Kunst et al. 1998. *Soc Sci Med*]. *Res Soc Stratification Mobility* 8: 3–84.

Gliksman M, Kawachi I, Hunter D, et al. 1995. Childhood socioeconomic status and risk of cardiovascular disease in middle aged US women: a prospective study. *J Epidemiol Community Health* 49: 10–15.

Goldblatt P. 1989. Mortality by social class, 1971–85. *Population Trends* 56: 6–15.

Goldblatt P, Fox J. 1979. Household mortality from the OPCS Longitudinal Study. *Population Trends* 14: 21–7.

Goldstein H. 1995. *Multilevel Statistical Models*. London: Edward Arnold.

Graham H. 1994. *When Life's a Drag*. London: Her Majesty's Stationery Office.

Green M. 1983. Use of predictive value to adjust relative risk estimates biased by misclassification of outcome status. *Am J Epidemiol* 117: 98–105.

Greenland S. 1980. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 112: 564–9.

Greenland S. 1992. Divergent biases in ecologic and individual-level studies. *Stat Med* 11: 1209–23.

Greenland S, Morgenstern H. 1989. Ecological bias, confounding, and effect modification. *Int J Epidemiol* 18: 269–74.

Greenland S, Pearl J, Robins J. 1999. Causal diagrams for epidemiologic research. *Epidemiol* 10: 37–48.

Greenland S, Robins J. 1994. Invited commentary: ecologic studies – biases, misconceptions, and counterexamples. *Am J Epidemiol* 139: 747–60.

Haan M, Kaplan G, Camacho T. 1987. Poverty and health. *Am J Epidemiol* 125: 989–98.

Harding S. 1995. Social class differences in mortality of men: recent evidence from the OPCS longitudinal study. *Population Trends* 80: 31–7.

Harding S, Brown J, Rosato M, et al. 1999. Socioeconomic differentials in health: illustrations from the Office for National Statistics Longitudinal Study. *Health Statistics Quarterly* 1: 5–15.

Hemingway H, Marmot M. 1999. Psychosocial factors in the aetiology and prognosis of coronary heart disease: systematic review of prospective cohort studies. *BMJ* 318: 1460–7.

Hill A. 1965. The environment and disease: association or causation? *Proc R Soc Med* 58: 295–300.

Hlatky M, Lam L, Lee K, et al. 1996. Job strain and the prevalence and outcome of coronary artery disease. *Circulation* 92: 327–33.

Houle C, Berthelot J-M, David P, et al. 1996. *Project on Matching Census 1986 Database and Manitoba Health Care Files: Private Households Component*. Ottawa: Statistics Canada.

House J, Kessler R, Herzog A. 1990. Age, socioeconomic status, and health. *Millbank Q* 68: 383–411.

Howard G, Anderson R, Johnson N, et al. 1997. Evaluation of social status as a contributing factor to the stroke belt region of the United States. *Stroke* 28: 936–40.

Howard G, Anderson R, Sorlie P, et al. 1994. Ethnic differences in stroke mortality between non-Hispanic whites, Hispanic whites, and blacks. *Stroke* 25: 2120–25.

Howard G, Russell G, Anderson R, et al. 1995. Role of social class in excess black stroke mortality. *Stroke* 26: 1759–63.

Howden-Chapman P, Blakely T, Blaiklock A, et al. 2000. Closing the health gap. *NZ Med J* 113: 301–2.

Howden-Chapman P, Isaacs N, Crane J, et al. 1996. Housing and health: the relationship between research and policy. *International Journal of Environmental Health Research* 6: 173–85.

Humphreys K, Carr-Hill R. 1991. Area variations in health outcomes: Artefact or ecology. *Int J Epidemiol* 20: 251–8.

Iversen G. 1974. Recovering individual data in the presence of group and individual effects. *Am J Sociol* 79: 420–34.

Iversen L, Andersen O, Andersen P, et al. 1987. Unemployment and mortality in Denmark, 1970–80. *BMJ* 295: 879–84.

Jaro M. 1995. Probabilistic linkage of large public health data files. *Stat Med* 14: 491–8.

Jenkins C. 1983. Social environment and cancer mortality in men. *N Engl J Med* 308: 395–8.

Jensen J. 1978. *Minimum Income Levels and Income Equivalence Scales.* Wellington: Department of Social Welfare.

Jensen J. 1988. *Income Equivalences and the Estimation of Family Expenditure on Children.* Wellington: Department of Social Welfare (unpublished).

Kaplan G, Pamuk E, Lynch J, et al. 1996. Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ* 312: 999–1003.

Kaplan H. 1991. Social psychology of the immune system: a conceptual framework and review of the literature. *Soc Sci Med* 33: 909–23.

Karasek R, Theorell T. 1989. *Healthy Work.* New York: Basic Books.

Kaufman J, Cooper R. 1999. Seeking causal explanations in social epidemiology. *Am J Epidemiol* 150: 113–20.

Kaufman J, Long A, Liao Y, et al. 1998. The relation between income and mortality in US. Blacks and Whites. *Epidemiol* 9: 147–55.

Kawachi I, Kennedy BP, Glass R. 1999. Social capital and self-rated health: a contextual analysis. *Am J Public Health* 89: 1187–93.

Kawachi I, Kennedy BP, Lochner S, et al. 1997. Social capital, income inequality, and mortality. *Am J Public Health* 87: 1491–8.

Kawachi I, Marshall S, Pearce N. 1991. Social class inequalities in the decline of coronary heart disease among New Zealand men, 1975–1977 to 1985–1987. *Int J Epidemiol* 20: 393–8.

Kennedy B, Kawachi I, Glass R, et al. 1998. Income distribution, socioeconomic status, and self-rated health in the United States: multilevel analysis. *BMJ* 317: 917–21.

Kennedy B, Kawachi I, Prothrow-Stith D. 1996. Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States. *BMJ* 312: 1004–7.

Kirkwood B. 1988. *Essentials of Medical Statistics.* Oxford: Blackwell Scientific Publications.

Kitagawa E, Hauser P. 1973. *Differential Mortality in the United States: A study in socioeconomic epidemiology.* Cambridge, MA: Harvard University Press.

Kogevinas M, Pearce N, Susser M, et al. 1997. *Social Inequalities and Cancer.* Lyon: IARC Scientific Publications.

Koskinen S, Martelin T. 1994. Why are socioeconomic mortality differences smaller among women than among men? *Soc Sci Med* 38: 1385–96.

Koskinen S, Martelin T, Valkonen T. 1996. Socioeconomic differences in mortality among diabetic people in Finland: five year follow up. *BMJ* 313: 975–8.

Krieger N. 1994. Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med* 39: 887–903.

Krieger N. 1999. Sticky webs, hungry spiders, buzzing flies, and fractal metaphors: on the misleading juxtaposition of 'risk factor' versus 'social' epidemiology. *J Epidemiol Community Health* 53: 678–80.

Krieger N. 2001. A glossary for social epidemiology. *J Epidemiol Community Health* 56: 693–700.

Krieger N, Chen J, Selby J. 1999. Comparing individual-based and household-based measures of social class to assess class inequalities in women's health: a methodological study of 684 US women. *J Epidemiol Community Health* 53: 612–23.

Krieger N, Davey Smith G. 2000. Re: Seeking causal explanations in social epidemiology [letter]. *Am J Epidemiol* 151: 831–2.

Krieger N, Williams D, Moss N. 1997. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health* 18: 341–78.

Krieger N, Zierler S. 1996. What explains the public's health? – A call for epidemiological theory. *Epidemiol* 7: 107–9.

Kuh D, Ben-Shlomo Y. 1997. *A Life Course Approach to Chronic Disease Epidemiology*. Oxford: Oxford University Press.

Kuh D, Power C, Blane D, et al. 1997. Social pathways between childhood and adult health. In D Kuh, Y Ben-Shlomo (eds) *A Life Course Approach to Chronic Disease Epidemiology* (pp.169–98). Oxford: Oxford University Press.

Kunst A. 1997. *Cross-national Comparisons of Socioeconomic Differences in Mortality [thesis]*. Rotterdam: Erasmus University.

Kunst A, Cavelaars A, Groenhof F, et al. 1996. *Socioeconomic Inequalities in Morbidity and Mortality in Europe: A comparative study*. Rotterdam: Department of Public Health, Erasmus University.

Kunst A, del Rios M, Groenhof F, et al. 1998a. Socioeconomic inequalities in stroke mortality: an international overview. *Stroke* 29: 2285–91.

Kunst A, Groenhof F. 1996. Potential sources of bias in 'unlinked' cross-sectional studies. In A Kunst, A Cavelaars, F Groenhof, et al (eds) *Socioeconomic Inequalities in Morbidity and Mortality in Europe: A comparative study* (pp.99–146). Rotterdam: Department of Public Health, Erasmus University.

Kunst A, Groenhof F, Anderson O, et al. 1999. Occupational class and ischaemic heart disease mortality in the United States and 11 European countries. *Am J Public Health* 89: 47–53.

Kunst A, Groenhof F, Borgan J-K, et al. 1998b. Socioeconomic inequalities in mortality: methodological problems illustrated with three examples from Europe. *Rev Epidemiol Sante Publique* 46: 467–79.

Kunst A, Groenhof F, Mackenbach J, et al. 1998c. Mortality by occupational class among men 30–64 years in 11 European countries. *Soc Sci Med* 46: 1459–76.

Kunst A, Groenhof F, Mackenbach J, et al. 1998d. Occupational class and cause specific mortality in middle-aged men in 11 European countries: comparison of population based studies. *BMJ* 316: 1636–42.

Kunst A, Mackenbach J. 1994a. International variation in the size of mortality differences associated with occupational status. *Int J Epidemiol* 23: 742–50.

Kunst A, Mackenbach J. 1994b. The size of mortality differences associated with educational level in nine industrialized countries. *Am J Public Health* 84: 932–7.

Lahelma E, Valkonen T. 1990. Health and social inequities in Finland and elsewhere. *Soc Sci Med* 31: 257–65.

Lantz P, House J, Lepkowski J, et al. 1998. Socioeconomic factors, health behaviours, and mortality. *JAMA* 279: 1703–8.

Last J. 1995. *A Dictionary of Epidemiology*. New York: Oxford University Press.

Leclerc A, Lert F, Fabien C. 1990. Differential mortality: some comparisons between England and Wales, Finland and France, based on inequality measures. *Int J Epidemiol* 19: 1001–10.

Leino-Arjas P, Liira J, Mutanen P, et al. 1999. Predictors and consequences of unemployment among construction workers: prospective cohort study. *BMJ* 319: 600–5.

Leon D. 1998. Commentary: Unequal inequalities across Europe. *BMJ* 316: 1642.

Leon D, Lithell H, Vagero D, et al. 1998. Reduced fetal growth rate and increased risk of death from ischaemic health disease: cohort study of 15,000 Swedish men and women born 1915–29. *BMJ* 317: 241–5.

Lewis G, Sloggett A. 1998. Suicide, deprivation, and unemployment: record linkage study. *BMJ* 317: 1283–6.

Liberatos P, Link B, Kelsey J. 1988. The measurement of social class in epidemiology. *Epidemiol Rev* 10: 87–121.

Link B, Phelan J. 1995. Social conditions as fundamental causes of disease. *J Health Social Behavior* (extra issue): 80–94.

Liu K. 1988. Measurement error and its impact on partial correlation and multiple linear regression analyses. *Am J Epidemiol* 127: 864–74.

Lochner K, Pamuk E, Makuc D, et al. 2001. State-level income inequality and individual mortality risk: a prospective, multilevel study. *Am J Public Health* 91: 385–91.

Lund E, Bona K. 1993. Reduced breast cancer mortality among fisherman's wives in Norway. *Cancer Causes Control* 4: 283–7.

Lynch J, Kaplan G. 2000. Socioeconomic position. In L Berkman, I Kawachi (eds) *Social Epidemiology* (pp.13–35). New York: Oxford University Press.

Lynch J, Kaplan G, Cohen R, et al. 1994. Childhood and adult socioeconomic status as predictors of mortality in Finland. *Lancet* 43: 524–7.

Lynch J, Kaplan G, Cohen R, et al. 1996. Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? *Am J Epidemiol* 144: 934–42.

Lynch J, Kaplan G, Pamuk E, et al. 1998. Income inequality and mortality in metropolitan areas of the United States. *Am J Public Health* 88: 1074–80.

Lynch J, Kaplan G, Salonen J. 1997a. Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Soc Sci Med* 44: 809–19.

Lynch J, Kaplan G, Shema S. 1997b. Cumulative impact of sustained economic hardship on physical, cognitive, psychological, and social functioning. *N Engl J Med* 337: 1889–95.

Lynch J, Krause N, Kaplan G, et al. 1997c. Workplace demands, economic reward and progression of carotid atherosclerosis. *Circulation* 96: 302–8.

Lynch J, Krause N, Kaplan G, et al. 1997d. Workplace conditions, socioeconomic status, and the risk of mortality and acute myocardial infarction: the Kuopio ischaemic heart disease risk factor study. *Am J Public Health* 87: 617–22.

Lynge E, Andersen O. 1997. Unemployment and cancer in Denmark, 1970–1975 and 1986–1990. In M Kogevinas, N Pearce, M Susser, et al (eds) *Social Inequalities and Cancer* (pp.353–9). Lyon: IARC Scientific Publications.

Macintyre S. 1997. The Black report and beyond what are the issues? *Soc Sci Med* 44: 723–45.

Macintyre S, Ellaway A, Der G, et al. 1998. Do housing tenure and car access predict health because they are simply markers of income or self-esteem? A Scottish study. *J Epidemiol Community Health* 52: 657–64.

Mackenbach J. 1995. Public health epidemiology. *J Epidemiol Community Health* 49: 333–4.

Mackenbach J, Kunst A. 1997. Measuring the magnitude of socioeconomic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Soc Sci Med* 44: 757–71.

Mackenbach J, Kunst A, Cavelaars A, et al. 1997. Socioeconomic inequalities in morbidity and mortality in western Europe. *Lancet* 349: 1655–9.

Mackenbach J, Kunst A, Groenhof F, et al. 1999. Socioeconomic inequalities in mortality among women and among men: an international study. *Am J Public Health* 89: 1800–6.

Mackenbach J, van de Mheen H, Stronks K. 1994. A prospective cohort study investigating the explanation of socioeconomic inequalities in health in the Netherlands. *Soc Sci Med* 38: 299–308.

Makela P, Valkonen T, Martelin T. 1997. Contribution of deaths related to alcohol use to socioeconomic variation in mortality: register based follow up study. *BMJ* 315: 211–16.

Marang-van de Mheen P, Davey Smith G, Hart C, et al. 1998. Socioeconomic differentials in mortality among men within Great Britain: time trends and contributory causes. *J Epidemiol Community Health* 52: 214–18.

Marmot M, McDowall M. 1986. Mortality decline and widening social inequalities. *Lancet* 274–6.

Marmot M, Shipley M. 1996. Do socioeconomic differences in mortality persist after retirement? 25 year follow-up of civil servants from the first Whitehall study. *BMJ* 313: 1177–80.

Marmot M, Theorell T. 1988. Social class and cardiovascular disease: the contribution of work. *Int J Health Serv* 18: 659–74.

Marmot MG, Bosma H, Hemingway H, et al. 1997a. Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet* 350: 235–9.

Marmot MG, Bosma H, Hemingway H, et al. 1997b. Contribution of job control to social gradient in coronary disease [letter – authors' reply]. *Lancet* 350: 1405.

Marmot MG, Shipley M, Rose G. 1984. Inequalities in death – specific explanations of a general pattern? *Lancet* 1003–6.

Marshall J, Hastrup J. 1996. Mismeasurement and the resonance of strong confounders: uncorrelated errors. *Am J Epidemiol* 143: 1069–78.

Marshall S, Kawachi I, Pearce N, et al. 1993. Social class differences in mortality from diseases amenable to medical intervention in New Zealand. *Int J Epidemiol* 22: 255–61.

Martelin T. 1994. Mortality by indicators of socioeconomic status among the Finnish elderly. *Soc Sci Med* 38: 1257–78.

Martelin T, Koskinen S, Valkonen T. 1998. Sociodemographic mortality differences among the oldest old in Finland. *J Gerontol B Psychol Sci Soc Sci* 53: S83–90.

Martikainen P. 1990. Unemployment and mortality among Finnish men, 1981–85. *BMJ* 301: 407–11.

Martikainen P, Valkonen T. 1996. Excess mortality of unemployed men and women during a period of rapidly increasing unemployment. *Lancet* 348: 909–12.

Martikainen P, Valkonen T. 1998a. Do education and income buffer the effects of death of spouse on mortality? *Epidemiol* 9: 530–4.

Martikainen P, Valkonen T. 1998b. The effects of differential unemployment rate increases of occupation groups on changes in mortality. *Am J Public Health* 88: 1859–61.

Martikainen P, Valkonen T. 1999. Bias related to the exclusion of the economically inactive in studies on social class differences in mortality. *Int J Epidemiol* 28: 899–904.

MatchWare Technologies I. 1998. *Automatch Generalised Record Linkage System, Version 4.2: User's manual.* Kennebunk, Maine: MatchWare Technologies, Inc.

McCarthy M. 1999. Transport and health. In M Marmot, R Wilkinson (eds) *Social Determinants of Health* (pp.17–43). Oxford: Oxford University Press.

McDonough P, Duncan G, Williams D, et al. 1997. Income dynamics and adult mortality in the United States, 1972 through 1989. *Am J Public Health* 87: 1476–83.

McEwen B. 1998. Protective and damaging effects of stress mediators. *N Engl J Med* 338: 171–9.

McEwen B, Seeman T. 1999. Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostatis and allostatic load. In N Adler, M Marmot, B McEwen, et al (eds) *Socioeconomic Status and Health in Industrial Nations: Social, psychological and biological pathways* (pp.30–47). New York: The New York Academy of Sciences.

Miller J, Korenmand S. 1994. Poverty and children's nutritional status in the United States. *Am J Epidemiol* 140: 233–43.

Ministry of Health. 1999. *Our Health, Our Future.* Wellington, New Zealand: Ministry of Health.

Montgomery S, Bartley M, Cook D, et al. 1996. Health and social precursors of unemployment in young men in Great Britain. *J Epidemiol Community Health* 50: 415–22.

Montgomery S, Cook D, Bartley M, et al. 1998. Unemployment, cigarette smoking, alcohol consumption and body weight in young British men. *Eur J Pub Health* 8: 21–7.

Montgomery S, Cook D, Bartley M, et al. 1999. Unemployment pre-dates symptoms of depression and anxiety resulting in medical consultation in young men. *Int J Epidemiol* 28: 95–100.

Morgan M. 1983. Measuring social inequality: occupational classifications and their alternatives. *Community Med* 5: 116–24.

Morgenstern H. 1995. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health* 16: 61–81.

Morgenstern H. 1998. Ecologic Studies. In K Rothman, S Greenland (eds) *Modern Epidemiology* (pp.459–80). Philadelphia: Lippincott-Raven.

Morris J, Cook D. 1991. A critical review of the effect of factory closures on health. *Br J Indust Med* 48: 1–8.

Morris J, Cook D, Shaper A. 1994. Loss of employment and mortality. *BMJ* 308: 1135–9.

Moser K, Fox A, Jones D. 1984. Unemployment and mortality in the OPCS longitudinal study. *Lancet* 1324–8.

Moser K, Fox A, Jones D, et al. 1986. Unemployment and mortality: further evidence from the OPCS longitudinal study 1971–81. *Lancet* 365–266.

Moser K, Goldblatt P, Fox A, et al. 1987. Unemployment and mortality: comparison of the 1971 and 1981 longitudinal study census samples. *BMJ* 294: 86–90.

Moser K, Pugh H, Goldblatt P. 1988. Inequalities in women's health: looking at mortality differentials using an alternative approach. *BMJ* 296: 1221–4.

Muntaner C. 1999. Invited commentary: social mechanisms, race, and social epidemiology. *Am J Epidemiol* 150: 121–6.

Murray J, Lopez A. 1996. *The Global Burden of Disease: A comprehensive assessment of the mortality from disease, injuries, and risk factors in 1990 and projected to 2020.* Boston: Harvard School of Public Health.

Mustard C, Derksen S, Berthelot M, et al. 1997. Age-specific education and income gradients in morbidity and mortality in a Canadian province. *Soc Sci Med* 45: 383–97.

National Health Committee. 1998. *The Social, Cultural and Economic Determinants of Health in New Zealand: Action to improve health.* Wellington, New Zealand: National Health Committee.

Newcombe H. 1988. *Handbook of Record Linkage: Methods for health and statistical studies, administration and business.* Oxford: Oxford University Press.

Nordic Statistical Secretariat. 1988. *Occupational Mortality in the Nordic Countries 1971–1980.* [Cited in: F Diderichsen, J Hallqvist. 1997. Trends in occupational mortality among middle-aged men in Sweden 1961–1990. *Int J Epidemiol* 26: 782–7.] Copenhagen: Nordic Statistical Secretariat.

Norstrom T. 1988. Deriving relative risks from aggregate data. 1. Theory. *J Epidemiol Community Health* 42: 333–5.

Pappas G, Queen S, Hadden W, et al. 1993. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl J Med* 329: 103–9.

Pearce N. 1996. Traditional epidemiology, modern epidemiology, and public health. *Am J Public Health* 86: 678–83.

Pearce N, Bethwaite P. 1997. Social class and male cancer mortality in New Zealand, 1984–87. *NZ Med J* 110: 200–2.

Pearce N, Cryer P. 1986. Analysis of the component of a linear trend in proportions. *Am J Epidemiol* 124: 127–33.

Pearce N, Davis P, Smith A, et al. 1983a. Mortality and social class in New Zealand. I: Overall male mortality. *NZ Med J* 96: 281–5.

Pearce N, Davis P, Smith A, et al. 1983b. Mortality and social class in New Zealand. II: Male mortality by major disease groupings. *NZ Med J* 96: 711–16.

Pearce N, Davis P, Smith A, et al. 1984. Mortality and social class in New Zealand. III: Male mortality by ethnic group. *NZ Med J* 97: 31–5.

Pearce N, Davis P, Smith A, et al. 1985. Social class, ethnic group, and male mortality in New Zealand, 1974–78. *J Epidemiol Community Health* 39: 9–14.

Pearce N, Howard J. 1986. Occupation, social class and male cancer mortality in New Zealand, 1974–78. *Int J Epidemiol* 15: 456–62.

Pearce N, Marshall S, Borman B. 1991. Undiminished social class mortality differences in New Zealand men. *NZ Med J* 104: 153–6.

Pearce N, Pomare E, Marshall S, et al. 1993. Mortality and social class in Maori and non-Maori New Zealand men: changes between 1975–77 and 1985–87. *NZ Med J* 106: 193–6.

Phillips A, Davey Smith G. 1991. How independent are 'independent' effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 44: 1223–31.

Phillips A, Davey Smith G. 1992. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med* 11: 953–61.

Phillips A, Wannamathee S, Walker M, et al. 1996. Life expectancy in men who have never smoked and those who have smoked continuously: 15-year follow up of large cohort of middle aged British men. *BMJ* 313: 907–8.

Pocock J, Shaper A, Cook D, et al. 1980. British regional heart study: geographic variations in cardiovascular mortality, and the role of water quality. *BMJ* 1243–9.

Pocock S, Shaper A, Cook D, et al. 1987. Social class differences in ischaemic heart disease in British men. *Lancet* 197–201.

Pomare E, Keefe-Ormsby V, Ormsby C, et al. 1995. *Hauora: Maori Standards of Health III*. Wellington: Eru Pomare Maori Health Research Centre.

Poole C, Kaufman J. 2000. What does standard adjustment for downstream mediators tell us about social effect pathways? [Conference poster, Society for Epidemiologic Research, 15–17 June 2000, Seattle]. *Am J Epidemiol* 151, Abstract 208.

Power C, Matthews S, Manor O. 1996. Inequalities in self rated health in the 1958 birth cohort: lifetime social circumstances or social mobility. *BMJ* 313: 449–53.

Riley MC. i. D-R, AV. 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Pub Health* 88: 216–22. (1963). Special problems of sociological analysis, *Sociological Research I: A Case Approach* (pp.700–25). New York: Harcourt, Brace, and World Inc.

Robins J, Greenland S. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiol* 3: 143–55.

Robinson W. 1950. Ecological correlations and the behavior of individuals. *Am Sociol Rev* 15: 351–7.

Rogot E, Sorlie P, Johnson N. 1986. Probabilistic methods in matching census samples to the national death index. *J Chron Dis* 39: 719–34.

Rogot E, Sorlie P, Johnson N. 1992a. Life expectancy by employment status, income, and education in the national longitudinal mortality study. *Public Health Rep* 107: 457–61.

Rogot E, Sorlie P, Johnson N, et al. 1992b. *A Mortality Study of 1.3 million Persons by Demographic, Social and Economic Factors: 1979–1085 follow-up*. Bethesda: National Institutes of Health.

Rosso S, Faggiano F, Zanetti R, et al. 1997. Social class and cancer survival in Turin. *J Epidemiol Community Health* 51: 30–4.

Rothman K, Greenland S. 1998. *Modern Epidemiology*. Philadelphia: Lippincott-Raven.

Salmond C, Crampton P. 2000. Deprivation and Health. In P Howden-Chapman (ed) *Social Inequalities in Health: New Zealand 1999* (pp.9–64). Wellington, New Zealand: Ministry of Health.

Salmond C, Crampton P, Hales S, et al. 1999. Asthma prevalence and deprivation: a small area analysis. *J Epidemiol Community Health* 53: 476–80.

Salmond C, Crampton P, Sutton F. 1998. NZDep91: A New Zealand index of deprivation. *Aust NZ J Public Health* 22: 835–7.

Sapolsky R, Alberts S, Altmann J. 1997. Hypercortisolism associated with social subordinance or social isolation among wild baboons. *Arch Gen Psychiatry* 54: 1137–43.

Schnall P, Landsbergis P. 1994. Job strain and cardiovascular disease. *Annu Rev Public Health* 15: 381–411.

Schrijvers C, Stronks K, van de Mheen H, et al. 1999. Explaining educational differences in mortality: the role of behavioural and material factors. *Am J Public Health* 89: 535–40.

Schrijvers C, van de Mheen H, Stronks K, et al. 1998. Socioeconomic inequalities in health in the working population: the contribution of working conditions. *Int J Epidemiol* 27: 1011–18.

Selvin H. 1958. Durkheim's suicide and problems of empirical research. *Am J Sociol* 63: 607–19.

Shaper A. 1984. Geographic variations in cardiovascular mortality in Great Britain. *Br Med Bull* 40: 366–73.

Shivley C, Clarkson T. 1994. Social status and coronary artery atherosclerosis in female monkeys. *Arterioscler Thromb* 14: 721–6.

Shkolinov V, Leon D, Adamets S, et al. 1998. Educational level and adult mortality in Russia: An analysis of routine data 1979 to 1994. *Soc Sci Med* 47: 357–69.

Shouls S, Congdon P, Curtis S. 1996. Modelling inequality in reported long term illness in the UK: combining individual and area characteristics. *J Epidemiol Community Health* 50: 366–76.

Siegrist J. 1996. Adverse health effects of high effort/low reward conditions. *J Occup Health Psychol* 1: 27–41.

Siegrist J, Peter R, Junge A, et al. 1990. Low status control, high effort at work and ischemic heart disease: prospective evidence from blue-collar men. *Soc Sci Med* 31: 1127–34.

Sihvonen A, Kunst A, Lahelma E, et al. 1998. Socioeconomic inequalities in health expectancy in Finland and Norway in the late 1980s. *Soc Sci Med* 47: 303–15.

Singh G, Yu S. 1996a. Trends and differentials in adolescent and young adult mortality in the United States, 1950 through 1993. *Am J Public Health* 86: 560–4.

Singh G, Yu S. 1996b. US childhood mortality, 1950 through 1993: Trends and socioeconomic differentials. *Am J Public Health* 86: 505–12.

Sloggett A, Joshi H. 1994. Higher mortality in deprived areas: community or personal disadvantage? *BMJ* 309: 1470–4.

Sloggett A, Joshi H. 1998. Deprivation indicators as predictors of life events 1981–1992 based on the UK ONS longitudinal study. *J Epidemiol Community Health* 52: 228–33.

Soobader M-J, LeClere F. 1999. Aggregation and the measurement of income inequality: effects on morbidity. *Soc Sci Med* 48: 733–44.

Sorlie P, Backlund E, Johnson N, et al. 1993. Mortality by Hispanic status in United States. *JAMA* 270: 2464–8.

Sorlie P, Backlund E, Keller J. 1995. US mortality by economic, demographic, and social characteristics: the national longitudinal mortality study. *Am J Public Health* 85: 949–56.

Sorlie P, Rogot E. 1990. Mortality by employment status in the national longitudinal mortality study. *Am J Epidemiol* 132: 983–92.

Sorlie P, Rogot E, Anderson R, et al. 1992a. Black-white mortality differences by family income. *Lancet* 340: 346–50.

Sorlie P, Rogot E, Johnson N. 1992b. Validity of demographic characteristics on the death certificate. *Epidemiol* 3: 181–4.

Statistics New Zealand. 1999. *New Zealand Now: Incomes.* Wellington, New Zealand: Statistics New Zealand.

Statistics Sweden. 1982. Dodsfallsregister 1961–1970. SCB PM 1981:5 [in Swedish] [cited in: F Diderichsen, J Hallqvist. 1997. *Int J Epidemiol* 26: 782–7.] Stockholm: Statistics Sweden.

Stefansson C-G. 1991. Long-term unemployment and mortality in Sweden, 1980–1986. *Soc Sci Med* 32: 419–23.

Stronks K, Van De Mheen H, Van Den Bos J, et al. 1997. The interrelationship between income, health and employment status. *Int J Epidemiol* 26: 592–600.

Suadicani P, Hein P, Gyntelberg F. 1993. Are social inequalities as associated with the risk of coronary heart disease a result of psychosocial working conditions? *Atherosclerosis* 101: 165–75.

Sundquist J, Johansson S. 1997. Indicators of socioeconomic position and their relation to mortality in Sweden. *Soc Sci Med* 45: 1757–66.

Susser M. 1994. The logic in ecological: I. The logic of analysis. *Am J Public Health* 84: 825–9.

Susser M. 1998. Does risk-factor epidemiology put epidemiology at risk? Peering into the future. *J Epidemiol Community Health* 52: 608–11.

Susser M, Susser E. 1996a. Choosing a Future for Epidemiology: I. Eras and paradigms. *Am J Public Health* 86: 668–73.

Susser M, Susser E. 1996b. Choosing a Future for Epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *Am J Public Health* 86: 674–7.

Tobias M, Christie S. 1998. Standard ICD groupers for population health: Draft 21/10/98. Wellington, New Zealand: Ministry of Health.

Turrell G, Mengersen K. 2000. Socioeconomic status and infant mortality in Australia: a national study of small urban areas, 1985–89. *Soc Sci Med* 50: 1209–25.

Turrell G, Oldenburg B, McGuffog I, et al. 1999. *Socioeconomic Determinants of Health: Towards a national research program and a policy and intervention agenda.* Canberra: Centre for Public Health Research, School of Public Health, Queensland University of Technology (in association with the Health Inequalities Research Collaboration).

Vagero D, Erikson R. 1997. Socioeconomic inequalities in morbidity and mortality in western Europe. *Lancet* 350: 516.

Vagero D, Leon D. 1994. Effect of social class in childhood and adulthood on adult mortality. 343.

Vagero D, Lundberg O. 1989. Health inequalities in Britain and Sweden. *Lancet* 1 July, 35–6.

Valkonen T. 1993. Problems in the measurement and international comparisons of socioeconomic differences in mortality. *Soc Sci Med* 36: 409–18.

Valkonen T, Sihvonen A, Lahelma E. 1997. Health expectancy by level of education in Finland. *Soc Sci Med* 44: 801–8.

van de Mheen D. 1998. *Inequalities in Health, to be continued? A life-course perspective on socioeconomic inequalities in health*. Erasmus University.

van de Mheen H, Stronks K, Schrijvers C, et al. 1999. The influence of adult ill health on occupational class mobility and mobility out of and into employment in The Netherlands. *Soc Sci Med* 49: 509–18.

van Doorslaer E, Wagstaff A, Bleichrodt H, et al. 1997. Income-related inequalities in health: some international comparisons. *Journal of Health Economics* 16: 93–112.

Victora C, Huttly S, Fuchs S, et al. 1997. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *Int J Epidemiol* 26: 224–7.

Westerling R, Gullberg A, Rosen M. 1996. Socioeconomic differences in 'avoidable' mortality in Sweden 1986–1990. *Int J Epidemiol* 25: 560–7.

Wilkinson R. 1992. Income distribution and life expectancy. *BMJ* 304: 165–8.

Wilkinson R. 1996. *Unhealthy Societies: The afflictions of inequality*. London: Routledge.

Wilkinson R. 1999. Health, hierarchy, and social anxiety. In N Adler, M Marmot, B McEwen, et al (eds) *Socioeconomic Status and Health in Industrial Nations: Social, psychological and biological pathways* (pp.48–63). New York: The New York Academy of Sciences.

Wilson N. 2000. Labour Force Status and Health. In P Howden-Chapman, M Tobias (eds) *Social Inequalities in Health in New Zealand: New Zealand 1999*. Wellington, New Zealand: Ministry of Health.

Winkleby M, Jatulis D, Frank E, et al. 1992. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* 82: 816–20.

Woodward A. 1996. What makes populations vulnerable to ill health? *NZ Med J* 109: 265–7.

Wright E. 1995. The class analysis of poverty. *Int J Health Serv* 25: 85–100.

Yen I, Kaplan G. 1999. Neighbourhood social environment and risk of death: multilevel evidence from the Alameda County Study. *Am J Epidemiol* 149: 898–907.

Appendix A: Socioeconomic Mortality Gradients – Literature Review

Box 5: Overview of Appendix A

Following a brief overview of socioeconomic position generally, separate sections are presented for education, labour force status, income, and asset wealth (car access and housing tenure). Each section commences with a description of the socioeconomic factor and its theoretical reason for being associated with mortality, followed by a *selected* review of empirical studies describing aspects of the association of each socioeconomic factor with mortality *relevant to this report* and the NZCMS.

Education

Education is both a marker of socioeconomic position generally at one point of the life-course, and a marker of specific factors that may be causally relevant: knowledge of healthy lifestyles, and an attitude to life that involves investing in one's future. Education has been commonly used as a measure of socioeconomic mortality gradients in the US and in European cross-national comparisons. Higher educational attainment has been consistently associated with lower mortality, including independently of income and material factors (although the latter are likely pathways from education to mortality risk).

Labour force status

Non-labour force status (eg, retired, not seeking work) is associated with mortality largely due to health selection – if you are in poor health and about to die, then you are probably unable to work. Some of the association is also probably due to other factors such as confounding by socioeconomic position.

Unemployment (ie, actively seeking and available for work) might be related to health through one of four mechanisms: material disadvantage, psychosocial stress, health-related behaviour, and as part of the life-course. The association might also be spurious due to health selection, a point that is debated in the literature.

Occupational class

Occupational class is the most commonly used socioeconomic factor to measure socioeconomic mortality gradients. It is usually operationalised in a Weberian manner, being a measure of the education requirements and monetary payoffs of a particular occupation, which in turn presumably reflects access to resources and life chances that affect health. Occupational class mortality gradients have been documented in numerous contexts. An important bias may arise when *current* occupation is used to assign occupational class.

Income

Income is considered to influence health outcomes by facilitating access to medical care, good housing, good diet, good working conditions, less exposure to noxious environment, and social amenities – although this materialistic explanation has recently been challenged by a psychosocial explanation. Income has been strongly associated with mortality. It is unclear whether this association is causal, or whether income is a proxy for other socioeconomic factors or exposures earlier in the life-course that are actually causing the apparent association with mortality. That said, income measured at one point in time is highly correlated with income earlier in life. The observed income–mortality gradient may also be biased by health selection, whereby poor health causes a drop in income.

Asset wealth

Asset wealth is difficult to measure directly – hence proxies like car access and housing tenure have been used. They are both strong predictors of health status.

There is varying use of the terms social class, socioeconomic position and socioeconomic status in the health literature. For example, social class is often used as an umbrella term that includes a number of possible socioeconomic factors (eg, income, education, occupational scales), particularly in the US (Berkman and Macintyre 1997; Krieger et al 1997). In this report, the framework as outlined by Davis et al (Davis et al 1997) has been followed. At the first level of categorisation, 'social class' is differentiated from 'socioeconomic status'.

> *'A 'class', in Marx's original sense of the word, may be defined as a group of people who share a common economic situation, based upon their relationship to the means of production, and whose interests inevitably conflict with those of others. Indeed the term 'class' is arguably rendered meaningless when removed from the context of the full theoretical arsenal of Marxism.*
>
> '*Socioeconomic status', on the other hand, must be understood in terms of* socioeconomic stratification, *the patterned unequal distribution of opportunities, advantages, resources and power among subgroups of a given population. Distinct 'socioeconomic strata' may thus be said to exhibit differential life chances, living standards and associated cultural practices. While it is common place for researchers to use the term 'class' to refer to the type of social stratification measured in their research, it is clear that socioeconomic status is in fact the construct they most seek to operationalise.*' (Davis et al 1997, p.8, non-italics in original).

Regarding the last point made by Davis et al, occasional studies attempt to examine socioeconomic inequalities in health from a Marxist class perspective (eg, Wright 1995), but the majority of studies, even when using the term 'social class' (eg, Davey Smith et al 1998b; Marmot and Theorell 1988; Pearce et al 1991), are more correctly examining socioeconomic stratification. That is, the majority of studies use a Weberian framework, in which society is viewed as being stratified by more than just access to the means of production, for example prestige, social status, and factors influencing life chances such as education (Edgell 1993; Liberatos et al 1988; Lynch and Kaplan 2000).

Recently, social epidemiologists have recommended the use of the term 'socioeconomic position' rather than 'socioeconomic status' to represent the overarching concept (Krieger 2001; Lynch and Kaplan 2000). This recommendation arises from the fact that the word 'status' is closely aligned with prestige or social hierarchy measures, whereas 'position' more comfortably subsumes all of income, education, occupational class, material and social deprivation, absolute and relative poverty and so on. This report follows this recommendation and as such updates the use of the term socioeconomic status in the PhD report on which this report is based.

It is important to consider whether socioeconomic position is best viewed as one underlying construct or dimension, or whether the separate socioeconomic factors (eg, income and education) are best viewed as separate socioeconomic exposures with independent relationships with health. For example, assume that low-income people have twice the mortality of high-income people. First, we could interpret this finding as meaning that low socioeconomic position people have twice the mortality of high socioeconomic position people. That is, we are interpreting income as a proxy for the underlying construct of socioeconomic position. Alternatively, we could interpret this finding as meaning that low-income people have twice the mortality of high-income people, independently of other factors (including other socioeconomic factors). That is, we are assuming that if we experimentally manipulated peoples' incomes such that everyone became a high-income earner, but nothing else changed, then the people with previously low incomes would now have the same mortality rate as the previously high-income people.

It may be argued that the consideration of the independent effects of separate socioeconomic factors on health is of little use, if not meaningless, as:

- in real-life (dis)advantage on separate socioeconomic factors cluster within individuals, and it is thus contrived to consider the effect of, say, income on mortality, independent of other socioeconomic factors
- even if the independent effect of separate socioeconomic factors could be estimated, the joint effects of (dis)advantage on multiple socioeconomic factors are unlikely to be additive due to the synergistic or multiplicative effects of multiple (dis)advantage
- separate socioeconomic factors cannot be assigned a discrete meaning; rather the underlying construct of socioeconomic position is an organising principle in society over which the exposures that actually affect health (eg, self efficacy, psychosocial stress, risk behaviours) are unevenly distributed.

Therefore, it may be argued that it is more parsimonious and realistic to consider socioeconomic position as a single construct. The corollary of this is that we should create composite indices of socioeconomic position, and measure multiple socioeconomic factors only to gain accuracy for this composite measure.

Creating such composite indices of socioeconomic position means that the 'determinants of health' panel in the upstream determinants of health in Figure 2 is reduced to a single measure of socioeconomic position. It would still be feasible to determine whether 'government policies' affect the socioeconomic construct, to determine whether 'government policies' modify the observed effect of the socioeconomic construct on health, and finally, to determine what pathways of midstream and downstream determinants appear to mediate the effect of the socioeconomic construct on health. However, it would not be feasible to determine the effect on health for separate socioeconomic factors. (Further problems also arise with comparisons between studies that will invariably use different composite indices.)

The position taken in this report and by the NZCMS more generally is that, in most circumstances, it is desirable to estimate the independent effect of separate socioeconomic factors on health. For example, it is a relevant policy (and aetiological) question to ask: 'Would income maintenance affect coronary heart disease mortality, independent of other socioeconomic factors?' 'Would policies to decrease unemployment reduce the suicide rate, independent of other socioeconomic factors?'; and (over the long term) 'Would raising the educational status of New Zealanders affect mortality?'. Reviews of the measurement of socioeconomic position highlight that most researchers and theoreticians advocate the use of multiple measures of socioeconomic position – not composite indices (Davis et al 1997; Liberatos et al 1988). Moreover, most studies of the association of numerous socioeconomic factors (eg, income, education, and occupational class) with health outcomes find that while there is some overlap between socioeconomic factors in the relationship with health status, there are clear independent contributions from each socioeconomic factor in most studies (Liberatos et al 1988; Sorlie et al 1995).

In the remainder of this section, the association of separate socioeconomic factors with mortality: education, labour force status, occupational class, car access, housing tenure and income are considered. The ordering of these socioeconomic factors approximates a life-course chronology of socioeconomic exposure among 25–64 year olds. At age 25 years, most people have completed their education. Occupational class captures the potential to convert education into income and asset accumulation, and income (in most studies) applies to a period immediately preceding the commencement of follow-up. As described above these socioeconomic factors also capture different elements of socioeconomic position, in addition to being measures of socioeconomic position at different points of the life-course. In using this framework of socioeconomic factors upon which to structure this section, the life-course position of each variable and the possible different causal relation between each variable and mortality have been emphasising. It is conceded, however, that much of the literature treats each socioeconomic factor as exchangeable. If it is true that individual socioeconomic factors are no more than markers of some larger entity, then the causal mechanisms connecting each socioeconomic factor to mortality will be essentially the same.

It is not my intention to systematically review the entire literature for each socioeconomic factor – the literature is simply too voluminous, and the review would become too repetitious. Rather, priority has been given to the following types of studies:

- The New Zealand studies by Pearce and colleagues on occupational class, the only robust evidence in New Zealand on the association of individual-level measures of socioeconomic position with mortality. Following the development of indices of small area deprivation (Crampton et al 1997; Salmond et al 1998), Salmond, Crampton and colleagues have conducted analyses of the association of deprivation and mortality in New Zealand (Salmond and Crampton 2000; Salmond et al 1999). Analyses in the NZCMS will make extensive use of this small area measure of deprivation in multilevel analyses to examine the possible ecological effect of deprivation on mortality controlling for individual-level socioeconomic factors. However, this report focuses on the individual-level measures of socioeconomic position as a first step, and treats the small area deprivation index principally as a validation tool.
- Studies with a similar design to the NZCMS, ie, linked census–mortality studies such as the OPCS Longitudinal Study and the US NLMS.
- Studies of socioeconomic mortality gradients in developed countries – a great deal has been written on social factors and health in developing countries, but this work will not be reviewed here.
- Cross-national comparison studies, particularly those by Kunst, Mackenbach and colleagues in the European Union Working Group on Socioeconomic Inequalities in Health. Cross-national comparison studies provide both a summary of the socioeconomic mortality gradient for a number of countries, *and* point to variations in socioeconomic mortality gradients between countries. 'Large dissimilarities would imply that socioeconomic inequalities in health are highly sensitive to specific national circumstances' (Kunst et al 1998d, p.1636). If socioeconomic mortality gradients vary across countries then this may provide clues as to how socioeconomic mortality gradients are modified by the most upstream determinants in Figure 2 (eg, government policies), or may provide clues as to the varying importance across countries of specific causes of death (eg, heart disease) and pathways (eg, alcohol consumption).
- Studies that examine the independent effects of two or more socioeconomic factors considered simultaneously. Such studies may give clues as to the aetiology of socioeconomic mortality gradients, and guidance to policy-makers as to which socioeconomic factor(s) to target for intervention.
- Studies that examine possible pathways between socioeconomic position and mortality (eg, psychosocial, behavioural, and materialistic). While midstream or pathway variables are unavailable for analysis in the NZCMS (except smoking in the 1981 and 1996 censuses), it is important to have an understanding of the likely pathways for the observed socioeconomic mortality gradients.

Regarding the last bullet point it must be noted that there is a vast and growing literature on the midstream factors that may mediate the association of socioeconomic position with health. Types of approaches that researchers have taken in tackling this topic include a life-course approach (eg, Blane 1999; Kuh and Ben-Shlomo 1997), focusing on psychosocial pathways (eg, Marmot et al 1997a) including the possible physiological changes effected by psychosocial factors (eg, Brunner 1997; Brunner and Marmot 1999; McEwen 1998; Sapolsky et al 1997; Shivley and Clarkson 1994), focusing on health related behaviours (eg, Emmons 2000; Lynch et al 1997a), and focusing on material factors (eg, Schrijvers et al 1999). Moreover, many of these pathway mechanisms have a high degree of overlap between specific socioeconomic factors. It is not possible in this report to do full justice to these areas of research.

The literature reviews for each socioeconomic factor are structured in three parts:

- a general overview, including the meaning of the socioeconomic factor and theoretically plausible mechanisms of action
- a review of empirical findings of the 'highlighted' studies
- a brief conclusion.

1 Education

Education has a distinct empirical advantage as a measure of socioeconomic position: it is fixed relatively early in life, and therefore is not prone to health selection effects. Disadvantages include the secular trends in educational attainment over time that make comparisons between cohorts difficult and the fact that it is not necessarily related closely to adult socioeconomic position. An additional problem is that education often fails to evenly stratify people – many people are assigned to one category of education (eg, nil formal qualifications), and only a few have attained the highest educational attainment. However, this problem may be largely historical now with the increasing educational attainment of populations in recent decades.

Education may be related to mortality in the following ways. First, education is a marker of parental socioeconomic position and socioeconomic conditions of childhood and early adulthood, giving it a firm place in the life-course (Lynch and Kaplan 2000). It may be that childhood and early adulthood socioeconomic factors (as indicated by education) affect adult mortality by latent effects, or by adding to a life-course of cumulative social (dis)advantage, or by determining adult socioeconomic factors which in turn affect mortality (Kuh et al 1997).

Second, education may be thought of as a measure of 'health capital' – education may influence health outcomes through its influence on choice of lifestyle behaviours (eg, exercise, diet), problem-solving capacity, and values (eg, importance of preventive health-related behaviour) (Liberatos et al 1988).

> '*Exposure to formal education involves gathering facts, learning concepts, and finding out how to access information. It may provide a set of cognitive resources that have broad potential to influence health.*' (Lynch and Kaplan 2000, p.22).

Winkleby et al found that education was more strongly related to risk factors for cardiovascular disease than income and occupation and that finding offers support to this 'health capital' mechanism (Winkleby et al 1992).

Third, people with different educational attainment vary in unobservable ways, including time preferences for investing in their future (including their future health) (Fuchs 1979). Thus, some third factor may be associated with both higher education and lower mortality risk (ie, confounding).

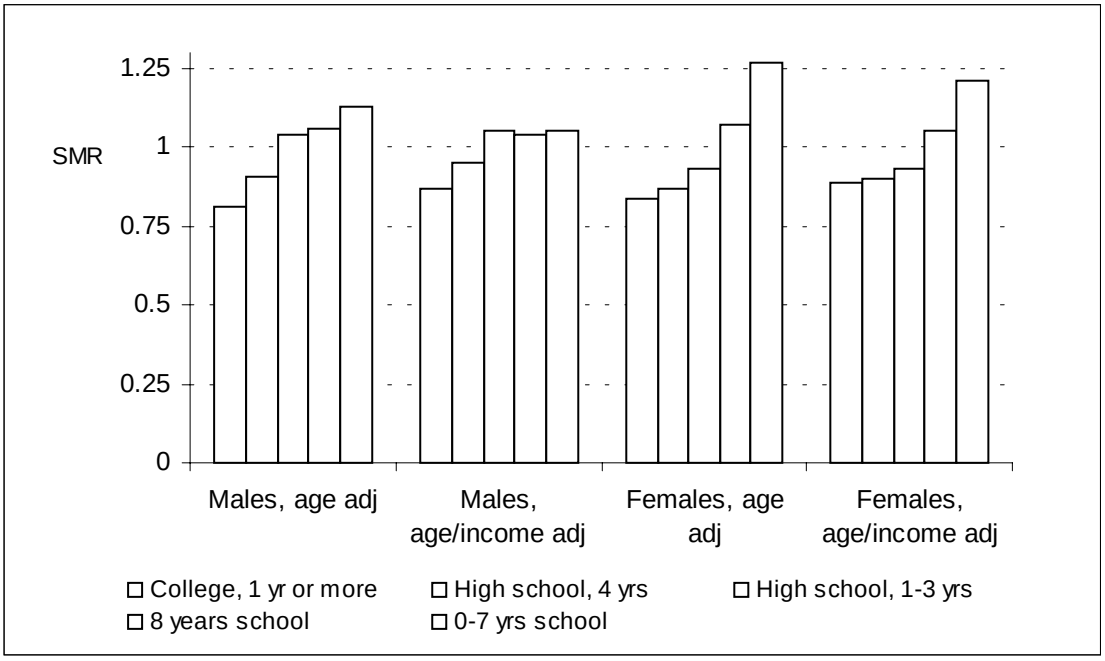
1.1 Findings

1.1.1 The association of education with mortality

The OPCS Longitudinal Study is limited with regard to education as a measure of socioeconomic position, as educational qualifications below A-levels were not elicited in the 1971 census. Thus, 81.3% and 85.3% of males and females aged 18–70 at the 1971 census have no elicited education (Fox and Goldblatt 1982). For the approximately 15% of the 18–70 year old population with A-levels or higher, the SMRs compared to the total population were notably reduced: 69 and 77 for those with a university degree, for males and females respectively; 83 and 88 for those with a qualification higher than A-levels (but not a university degree); and 93 and 78 for those with A-levels.

Education is more commonly and comprehensively used as a measure of socioeconomic position in the USA and Europe. Hauser and Kitagawa (1973) reported a 64% greater mortality among 25–64 year old men in 1960 with 0–4 years of school education compared to four years or more of college education, using data from the 1960 US Matched Records Study. The comparable difference for females was 105%, although this was mainly due to a much lower than expected mortality among the females with four or more years of college education – otherwise the socioeconomic mortality gradients were similar across sexes. The education–mortality gradient was greater at younger ages. Figure 16 shows these SMR results, but with the categories of education aggregated somewhat to allow further standardisation by family income. Standardising by family income explained less than 50% of the socioeconomic mortality gradient by education for each sex.

Figure 16: Standardised mortality ratios (SMRs) for 25–64 year olds by education in the 1960 Matched Records Study, standardised by age and by age and family income

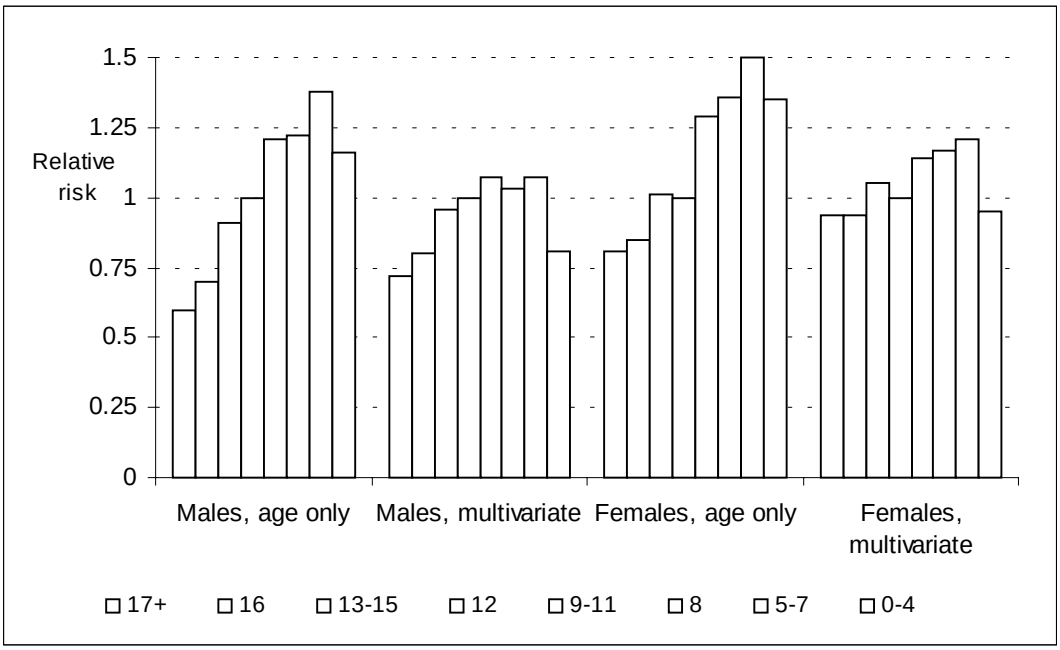


Source: Kitagawa and Hauser 1973.

Comparing the above results from the 1960 Matched Records Study with mortality differentials by education during 1971–84, Feldman et al concluded that (relative) socioeconomic mortality gradients for males had increased in the US within two decades (Feldman et al 1989). Among men aged 55–64 years, the excess mortality for the low versus highly educated increased from 17% to 50% between the two time periods. Much of this increased socioeconomic mortality gradient for males was due to an increasing gradient for heart disease – in 1960 there was no apparent gradient by education for heart disease deaths. For females, Feldman et al reported little change over the two decades. Feldman et al also found that controlling for family income and other baseline risk factors (ie, smoking, BMI, blood pressure and serum cholesterol) did not explain much of the heart disease gradient in 1971–84, corroborating the findings emerging from the Whitehall Study in the UK for occupational status (Marmot et al 1984). Pappas et al have also demonstrated increasing mortality gradients over time in the US, using the Matched Records Study as the baseline measurement (Pappas et al 1993).

The US National Longitudinal Mortality Study measured the association of education with mortality during the 1980s (Sorlie et al 1995). Adjusting for age only, strong associations of the number of years of education with mortality were found for males and females aged 25–44 and 45–64 year olds, but somewhat diminished associations were found for those aged 65 years and older. Figure 17 shows the association for 45–64 year males and females adjusting for age only, and also adjusting for age, race, labour force status, income, marital status, and household size. This latter multivariate adjustment explained much of the education association. Some of the multivariate adjustment may be considered as controlling for confounders (eg, race, marital status and household size), some as adjusting for confounders/mediators of the association of education with mortality (eg, labour force status, and family income), and possibly also introducing some bias from differential health selection (ie, controlling for labour force status).

Figure 17: Relative risk of death by years of education among 45–64 year olds in the US National Longitudinal Mortality Study, adjusted for age only and age, race, labour force status, income, marital status, and household size



Source: Sorlie et al 1995.

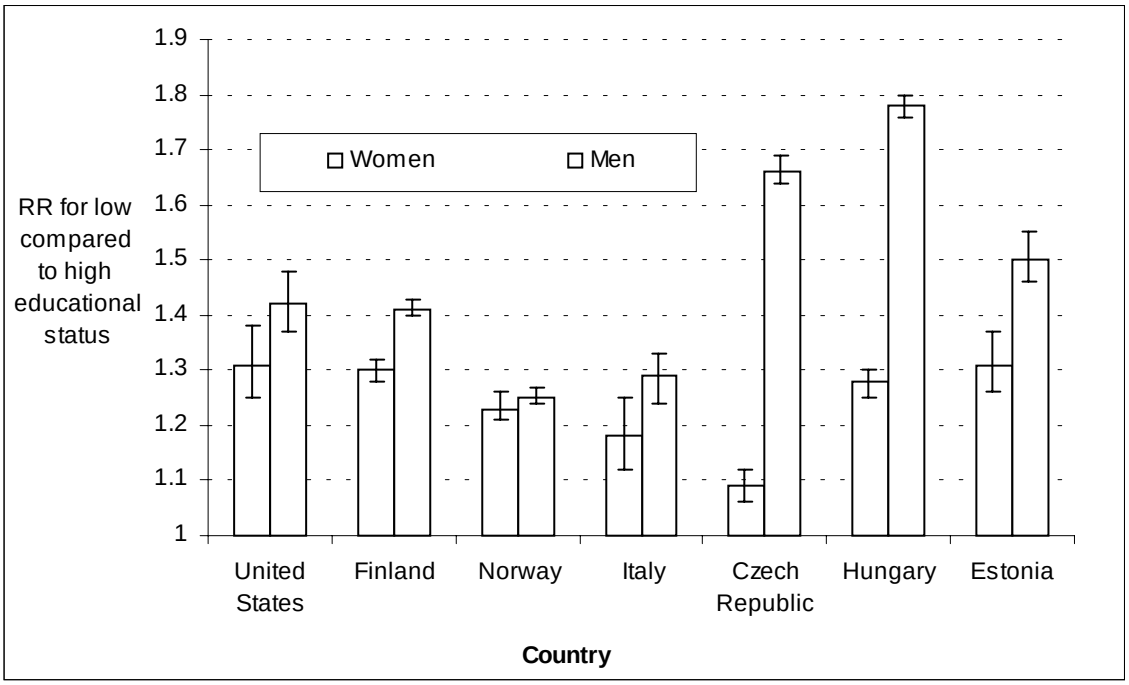
Note: Reference group for education was 12 years of education.

Educational inequalities in mortality tend to be greater at younger ages – at least when measured in terms of relative risks. An extreme example of this variation by age is a study of deaths occurring in 1993–94 in Barcelona and Madrid that found that 25–34 year old males and females with no education had five to seven times the mortality rate of 25–34 year olds with secondary or higher education (Borrell et al 1999). This very high educational inequality among young adults was probably due to AIDS, highlighting the rapid social patterning of emerging diseases. Among older age groups, the educational inequalities were more modest. For example, among 55–64 year olds there was an approximately 10% excess mortality rate among those with no education compared to those with a secondary or higher education.

1.1.2 Cross-national comparison studies of educational mortality gradients

Perhaps the most authoritative overview of the association of education with mortality is a cross-national comparison study conducted by Mackenbach, Kunst, and others (Mackenbach et al 1999). Seven countries were included with data mostly from the 1980s: the United Sates, Finland, Norway, Italy, the Czech Republic, Hungary and Estonia. To maximise international comparability, education was aggregated into two broad categories: up to higher secondary school, and higher secondary school and tertiary. Comparisons were for females and males aged 20–74 – except the Czech Republic where the age range was 25–64.

Figure 18: Relative risk of death by educational attainment among men and women aged 20–74



Source: Mackenbach et al 1999.

Note: age range 20–64 years for Czech Republic. Error bars are 95% confidence intervals.

With the exception of a relatively low educational difference in all-cause mortality among females in the Czech Republic (rate ratio 1.09), the educational disparity was approximately the same for females in the remaining countries (rate ratios 1.18 to 1.31). For males there was greater variation in educational inequalities, with rate ratios ranging from 1.25 in Norway to 1.66 for the Czech Republic and 1.78 in Hungary. Thus, all-cause mortality inequalities by education tended to be smaller among females. Much of this sex difference was due to the causes of death: females were more likely to die of cancer, and educational inequalities in cancer were relatively small. However, there was also an interesting pattern of educational inequalities being greater among *females* for cardiovascular diseases, and greater among *males* for lung cancer, respiratory diseases, and external causes. These findings suggest a different patterning of health-related behaviours by sex and educational status – an expectation supported by cross-national studies of educational inequalities in health-related behaviours (Cavelaars 1998; Cavelaars et al 2000; Cavelaars et al 1997). The question then is 'why does the social patterning of health-related behaviours vary by sex?'. Mackenbach et al (1999) hypothesise that males may tend to respond to the material disadvantage and psychosocial stress associated with low educational status by taking up relatively dangerous behaviours (eg, smoking, and excessive alcohol consumption). Conversely, females respond differently (eg, obesity rather than excess alcohol consumption) due to access to different types of behavioural responses. This hypothesis may explain both some of the generally greater educational inequalities in all-cause mortality for males compared to females, and some of the differences by cause of death (eg, cardiovascular disease).

An additional observation from the cross-national comparison study of Mackenbach et al (1999) is that, for males at least, there was a tendency for the educational mortality gradients to be greatest in the former communist countries. Likewise, educational inequalities in mortality in Russia have been found to be as large, if not larger, than in western countries (Shkolinov et al 1998). These findings are of importance for two reasons. First, it illustrates that large, if not larger, socioeconomic mortality gradients are present in a political and social system that ostensibly aimed to annul inequalities. Second, it provides an interesting natural experiment: in communist countries the correlation between education and income is not as strong as in the west. For example, a medical career required a long education but did not result in a high income. Moreover, income differentials were less in the communist countries. Income and material factors are therefore unlikely to explain the educational mortality gradients in the former communist countries. Some direct evidence to support material factors being an unlikely pathway between education and health in the former communist countries comes from a cross-sectional survey of cardiovascular risk factors in the Czech Republic (Bobak et al 1999). This survey found that material factors were inconsistently associated with cardiovascular risk factors, and that material factors did not explain the strong association of education with risk factors. On the contrary, many studies in western countries suggest that educational mortality gradients are largely 'explained' by income or material factors (Backlund et al 1999; Lantz et al 1998; Schrijvers et al 1999).

Bobak et al (2000) extended the work in the Czech Republic with a case-control study of myocardial infarction (Bobak et al 2000) – a part of the MONICA studies. They found a strong independent association of education with myocardial infarction, but not so for material factors (measured by the proxy of car ownership). When two exposures are both highly correlated and measured with error (as are education and income in western countries) it is not possible to reliably discriminate the independent effects of either variable (Davey Smith and Phillips 1992; Phillips and Davey Smith 1991). Thus this Czech study is important due to the lack of correlation of education and car ownership, and suggests that education is more important than material factors. Thus, educational inequalities in mortality may not be due to a life-course pathway whereby education determines adult income and material factors which, in turn, affects mortality. Rather, it may be that educational inequalities are mediated by other 'health capital' (including health-related behaviours) or psychosocial pathways as described at the beginning of this section. The two notable limitations to this study are whether it is generalisable to other countries, and whether car ownership is an adequate measure of material circumstances.

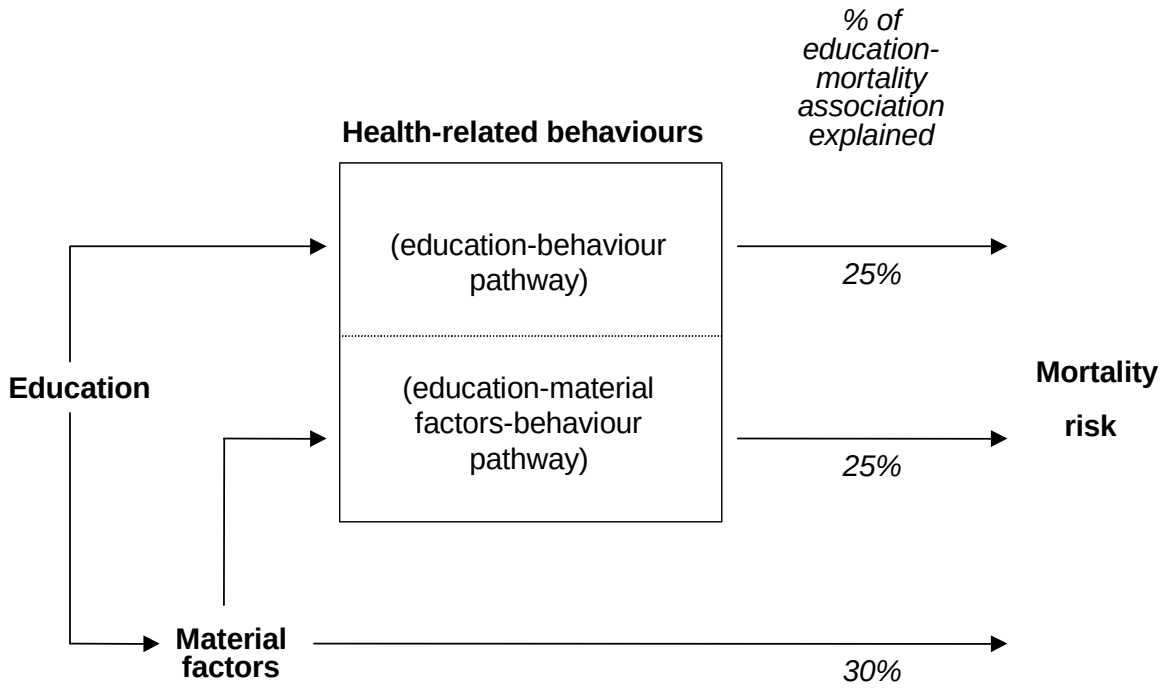
1.1.3 Material versus behavioural pathways from education to mortality: the Dutch Longitudinal Study on Socioeconomic Health Differences

Returning to western countries, several studies have demonstrated the uneven distribution of behaviours by socioeconomic group, but it is not enough to explain socioeconomic mortality gradients (Davey Smith et al 1994; Davey Smith et al 1990; Doornbos and Kromhout 1990; Lantz et al 1998; Lynch et al 1997a). Schrijvers et al examined how much of the association of education with mortality was explained by health-related behaviours versus materialist factors, together and independently (Schrijvers et al 1999). They used the Longitudinal Study on Socioeconomic Health Differences in the Netherlands, an aselect sample of 19,000 15–74 year olds, followed-up from 1991–96. Education was categorised at four levels: higher vocational school and university; intermediate vocational school and intermediate or higher secondary school; lower vocational school and lower secondary school; and primary school only. The rate ratios of mortality adjusting for confounders (age, sex, marital status, religious

affiliation, degree of urbanisation) were 1.00 (reference), 1.28 (0.85 to 1.93), 1.48 (1.02 to 2.13), 1.64 (1.13 to 2.40) for the four levels of education, respectively. The behavioural factors included were alcohol consumption, smoking, BMI, and physical activity. The material factors included were financial problems in the preceding year, employment status (working, unemployed, disabled, retired, and housewife), and a proxy for income (combined index of health insurance, housing tenure, and car access). By conducting separate regression models adding either the behaviours or the material factors, and adding both the behaviours and material factors, Schrijvers et al estimated the contribution of behavioural and materialist factors to the educational mortality gradient. As shown in Figure 19, they estimated that about 30% of the association of education with mortality was explained directly by material factors, a quarter was explained directly by health-related behaviours, and another quarter by a pathway through material factors to behaviour. Thus, perhaps over 50% of the education-mortality association was explained in total by material factors. However, as discussed in Chapter 1, analyses such as those by Schrijvers et al (1999) are subject to many possible sources of error, including:

- assuming that the reduction in the association of education with mortality due to controlling for material factors and behaviours was due to allowing for pathways including these variables may be an overestimate. The reduction may have been due to both the intermediary *and* confounding components of the material and behavioural variables
- conversely, there was undoubtedly measurement error of the behaviours and material factors, probably causing an underestimation of the contribution of behaviour and material factors to the educational mortality gradient
- finally, if psychosocial variables (or other variables) confound the association of material and behavioural factors with mortality, then excluding psychosocial variables may cause an overestimation of the contribution of behavioural and material factors.

Figure 19: Percentage of the association of education with mortality due to behavioural and material pathways in the Dutch Longitudinal Study on Socioeconomic Health Differences



Source: Schrijvers et al 1999.

1.2 Conclusion

Education is a strong underlying determinant of health and mortality risk. It is unclear whether this association represents:

- a sorting process of youth consequent on family circumstances, intelligence and a predisposition to invest in the future, where these (confounding) characteristics are the true underlying determinants of future mortality risk
- a pathway to increased access to material resources that affects mortality risk
- a pathway to healthier behaviours and lifestyles consequent on improved knowledge obtained through education
- or other processes (eg, psychosocial influences).

As a measure of socioeconomic position, education has the advantage of (usually) being fixed by early adulthood, thus representing a stable measure of socioeconomic position and a measure of socioeconomic position from earlier in the life-course. Disadvantages of education include the varying distribution of education between age cohorts, and measurement imprecision (eg, years of education rather than educational achievement).

With the increasingly recognised problems using occupational class for cross-national comparison studies, it is likely that education will be used more commonly in cross-national comparison studies in the future.

2 Labour force status

Mortality risk is markedly elevated among some categories of the labour force (eg, the permanently sick, and the retired). This elevated mortality risk is undoubtedly due in large part to health selection. However, there is also usually an elevated mortality risk for the unemployed (but still in the workforce) versus the employed (eg, Martikainen 1990; Martikainen and Valkonen 1996; Morris et al 1994; Moser and Fox 1984; Moser et al 1986; Moser et al 1987). The interpretation of this mortality differential between the unemployed and employed is problematic: it could be due to a real effect of unemployment on mortality risk, or it could be due to residual health selection into unemployment rather than out of the active labour force entirely. Additionally, the elevated mortality risk for the unemployed compared to the employed may be due to confounding, whereby extraneous factors (eg, personality) are associated with both the chance of unemployment and mortality risk. Most commentators, however, have concluded that while health selection and confounding may explain some of the association of unemployment with excess mortality, there is an underlying causal relationship (Bartley 1994; Bartley et al 1999).

Bartley and Owen (1996) have provided direct evidence of the magnitude of health selection out of employment and into either unemployment or the non-active labour force (Bartley and Owen 1996). Using 20 years of British general household survey data, they demonstrate that people with limiting longstanding illness were concentrated in the non-active labour force, but not notably among the unemployed. Of particular interest, they convincingly demonstrated differential health selection out of the employed to the non-active labour force (but not into the unemployed) such that:

- lower socioeconomic people with a limiting longstanding illness were much less likely to retain employment that higher socioeconomic people with a limiting longstanding illness
- among the non-active labour force, a much greater proportion of lower socioeconomic people had a limiting longstanding illness compared to higher socioeconomic people
- during times of high population unemployment rates (eg, mid-1980s), the proportion of health selection out of the employed and into the non-active labour force that was differential by socioeconomic position was much greater.

Bartley proposes four causal mechanisms that may mediate the relationship between unemployment and health (Bartley 1994; Bartley et al 1999): poverty or material disadvantage; psychosocial stress; health-related behaviour; and unemployment as part of the life-course.

Financial strain and lower incomes among the unemployed compared to the employed have been found to explain some of the variation in health status between the employed and unemployed. The need to borrow money after being made unemployed has also been linked to deterioration in physical health (Bartley et al 1999). Based on these findings, we would expect deterioration in the health of the unemployed if welfare benefits were reduced.

Loss of employment is a stressful life event, resulting in human reactions similar to those experienced in bereavement. One loses social support, self-efficacy, and a daily structure that accompanies employment, and one gains the new stressors of possible social isolation, exclusion and financial hardship. Such chronic stressors are associated with health status (Brunner and Marmot 1999; Ferrie et al 1995; Hemingway and Marmot 1999; Kaplan 1991; McEwen 1998), suggesting they may mediate some of the association of unemployment with health.

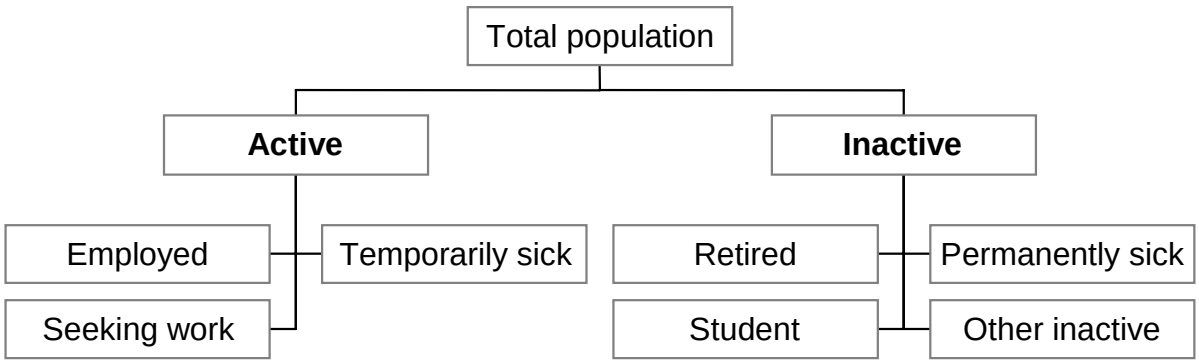
Health-related behaviours may be either confounders or mediators of the association of unemployment with health. Some of the studies of the association of unemployment with mortality imply they are confounders (eg, Morris et al 1994). There is not convincing evidence that loss of employment leads to deterioration in health-related behaviours (Morris and Cook 1991). Indeed, mass redundancy of construction workers in Finland saw an improvement in some health-related behaviours (decreased alcohol consumption and increased exercise) associated with long-term unemployment (Leino-Arjas et al 1999). It may be argued, however, that the acquisition of health-related behaviours must be seen in the context of life-course that includes past exposure to unemployment. For example, Montgomery et al have reported that there were significant associations of health-related behaviour at age 33 with previous bouts of unemployment, controlling for health-related behaviour at age 16 and other socioeconomic variables (Montgomery et al 1998). More directly, the ultimate self-destructive behaviour of suicide is strongly associated with unemployment (Lewis and Sloggett 1998; Moser et al 1984).

2.1 Findings

2.1.1 All categories of labour force

Figure 20 presents the categories of 'economic activity' derived from the 1971 census in the OPCS Longitudinal Study. The active–inactive economical activity distinction is similar to the labour force–non-labour force distinction.

Figure 20: 'Economic activity' categories used in the OPCS LS (1971 census)



Not surprisingly, the standardised mortality ratios (SMRs) for the permanently sick in the OPCS LS following the 1971 and 1981 censuses were markedly elevated (Bethune 1996; Fox and Goldblatt 1982) (eg, 332 for males aged 15–64 and 588 for females aged 15–59 in the 10 years following the 1981 census, where 100 is the SMR for all labour force status groups combined (Bethune 1996)). Likewise, SMRs were elevated for temporarily sick people in the 10 years following the 1981 census (332 and 558, respectively) and retired people (163 and 218, respectively) (Bethune 1996). These elevated SMRs are undoubtedly due, in large part at least, to health selection. Fox et al (1985) also report the trend over the 10 years following the 1971 census in the males 15–64 year old SMRs for the 'unoccupied' (essentially the 'permanently sick' and 'other inactive' categories (Harding 1995)). The SMR was in excess of 300 in the first two years following the census, but then steadily tracked down to about 150 in 1980 before rising again to nearly 300 in 1981 (Fox et al 1985). Assuming that the 1981 SMR was an aberrantly high result due to small numbers, then these results of Fox et al suggest a decrease in health selection with increasing time between measurement of economic (in)activity and mortality. Such a decrease in the apparent effect of health selection over time is exactly what one would expect – the excess mortality for the unoccupied should reduce over time as those selected into the unoccupied due to poor health die or return to good health, *and* some of the initially occupied become sick. However, it is also possible that some of the tracking down of the mortality rate over time was due to changing labour force status over time. For example, people in the non-labour force (ie, economically inactive) initially may move into employment with a lower mortality rate, and some people in the labour force (ie, economically active) initially may move in the opposite direction incurring a higher mortality rate.

2.1.2 Unemployment

The pattern of mortality differentials between the unemployed and employed differs to those described above. A seminal study is that by Moser et al (1984; Table 48). Using data from the OPCS Longitudinal Study for the 10-year period following the 1971 census, the age-adjusted SMR for unemployed males aged 15–64 years was 136, and additionally controlling for social class it was 121 (Moser et al 1984). (The UK census elicits most recent occupation for the unemployed, allowing an assignment of social class to the currently unemployed.) The SMRs, either age-adjusted or age and social class-adjusted, were higher in 1976–81 than 1971–75 (Table 48). If the association of unemployment with mortality was caused by health selection, one would expect the SMRs to *decrease* over time (as explained above). The results reported by Moser et al show no such decrease, and if anything an increase over time. Thus, this study offers strong evidence against the elevated mortality among unemployed men compared to all men being due to health selection. However, the possibility of confounding remains.

Table 48: SMRs of men unemployed at the 1971 census, aged 15–64 at death, in the OPCS LS

Time period	Age-adjusted		Age and social class-adjusted	
	SMR	(95% CI)	SMR	(95% CI)
1971–81	136	(122–152)	121	(108–135)
1971–75	129	(110–150)	122	(104–142)
1976–81	144	(122–168)	123	(105–143)

Source: Moser et al 1986.

Further studies using the OPCS Longitudinal Study have added evidence to support an effect of unemployment on mortality that is not due to health selection, including:

- Analyses by Moser et al that demonstrated an elevated SMR among the spouses and other female members of households with an unemployed male (Moser et al 1984; Moser et al 1986). As the health of the spouses and other female members of the household should not have caused the male to become unemployed due to health selection, Moser et al argue that the increased SMRs of the females supports an independent effect of unemployment. For example, both a loss of income and increase in stress resulting from the males unemployment could increase the mortality among all members of the household.
- Further analyses by Moser et al that found that the SMR for unemployed males following the 1971 and 1981 censuses were comparable (Moser et al 1987). If health selection were causing the association of unemployment with mortality, a weaker association may have been expected following the 1981 census due to the higher background unemployment rate in 1981. That is, one might expect that when unemployment rates are high (1981) 'ordinary' people are afflicted with unemployment, whereas when unemployment rates are low (1971) only the 'unordinary' (including the unhealthy) are afflicted by unemployment. As it might be expected that 'ordinary' people afflicted by unemployment were less likely to differ from 'ordinary' people remaining in employment on unmeasured confounders, this analysis also suggests that the excess mortality among the unemployed was not due to confounding or health selection.
- Moser et al also found that the SMRs for the unemployed actually *increased* for the first three years following both the 1971 and 1981 censuses (Moser et al 1987) (ie, a healthy worker effect). If health selection was responsible for the elevated mortality of the unemployed, one would expect the reverse pattern – a decrease in the unemployed SMRs in the years following the census.

In addition to the OPCS Longitudinal Study, other studies have also found an elevated mortality rate for the unemployed:

- Danish unemployed men and women aged 20–64 at the 1970 census had a 40–50 percent excess mortality rate compared to the unemployed in the following 10 years, controlling for occupation, housing tenure, region and marital status (Iversen et al 1987; Martikainen 1990). The mortality rates among the unemployed were over twice those among the employed for accidents, suicide, and other diseases (ie, where 'other' excludes the former two causes of death, and cancer and cardiovascular disease).
- US unemployed men in the 1980s had a relative risk of mortality of 2.06 (25–44 years) and 1.36 (45–64 years) compared to the unemployed, falling to 1.60 and 1.16, respectively, when controlling for race, income, education, marital status, and household size (Sorlie et al 1995). All male relative risks were statistically significantly different from 1.00 at the 1% level. Among females, both the univariate and multivariate-adjusted relative risks were non-significantly different from 1.00.

- Finnish unemployed men aged 30–54 years at the 1980 census had a relative risk of mortality of 2.41 compared to the employed in the following five years. Controlling for occupational socioeconomic position, education, sick leave and marital status, the relative risk was moderately reduced to 1.93 (95% CI 1.82 to 2.05) (Martikainen 1990). As with the Danish study above (Iversen et al 1987), relative risk was greatest for injury-related deaths and 'other' diseases. Further, Martikainen reported that 'selection for unemployment based on age, socioeconomic states, and marital state was evident (ie, confounding) but no such selection was detected based on health [(e, health selection)'. Finally, the excess mortality among the unemployed increased with duration of unemployment prior to census night.
- Italian (Turin) men aged 15–59 years at the 1981 census had a relative risk of mortality of 2.02 compared to employed men in the following nine years. Controlling for housing tenure the relative risk decreased to 1.93 (1.79 to 2.27) (Costa and Segnen 1987).
- Men in the British Regional Heart Study with continuous employment in the late 1970s, but a period of unemployment *not due to health reasons* in the early 1980s, had a relative risk of mortality in the late 1980s of 1.59 (1.20 to 2.11), compared to men that had continuous employment through both the late 1970s and early 1980s (Morris et al 1994). The strength of this study was the ability to exclude those who stated they became unemployed due to health reasons, thus reducing health selection effects. Controlling further for the potential confounders of social class, town, smoking, alcohol, and pre-existing illness at the initial screening (1978–80), the relative risk reduced modestly to 1.47 (1.10 to 1.96). The relative risk was similarly raised for both cancer and circulatory disease deaths. Morris et al pointed to this latter non-specificity of association as partial evidence against a causal relationship between unemployment and mortality, based on their assertion of biological plausibility for an association of unemployment with circulatory disease, but not with cancer. Otherwise, Morris et al conclude that their study is consistent with a causal relation between unemployment and mortality.
- Lewis and Sloggett found that suicide for both males and females combined was strongly associated with unemployment (and other labour force characteristics) in the OPCS Longitudinal Study (Lewis and Sloggett 1998). Controlling for age, time period and sex, the unemployed had an odds ratio of 3.14 (95% CI 2.44 to 4.02) compared to the employed. Controlling for marital status, social class, education, car access and housing tenure, the odds ratio decreased moderately to 2.58 (95% CI 1.97 to 3.38). Interestingly, there was no association in the latter multivariate model of education or social class with suicide. There was, however, some residual association of no car access and communal housing tenure with elevated suicide risk.

More recently Martikainen and Valkonen have moved their stance to argue that causal attribution of an increased mortality risk to unemployment remains difficult because these studies have controlled only for a limited number of confounders (Martikainen and Valkonen 1996). Moreover, despite controlling for confounders, inevitable inaccuracy in the measurement of these confounders may leave residual confounding effects on the association of unemployment and mortality (Davey Smith and Phillips 1992; Greenland 1980; Marshall and Hastrup 1996). Following Moser et al (1987), Martikainen and Valkonen also hypothesis that during times of high background unemployment one would expect the unemployed to be more similar to the employed than during times of low unemployment. That is, when background unemployment is high, anyone may be

afflicted with unemployment, but when background unemployment is low, affliction by unemployment is a less random event that is probably correlated with other (unmeasured) risk factors for mortality. Consistent with this hypothesis, but contrary to the findings of Moser et al (1987), Martikainen and Valkonen (1996) observed the rate ratio of mortality for the unemployed compared to the employed to decrease with increasing background unemployment rates during the recession in the early 1990s in Finland. Thus, they concluded that at least some of the observed association of unemployment with mortality is due to unmeasured confounders and/or health selection. In an editorial commenting on Martikainen and Valkonens' study, Bartley concluded that these findings may be equally consistent with health selection out of the workforce altogether, not just from employed to unemployed (Bartley 1996). Specifically, the employed with health problems cling to their employment, but comparable people with health problems afflicted by unemployment do not even bother to categorise themselves as unemployed when the chances of regaining work are minimal (ie, when background unemployment rates are high), but rather exit the workforce altogether (Bartley 1996). That is, there may actually be health selection from unemployed to non-workforce that outweighs any health selection from employed to unemployed, causing Martikainen and Valkonen (1996) to observe a smaller differential between the unemployed and employed in a period of high background unemployment. While a plausible alternative explanation, neither Bartley's nor Martikainen and Valkonens' explanations address why the findings of Martikainen (1996) and Moser et al (1987) are contradictory, given the same type of natural experiments that were being studied. It is to be noted that independent of the background unemployment rate, the probability of becoming unemployed and the experience of unemployment (eg, provision of welfare) may vary between countries.

In a further study by Martikainen and Valkonen, they again found results inconsistent with an independent effect of unemployment on mortality (Martikainen and Valkonen 1998b). In Finland, unemployment had increased from 4–5% in the late 1980s to a peak of about 19% in 1994. Using 1985 and 1990 Finnish census records linked to unemployment information from the Ministry of Labour, and mortality records for the period 1987 to 1993, they tracked changes in mortality rates in three bands of occupations: occupations experiencing small (4.5–8.2% points), intermediate (5.0–15.9%) and large (10.0–31.2%) increases in unemployment over the given time period. If unemployment has a short-term impact on mortality, then one would expect the mortality rate to increase most over the period among the occupations with rapidly rising unemployment. This variation in mortality change between the three groups was not observed – all three groups experienced the same relative changes in mortality. Further, the results were similar by age group and cause of death – even accidental and violent causes of death that may be more plausibly linked to unemployment in the short-term. The advantage of this study design was that by using each occupational band as its own control, residual confounding between the employed and unemployed was overcome. The disadvantage was that, while a novel test of the unemployment–mortality hypothesis, there may have been too much 'noise' to detect any unemployment effect. That 'noise' would include the fact that the majority of people in each band of occupations remained employed, and the relatively small mortality 'signal' (say a 20–30% excess mortality) among the unemployed was unable to be detected. Other secular changes in mortality by occupational groups may also have obscured the signal.

Another recent study, although not measuring mortality as an outcome, found definite evidence of health selection into *long-term* unemployment (Leino-Arjas et al 1999). Leino-Arjas et al followed a cohort of 586 Finnish construction workers over a period of four years during the recession in the early 1990s. In the four-year follow-up period 78.7% of the workers experienced some unemployment in the four-year follow-up period, and 19.6% experienced greater than 24 months of unemployment. Despite the unemployment arising in the context of mass redundancies, diseases at baseline were strong predictors of greater than 24 months unemployment, with mental disorder and skin disease remaining as significant predictors when simultaneously controlling for age, previous unemployment, and health-related behaviours. It is not clear whether the definition of unemployment used by Leino-Argas et al excluded people who were either not actively seeking work or not available for work – actively seeking work and availability for work are usual criteria that have to met before being termed unemployed (as opposed to being out of the workforce).

2.2 Conclusion

Being out of the labour force is strongly associated with a higher mortality risk – undoubtedly much of this excess mortality risk is due to health selection.

Within the labour force, there is debate as to whether unemployment truly confers an excess risk of mortality compared to being employed, or whether the observed association is due to health selection and/or confounding. The most measured conclusion to make is that:

- health selection effects exist, but are probably not solely responsible for the excess mortality among the unemployed compared to the employed and will vary depending on how stringent the definition of unemployment is
- confounding of the association of unemployment with mortality exists, but it is unlikely to explain all of the unemployment–mortality association.

An additional point is the importance of context. In a community that buffers the material penalties of being unemployed, minimises stigma, and provides alternative and affirming activities, it is likely that losing one's job would be less health-threatening than the same event occurring in a punitive and unsupportive environment. Such contextual issues disclose challenges for social epidemiology, as it limits our ability to compare studies of the health effects of unemployment between contexts.

3 Occupational class

Occupational classifications of socioeconomic position have been the most widely used in social epidemiology, and arguably represent the key underlying socioeconomic stratification in society. In the United Kingdom, the Registrar General's occupational classification of socioeconomic position was first used in 1911 to analyse infant mortality rates (Morgan 1983).

The ranking of occupations into socioeconomic strata may be divided into:

> '*... those based on public opinion of their level of esteem and those based on educational requirements and monetary payoffs (since education and income are conceptualized as allocating persons to different lifestyles and power positions. The first is referred to as a* prestige *perspective and is conceptualized as part of Weber's status domain; the second is considered to be a* socioeconomic *approach and falls under Weber's class domain.*' (Liberatos et al 1988, p.89, non-italic emphasis in original).

According to Weber, societal position is based on three dimensions: class, status and power. The status dimension (and hence *prestige* occupational class scales) captures prestige and honour in the community. Status gains 'access to life chances, based on social and cultural factors such as family background, lifestyle and social networks' (Liberatos et al 1988, p.89). The class dimension (and hence the *socioeconomic* occupational class scales) has an economic base, reflecting 'ownership and control of resources and is indicated by measures of income' (Liberatos et al 1988, p.89). The majority of occupational class scales reflect more the socioeconomic dimension than the prestige dimension.

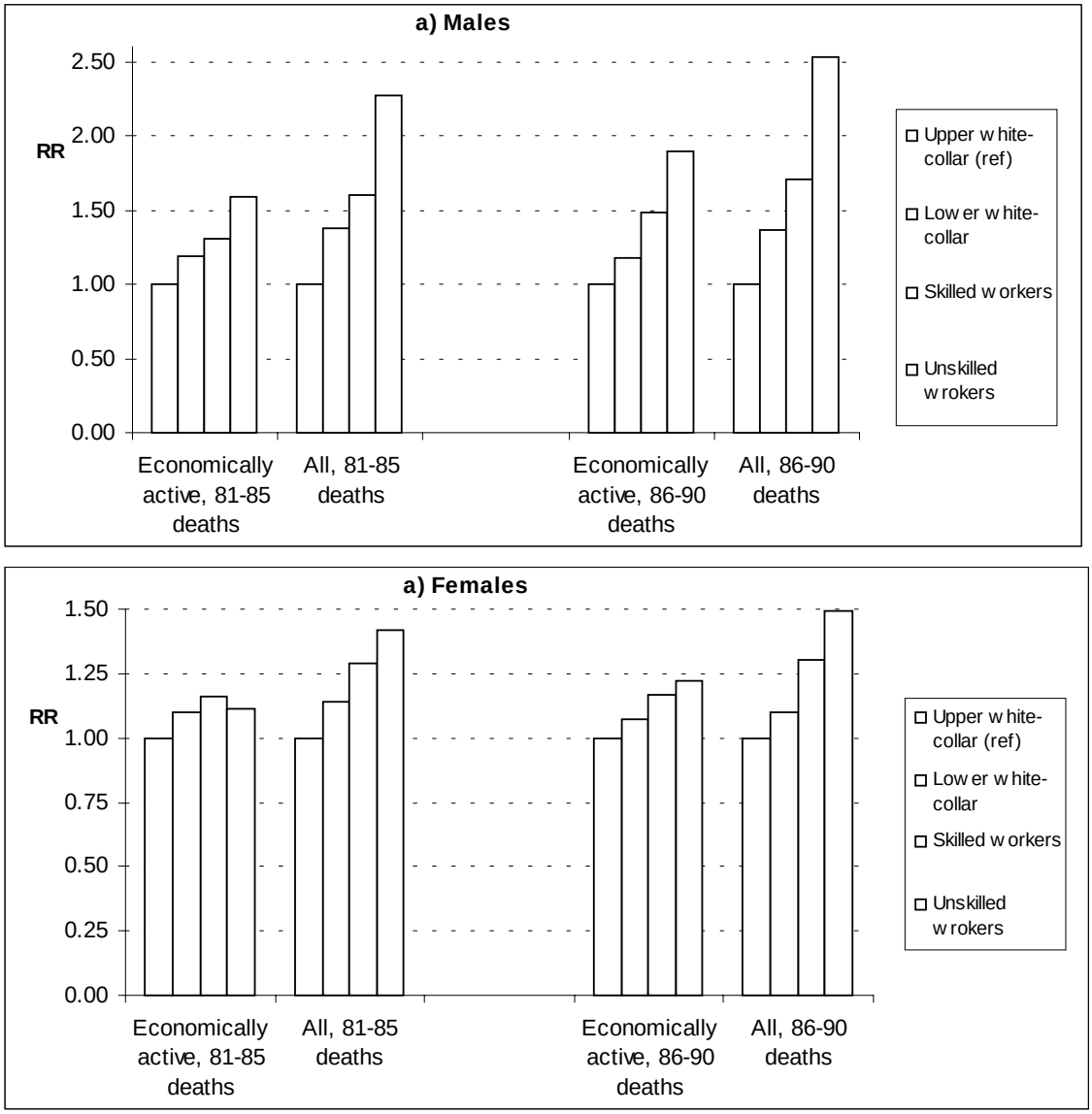
Occupational class measures can only be applied to those with a stated occupation, be it current, most recent, or usual. This may immediately exclude large sections of society depending on how occupational data is elicited. For example, in New Zealand Pearce and colleagues have been unable to investigate occupational class mortality gradients for females due to inadequate occupational data on death registration forms (Pearce et al 1991). Spousal occupation, or highest household occupation, have been used to circumvent the often encountered lack of occupational data for females (Krieger et al 1999), but this information is not available on mortality data in New Zealand.

Health selection is also a potential problem, in two ways. First, there may be health-related mobility up and down occupational classes (ie, drift health selection – see page 196). If this occurs, one would expect to see a steeper occupational class mortality gradient due to poor health causing a drift down the occupational classes. However, drift up and down occupational classes probably has a negligible effect on occupational class gradients (Fox et al 1985; Power et al 1996; van de Mheen et al 1999). Second, the probability of moving in and out of the workforce due to poor health may vary by occupational class (ie, *differential* health selection – see page 196). Differential health selection probably biases occupational class mortality gradients if data on *current* occupation only is available (Fox et al 1985; Goldblatt 1989; Kunst 1997; Kunst et al 1998b; Martikainen and Valkonen 1999; van de Mheen et al 1999). The direction of the bias is to underestimate the occupational class mortality gradient, as people in lower occupational classes are less likely to remain in the workforce if their health deteriorates compared to people in higher occupational classes (Bartley and Owen 1996). Fox et al concluded that after five years of follow-up in the OPCS Longitudinal Study this latter health selection effect was negligible (Fox et al 1985). Martikainen and Valkonen, however, conclude that the bias from differential health selection persists after five years of follow-up (Martikainen and Valkonen 1999). Figure 21 shows the age-standardised mortality rate ratios from the study by Martikainen and Valkonen. Two time periods were analysed, deaths in Finland during 1981–85 and during 1986–90, but for each analysis the occupational class determined at the 1980 census was used. The Finnish census elicits a previous occupation for those not currently economically active, allowing Martikainen and Valkonen to compare occupational class mortality gradients for the whole 35–64 year old populations to the gradients among just the economically active.

Restricting analyses to the economically active (ie, replicating the only possible occupational class analysis in many other countries when only current occupation is available) markedly underestimates the usual occupational class gradient (Figure 21):

- For male deaths in the first five years after the census, the excess mortality rate among unskilled workers compared to upper white-collar workers was underestimated by half.
- For female deaths in the first five years, the underestimate was even greater than half.
- In the second five years after the census, the underestimate was still apparent and sizeable for both sexes.

Figure 21: Age-standardised mortality rate ratio among 35–64 year old Finnish males and females during 1981–85 and 1986–90, by occupational class and economic activity measured in 1980



Source: Martikainen and Valkonen 1999

This diminution of mortality gradients may occur for socioeconomic factors other than occupational class, for example education. Using linked census–mortality studies from Finland, Norway, France and Italy (Turin), Kunst et al demonstrated that excluding the economically inactive reduced educational mortality gradients to the null by 43% to 90% among 50–59 year olds in the first five years of follow-up (Kunst et al 1998b). For example, among 50–59 year olds in Norway the mortality rate ratio for those with high compared to low educational status was 1.31 among the entire population, but excluding the economically inactive this rate ratio reduced by 52% to 1.15. Further analyses by Kunst et al also demonstrated that the underestimate of the educational mortality gradients was less for deaths occurring five to nine years after baseline: the bias to the null for the four linked census–mortality studies included was 11% to 35% during the latter time period. While the size of the health selection bias is reduced in the second five years of follow-up it is still not negligible, supporting Martikainen and Valkonens' (1999) findings described above for occupational class.

It is, however, incorrect to assume that the diminution of the above educational and occupational class mortality gradients is *solely* due to differential health selection. Labour force status *itself* may be an independent predictor (be it 'causal' or a proxy for something else) of mortality, and can therefore be a confounder of the association of occupational class or education with mortality. Undoubtedly, much of the attenuation of mortality gradients after exclusion of the economically inactive is due to differential health selection out of the labour force, but it is also plausible that labour force participation itself, in part at least, may confound the association. For example:

- being out of the labour force is not always due to poor health, but could also be due to chronic inability to find work resulting in foregoing the option of even self-identifying as 'unemployed and actively seeking work', a 'choice' not to be in the labour force, aptitude and so on. These non-health selection reasons for being out of the labour force would probably be independently associated with higher mortality risk, making labour force status a confounder of occupational class and education mortality gradients
- a subtle extension of the above is that exclusion from the labour force (be it health selection or not) is likely to be associated with socioeconomic factors *other* than just occupational or education, thus introducing confounding by other socioeconomic factors for any analysis that excludes the non-labour force participants.

Thus, the diminution of the occupational class and education mortality gradients observed by Martikainen and Valkonen and Kunst et al may not be solely due to differential health selection, but may be due in part at least to confounding by labour force status. Indeed, it is highly unlikely the gradients are entirely due to differential health selection. Moreover, it is plausible that the residual diminution during the fifth to tenth year of follow-up in Martikainen and Valkonens' study (Figure 21 above) (Martikainen and Valkonen 1999) may result only to a small extent from differential health selection, but may mostly result from confounding. (One way to determine the length of time over which health selection by labour force status operates is to plot the mortality risk over time for each stratum of education or occupational class among those in the labour force (eg, Fox et al 1985), and determine when the mortality risk lines become parallel (ie, stop diverging). When the mortality risk lines become parallel, differential health selection out of the labour force would no longer be affecting the mortality gradient.) However, whatever the relative contributions of differential health selection and confounding, the association of occupational class, education, and other socioeconomic factors with mortality restricted to the active labour force will underestimate the same association among all categories of the labour force combined.

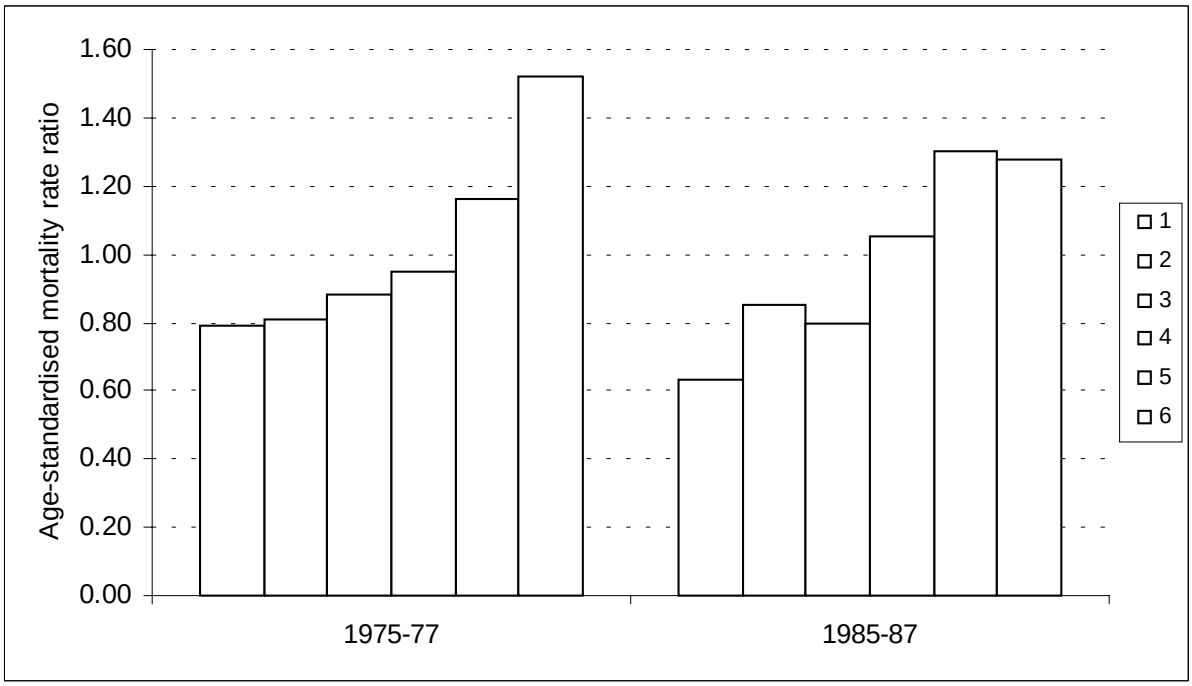
3.1 Findings

3.1.1 Occupational class mortality gradients in New Zealand

Pearce and colleagues have extensively analysed the association of occupational class with mortality among 15–64 year old males, for the two time periods 1974–78 and 1985–87. They used occupational data from the death registration forms for the numerator, and census occupational data for the denominator – an unlinked census–mortality study design. Analyses for females were not possible due to inadequate occupational data, and were also not possible for the retired as the New Zealand census elicits only current occupation. Analyses for 1995–97 have been submitted (personal communication, Neil Pearce, 2001).

Figure 22 shows the all-cause mortality rate ratios for the two time periods 1975–77 and 1985–87. As in other countries, there is a definite and, mostly, monotonic occupational class mortality gradient. The mortality among the lowest occupational class is approximately twice that in the highest occupational class in both periods – more precisely 1.9 times in 1975–77 and 2.0 times in 1985–87. Using a method for determining the linear trend in proportions (Pearce and Cryer 1986), the percentage change in mortality compared to the overall pooled mortality rate for each incremental decrease in occupational class was 13.4% in 1975–77 and 14.6% in 1985–87. Thus, assuming the relationship of mortality with occupational class is approximately linear, there was some suggestion the occupational mortality gradient had increased over a decade in relative terms. This increase over time is in keeping with findings internationally that gradients had tended to increase over time (Feldman et al 1989; Marmot and McDowall 1986; Pappas et al 1993). Note, though, that as the overall mortality decreased the absolute difference in age-standardised mortality rates across classes decreased. Beneath these summary estimates of the gradients, however, are different patterns between 1975–77 and 1985–87 – the gradient steepens with lower occupational class in 1975–77, whereas it is roughly linear, but not quite monotonic, in 1985–87. Some of these different patterns would have been due to changes in occupational coding on the unlinked census and mortality data over time (numerator–denominator bias). Also, for the 1985–87 period occupational class 6 had higher mortality than occupational class 5 for each age-group except 55–64 year olds – the overall 15–64 year old standardised rates were strongly influenced by the more numerous deaths in this oldest age group.

Figure 22: Age-standardised mortality ratios by Elley Irving occupational class, among 15–64 year old New Zealand males during 1975–77 and 1985–87



Source: Pearce et al 1991

The unlinked census–mortality study design used by Pearce and colleagues is prone to numerator–denominator bias. Findings from the OPCS Longitudinal Study in the UK suggested that the occupational mortality gradient is generally similar between unlinked and linked (with long follow-up) census–mortality studies (Fox and Goldblatt 1982; Fox et al 1985), thus arguing against substantial numerator–denominator bias. Prior to the NZCMS there has been no linked study against which to compare the results from Pearce et al's unlinked analyses – although the NZCMS is limited with regard to occupational class as economically inactive men are excluded and follow-up is for only three years. There are three caveats regarding possible numerator–denominator bias in the unlinked analyses by Pearce et al First, numerator–denominator bias arises because of inconsistency in the coding of occupation between census and mortality data. The nature of that inconsistency will vary between countries, and thus the magnitude of any numerator–denominator bias will vary between countries. For example, coding practises for occupation may vary both by country and census/mortality data. Kunst and Groenhof estimate that the numerator–denominator bias may be up to 20% for England and Wales and France, and that the bias could be in either direction (Kunst and Groenhof 1996). Second, within countries this coding inconsistency may vary over time, jeopardising time series comparisons. Between 1975–77 and 1985–87 in New Zealand, the coding of occupations had changed. Pearce et al (1991) put considerable effort into overcoming these changes, but there may still be some small bias that alters the interpretation of a small increase in occupational class gradients between 1975–77 and 1985–87. Third, while the presence of a gradient of mortality by occupational class is not an artefact of numerator–denominator bias, the actual pattern might be. For example, French linked census–mortality analyses find a monotonic gradient by occupational class, but unlinked analyses find the highest mortality for 'routine non-manual' (4th ranked class) and lower mortality for 'skilled manual' and 'unskilled manual' (5th and 6th ranked class) (Kunst et al 1998b). Kunst et al concluded that this apparent lack of a monotonic gradient in the unlinked analyses was due to numerator–denominator bias.

It is interesting to speculate whether the observed lower mortality in occupational class 6 compared to 5 among 55-64 year old men in 1985–87 reported by Pearce et al was also due to some residual numerator–denominator bias. A possible mechanism may be a tendency to 'promote the dead' when relatives report the deceased's usual occupation to the funeral director.

In addition to the broad picture of all-cause mortality at two points in time, the work of Pearce and colleagues has provided much information of occupational class mortality gradients by specific causes of disease, and the overlap of ethnicity and occupational class:

- The gradient was greater for diseases amenable to medical intervention than those non-amenable, with the ratio between occupational class 6 and 1 being about 3.5 for amenable deaths in both time periods (Marshall et al 1993).
- By cause of death, the strongest gradients in 1975–77 were for deaths from accidents, poisonings and violence, diseases of the respiratory system, endocrine, nutritional and metabolic diseases, diseases of the genitourinary system, and diseases of the of the digestive system (Pearce et al 1983b). Gradients were also present, but weaker, for cancers and coronary heart diseases.
- Occupational class mortality gradients existed for a range of cancers in 1984–87, particularly larynx, liver, buccal cavity/pharynx, oesophagus, lung and soft tissue sarcoma (Pearce and Bethwaite 1997). Inverse occupational gradients existed for rectal and colon cancers, malignant melanoma, and brain/nervous system cancers. Lung cancer accounted for 54.1% of the overall cancer gradient, and much of the overall cancer gradient was probably attributable to the differential distribution of smoking by occupational class. These patterns were consistent with previous analyses for 1974–78 cancer mortality (Pearce and Howard 1986), and results from other countries (Kogevinas et al 1997).
- Between 1975–77 and 1985–87 the gradient for all circulatory diseases increased (Kawachi et al 1991), mainly due to a doubling of the gradient for ischaemic heart disease. These results are consistent with the observed transition in other developed countries this century of ischaemic heart disease becoming increasingly concentrated among lower socioeconomic groups. Possible explanations offered by Kawachi et al for the increased IHD gradient, emphasising 'yet again the social and political nature of public health' (p.397), were: targeted marketing of high fat foods and tobacco to lower socioeconomic groups; differential uptake of health promotion and education material; differential uptake of health care; and psychosocial conditions associated with work.
- The relative excess of mortality for Maori compared to non-Maori decreased between 1975–77 and 1985–87, particularly for diseases amenable to medical treatment (Pearce et al 1993). In both time periods there were strong occupational class gradients among Maori and non-Maori.
- Depending on whether standardised rate ratios or standardised rate differences are used as the means of comparison, 20–30% and 30–40% of the Maori non-Maori disparity in mortality could be attributed to occupational class in 1975–77 and 1985–87, respectively (Pearce et al 1993). However, these percentages of the ethnic disparity attributable to socioeconomic position must be treated with considerable caution. First, occupational class only captures one element of socioeconomic position, and it is likely that much more of the ethnic disparity is attributable to socioeconomic position. Second, social standing in Maoridom is probably less aligned with occupational class than it is among non-Maori. Third, percentage

attribution may be sensitive to the standardisation process. Fourth, occupational class and ethnicity are not independent. Being Maori, via institutional and interpersonal racism and historical forces dating back to dispossession of land, taonga and other resources, is a determinant of socioeconomic position. Thus, socioeconomic position is a *pathway* between ethnicity and health status.

3.1.2 Cross-national comparison studies of occupational class mortality gradients

An often cited study comparing occupational class mortality gradients between countries is that by Vagero and Lundberg (1989), comparing deaths in the 1970s among males aged 20–64 in Britain and Sweden using the British Registrar General's classification (Vagero and Lundberg 1989). This study found that the mortality gradient was less in Sweden than in Britain, supporting the notion that a more egalitarian society has smaller health inequalities. However, with the benefit of hindsight, we now know that the study was flawed: the British data was derived from unlinked census–mortality data, while the Swedish data was derived from linked census–mortality data for *economically active* men. As described above, recent work has shown that the (crude) occupational class mortality gradient is underestimated among economically active men due to differential health selection and/or confounding by labour force status.

The most up-to-date and comprehensive cross-national comparison studies on occupational class mortality gradients are those by Kunst, Mackenbach and colleagues in the European Union Working Group on Socioeconomic Inequalities in Health (Kunst 1997; Kunst et al 1998a; Kunst et al 1999; Kunst et al 1998b; Kunst et al 1998c; Kunst et al 1998d; Mackenbach et al 1997). These studies draw upon data from both unlinked and linked census–mortality studies for deaths during the 1980s. Only males could be included due to inadequate data for females (Kunst 1997). As many countries as possible were included. Eleven European countries were included in all studies (England/Wales, Ireland, Finland, Sweden, Norway, Denmark, France, Switzerland, Italy (Turin), Spain and Portugal), and data from the NLMS in the United States was included in one study of ischaemic heart disease (Kunst et al 1999).

Perhaps the greatest contribution from recent analyses by Kunst, Mackenbach and colleagues is their meticulous attention to overcoming potential biases in previous cross-national comparison studies (Kunst et al 1998b). First, while unable to determine the exact direction of the bias, numerator–denominator bias was estimated to bias the observed manual non-manual mortality gradient by up to 20% in the four countries with only unlinked census–mortality data (Ireland, Spain, Portugal, Switzerland) (Kunst and Groenhof 1996; Kunst et al 1998b; Kunst et al 1998c).

Second, health selection bias arising from excluding the economically inactive when only data on current occupation was available was estimated, and adjusted for by a correction formula (Kunst 1997; Kunst et al 1998c). The correction formula assumes that the ratio of mortality for the economically inactive to the economically active is constant across occupational classes, an assumption found to be reasonable on the basis of sensitivity analyses. One external piece of information for each country and time period is required for the correction formula – the proportion of people assigned to each occupational class on the basis of usual occupation that were economically inactive at the time of the relevant census. This external data was obtained from other surveys within each country. (The Household Labour Force Survey in New Zealand could be used for this purpose – but was beyond the scope of this report.)

Third, bias that may arise from differing coding systems for occupations and occupational class was minimised by using where possible the original 'EGP' (Erikson, Goldthorpe and Portocarero) occupational class scheme for Sweden, England and Wales, and France (Erikson and Goldthorpe 1992). For Finland, Norway and Switzerland, the 'GLT' conversion algorithm for the EGP scheme was used (Ganzeboom et al 1989). However, for the remaining countries data could only be made comparable in three broad classes: non-manual and manual classes, and farmers and farm labourers (Kunst et al 1998c). Thus analyses were limited to these three broad classifications (usually just manual compared to non-manual). Sensitivity analyses for Sweden, England and Wales, and France suggested that the ranking of countries by occupational class mortality gradients did not differ between using the seven-level class scheme versus a manual non-manual dichotomy (Kunst et al 1998c). Considering the three above sources of bias together (numerator–denominator, economically inactive, occupational class coding), Kunst et al estimated that the possible residual systematic error in their results after adjustments was negligible for Sweden, England and Wales, and France, but possibly still large for Ireland, Spain and Portugal – other countries were intermediary. Taking Spain and 45–59 year old males as the most extreme example, if all three biases were the maximum of the estimated range of error, *and* all three biases acted in the same direction, then the manual non-manual rate ratio would have been out by 38% (Kunst et al 1998c). However, it is unlikely that all three biases would be maximal and acting in the same direction, so the 38% estimate is a (somewhat implausible) worst case scenario.

The general finding by Kunst, Mackenbach and colleagues for all-cause mortality in the 1980s was that the manual to non-manual mortality ratio was *little different across countries*, except in France where the inequalities were larger (Kunst et al 1998c; Kunst et al 1998d). This general finding was true for males aged 45–59 years, and 60–64 years. For example, among males aged 45–59 years the manual to non-manual rate ratio (with adjustment for the exclusion of men with unknown occupation) was highest in France (1.71), somewhat higher in Finland (1.53), and then between 1.33 and 1.44 for the remaining nine European countries (see Table 49). If the country specific rate ratios had not been adjusted for the exclusion of economically inactive males, the rate ratio would have appeared smaller in Sweden compared to England and Wales (in keeping with Vagero and Lundberg (1989)), and low by international standards for Italy and Spain. Thus, adjusting for the exclusion of economically inactive people made a substantial difference to cross-national comparisons. Within the likely bounds of the methods used to adjust the data, and excluding Finland and France, the countries were indistinguishable. Kunst et al concluded:

> '*This study underlines the similarities rather than the dissimilarities between European countries. There is no evidence that mortality differences are smaller in countries with more egalitarian socioeconomic and other policies.*' (Kunst et al 1998c, p.1459).

Table 49: Mortality rate ratio (95% confidence interval) comparing manual classes to non-manual classes for major groupings of cause of death among men aged 45–59, using 1980s data

Country	All-cause		Neoplasms		Cardiovascular	
Finland	1.53	(1.49–1.56)	1.39	(1.32–1.47)	1.48	(1.42–1.53)
Sweden	1.41	(1.38–1.44)	1.18	(1.13–1.23)	1.36	(1.31–1.40)
Norway	1.34	(1.30–1.39)	1.25	(1.18–1.33)	1.34	(1.27–1.40)
Denmark	1.33	(1.30–1.36)	1.21	(1.16–1.26)	1.28	(1.23–1.33)
England/Wales	1.44	(1.33–1.56)	1.21	(1.05–1.39)	1.52	(1.36–1.71)
Ireland	1.38	(1.30–1.46)	1.39	(1.24–1.55)	1.27	(1.17–1.38)
France	1.71	(1.66–1.77)	1.71	(1.61–1.82)	1.35	(1.26–1.45)
Switzerland	1.35	(1.29–1.39)	1.44	(1.35–1.54)	1.08	(1.01–1.15)
Italy	1.35	(1.28–1.42)	1.43	(1.31–1.55)	1.17	(1.07–1.28)
Spain	1.37	(1.34–1.39)	1.33	(1.29–1.38)	1.19	(1.15–1.22)
Portugal	1.36	(1.31–1.40)	1.12	(1.05–1.21)	1.03	(0.97–1.10)

Source: Kunst et al 1998d

Kunst and colleagues also report a range of subsidiary findings on occupational class gradients for *all-cause mortality* (Kunst et al 1998c):

- There was more between-country variation for rate ratios among 30–44 year old males. Finland, Sweden and Norway had the highest manual non-manual rate ratios (1.76, 1.66, and 1.65 respectively), Italy the lowest (1.35), and the other countries (Denmark, England and Wales, Ireland, Switzerland, Portugal – no data on 30–44 year olds for France and Spain) all had rate ratios of about 1.50.
- Among males aged 60–64 years, France again had the highest mortality rate ratio (1.50) compared to the other countries with 60–64 year old data (Finland, Norway, Denmark, England/Wales; rate ratios 1.21 to 1.33). Without adjustment for the exclusion of economically inactive males, the rate ratios for all three Nordic countries would have appeared much smaller than that in England and Wales.
- The above pattern across countries was robust to changing the measure from rate ratios to the index of dissimilarity, a measure of the percentage of deaths that needs to be redistributed between occupational classes to remove inequalities. (Both the index of dissimilarity and rate ratios are *relative* measure of inequalities.)
- It is an epidemiological truism that if rate ratios are constant across strata (in this case across countries) and the absolute rate among the unexposed (in this case the non-manual class) varies across strata, then the rate differences *must* vary across strata (Rothman and Greenland 1998). Put another way, if the background mortality rates vary between countries, then the ranking of countries will depend on whether one uses relative (eg, rate ratio) or absolute (eg, rate difference) measures of effect. Accordingly, a different pattern emerges when the difference in the probability of dying between the ages of 45 and 64 years between manual and non-manual classes was used (Kunst et al 1998c). France remained the most unequal country (11.5% difference in the probability of dying), but Finland and Ireland (9.8% and 8.1% respectively) also stood out as particularly unequal – a consequence of the higher overall mortality rates in these two countries. Switzerland, Italy, Spain, Portugal, Sweden, Norway, and Denmark had

comparably low inequalities (5.0% to 6.3%), and England and Wales was intermediary (7.5%).

- The cross-national comparisons of occupational class mortality gradients was broadly consistent with those for educational mortality gradients (page ? of this report) (Mackenbach et al 1999).

As already alluded to, the work by Kunst, Mackenbach and colleagues challenges previous studies that suggested occupational mortality gradients were smaller in more egalitarian countries (Kunst and Mackenbach 1994a; Leclerc et al 1990; Vagero and Lundberg 1989). Referring to the upstream determinants of health in the model by Turrell et al (1999) in Figure 2, previous studies supported the notion that more egalitarian government policies flattened the socioeconomic mortality gradients for the determinants of health, including occupational class. An important possible distinction between the previous studies and the most recent work by Kunst, Mackenbach and colleagues was that the former studies tended to use mortality data from the 1970s, whereas the latter studies by Kunst et al used mortality data from the 1980s.

Table 50 is taken from Kunst et al (Kunst et al 1998c), and demonstrates that the discrepancy is not a period effect, but due to the earlier studies not allowing for exclusion of the economically inactive in (predominantly) the Scandinavian countries. Using data from a previous cross-national comparison by Kunst and Mackenbach of occupational class mortality gradients in the 1970s (Kunst and Mackenbach 1994a), the observed manual non-manual mortality gradients for 45–69 year old males are shown in column 1. In column 2, the age range is restricted to 45–59 years. As many 60–69 year olds are economically inactive, this simple age restriction is a partial adjustment – accordingly the rate ratios increased most notably for Norway, Denmark and Sweden. In the third column, the 1970s rate ratios are adjusted using the correction formula developed by Kunst et al for the 1980s analyses (Kunst 1997; Kunst et al 1998c). The comparable adjusted results for the 1980s are shown in the fourth column.

Two findings stand out. First, inequalities increased in all countries between the 1970s and 1980s. Second, the relative ranking of countries was similar in the 1970s and 1980s: Norway, Denmark, Sweden, and England and Wales had roughly equivalent inequalities (although England and Wales was toward the high-end of the range in each period); France clearly had the greatest inequality in each time period; and Finland was intermediary in each time period.

Table 50: Male mortality rate ratios for manual compared to non-manual classes for linked census–mortality studies in the 1970s and 1980s, unadjusted and adjusted for exclusion of economically active

	Observed		Adjusted (for economically inactive)	
	1970s, 45–69 years	1970s, 45–59 years	1970s, 45–59 years	1980s, 45–59 years
Norway	1.05	1.12	1.18	1.34
Denmark	1.07	1.18	1.22	1.33
Sweden	1.09	1.18	1.26	1.41
England/Wales	1.20	1.21	1.25	1.44
Finland	1.35	1.39	1.40	1.53
France	1.39	1.44	1.61	1.71

Source: Kunst et al 1998c

It is possible to argue that any beneficial effect of egalitarian policies on health affects both high and low socioeconomic groups evenly. This even benefit would reduce the overall national mortality rate, reduce the *absolute* difference in mortality between high and low occupational classes, but not reduce the *relative* difference in mortality between occupational classes. The lower absolute difference in mortality in Sweden compared to England and Wales offers some support to this argument (Vagero and Erikson 1997), particularly from the perspective that the actual number of people affected is more important for public health policy that relative risks. However, it is still surprising that if egalitarian policies are good for health that they do not advantage lower socioeconomic groups more than higher socioeconomic groups.

The similarity of the occupational class gradients across countries for all-cause mortality re-emphasises the fundamental nature of socioeconomic determinants of health (Link and Phelan 1995) regardless of time or place, mortality is strongly patterned by the socioeconomic determinants of health. However, while the *all-cause* gradient appeared similar across the countries studied by Kunst, Mackenbach and colleagues, the occupational class gradients for *specific* causes of death did vary across countries (see Table 49) (Kunst et al 1998a; Kunst et al 1999; Kunst et al 1998d). The southern European countries tended to have higher inequalities in cancer mortality; the northern countries tended to have higher inequalities in cardiovascular mortality. Thus, despite different diseases (and presumably different health-related behaviours and psychosocial mechanisms) giving rise to socioeconomic mortality gradients within each country, the *overall* all-cause mortality gradient was similar across countries. This finding argues against a link between general socioeconomic 'stress' and susceptibility to disease (Leon 1998) if general stress and susceptibility were the causal mechanism, one would expect the pattern of cause-specific mortality to also be similar across countries. Instead, the evidence of Kunst, Mackenbach and colleagues argues that midstream factors (eg, smoking, diet) vary between countries (Leon 1998). For example, the socioeconomic (by education) inequalities in smoking and vegetable consumption across countries has been found to approximate that expected on the basis of occupational class inequalities in ischaemic heart disease mortality (Cavelaars et al 2000; Cavelaars et al 1997).

3.1.3 Occupational class mortality gradients from a life-course perspective

The majority of research on the independent effect of socioeconomic position at different points of the life-course on adult mortality has compared occupational class measures (eg, father's occupational class as a measure of childhood socioeconomic position, occupational class at first job, and most recent occupational class). The focus on occupation is probably a consequence of the best available data over the life-course being occupation.

A prospective observational study of nearly 6000 men aged 35–64 years at recruitment in the West of Scotland with 21 years of follow-up has provided important insights (Blane et al 1996; Davey Smith et al 1997; Davey Smith et al 1998a; Davey Smith et al 1998b). This study will be used as one possible example of research on occupational class mortality gradients measured at different points in the life-course. In this study, occupational class (manual, non-manual) was determined for each subject's father (ie, childhood occupational class), occupation at entry to labour market, and occupation at recruitment to the study. With these three measures of occupational class, it was possible to determine the association of cumulative occupational class (eg, manual class at all three times versus non-manual at all three times) on mortality, *and* whether occupational class at one point in the life-course (eg, childhood versus adulthood) was more strongly associated with (cause-specific) mortality. Findings from this study include:

- compared to men assigned as non-manual class at all three stages of the life-course, the relative risk of age-adjusted mortality for men assigned as: manual on one occasion was 1.29 (95% confidence interval 1.08 to 1.56); manual on two occasions was 1.45 (1.21 to 1.73); and manual on all three occasions was 1.71 (1.46 to 2.01). The strength of this association was strongest for cardiovascular diseases (relative risk 1.94 for manual on all three occasions), weakest for cancer (1.44), and intermediary for non-cardiovascular non-cancer deaths (1.75) (Davey Smith et al 1997)
- interestingly, occupational class in childhood was more strongly associated with cardiovascular disease mortality than occupational class derived from the current job (relative risks 1.58 versus 1.38), and conversely, occupational class for the respondents current job was more strongly associated with cancer mortality than childhood occupational class (relative risks 1.35 versus 1.26) (Davey Smith et al 1997). These findings fit with the hypothesis of foetal and early childhood origins of cardiovascular disease (Barker and Martyn 1992; Kuh and Ben-Shlomo 1997)
- at a finer level of cause-specific mortality, the association of childhood occupational class with stroke and stomach cancer, was unaltered after controlling for adult occupational class, attenuated for coronary heart disease and respiratory disease, and essentially eliminated for lung cancer (Davey Smith et al 1998a).

The findings of this West of Scotland study strongly suggest the importance of cumulative socioeconomic position over the life-course, and the possible variation in the critical exposure period by cause of death. An alternative explanation, however, is that measuring occupational class at three points of the life-course is akin to reducing measurement error of socioeconomic position in, say, adulthood. For example, among manual class males there would still have been variation on other socioeconomic factors. It may be that those manual men with the lowest incomes and least educational qualifications were also more likely to have had a father of manual occupational class. Thus, using multiple measures of occupational class over the life-course may be merely reducing variation during adulthood on other socioeconomic factors. While plausible as a source of error in the above (and other) studies on the life-course, it does not explain why different causes of death had varying associations with occupational class over the life-course. Also, it may be that within class variation on other socioeconomic factors is not just spuriously associated with occupational class earlier in the life-course, but that earlier occupational class causes this variation.

3.1.4 The effect of occupational class on mortality independent of other socioeconomic factors

It seems plausible to determine the independent effects of, say, income and education on mortality risk. One can conceptualise that education and income might exert different forces on mortality risk. For example, education may affect mortality risk by knowledge of healthy lifestyles, whereas income may affect mortality risk via material influences. However, it is more problematic to determine the independent effect of occupational class from, say, income and education. For example, many indices of occupational class capture processes that mediate the transition from education to income or wealth. The New Zealand Socioeconomic Index (NZSEI) is of this type, being calculated from educational and income data (Davis et al 1997; Davis et al 1999b). Thus, determining the independent effects of occupational class and other socioeconomic factors may not be particularly illuminating. Correspondingly, many authors compare and contrast socioeconomic factors other than occupational class, but few (eg, Davey Smith et al 1998b; Sorlie et al 1995) include occupational class in analyses simultaneously with other socioeconomic factors.

Sorlie et al (1995) assessed the association of occupational category on mortality in the US NLMS, controlling for just age and race, and then controlling additionally for employment status, income, education, marital status and household size (Sorlie et al 1995). The occupational categories were not strictly occupational classes, but rather groupings of professional, technical, service, farming, production, and operator occupations. Moderately sized associations were seen just controlling for age and race, but they were largely reduced to the null when controlling for the remaining socioeconomic factors. Conversely, the effect sizes for the other socioeconomic factors were little reduced following inclusion of the occupational categories.

Davey Smith et al (1998) compared the association of education and occupational class with mortality in a prospective observational study of 5749 men in workplaces in the west of Scotland (Davey Smith et al 1998b). Education was categorised by age at leaving full-time education: 12–14, 15–16, 17–18 and 19+ years. Occupational class was categorised according to the British Registrar-General's classification. Cause of death was broadly assigned to three groups: cardiovascular, cancer, and non-cardiovascular non-cancer. Mortality was strongly associated with each socioeconomic factor, with cardiovascular disease being the most strongly associated cause of death with education and non-cancer non-cardiovascular disease the most strongly associated cause of death with occupational class. Within strata of education, a strong association of occupational class with mortality persisted. However, within strata of occupational class the association of education with mortality was largely reduced to the null. The authors concluded that occupational class is a better discriminator of socioeconomic differences in mortality than education. Furthermore, they concluded that their findings supported a material explanation of socioeconomic mortality gradients (ie, as measured by occupational class) as opposed to a cultural explanation (ie, as measured by education). However, as reviewed above, any material-cultural distinction between education and occupational class as measures of socioeconomic position is not that clear-cut. Also, if education was measured with more error than occupational class, one would expect occupational class to dominate. Indeed, the measure of education was simply the age at leaving formal education, not the qualifications obtained or quality of education received. Perhaps a more reasonable interpretation (also made by Davey Smith et al) is that the stronger association of education with cardiovascular disease may reflect the importance of socioeconomic position in childhood as a determinant. However, one could still argue that rather than education and occupational class discriminating between stages of the life-course, it may that education is more strongly associated with cardiovascular disease due to health-related knowledge and behaviours.

3.1.5 Workplace pathways from occupational class to mortality

In addition to the underlying principle of socioeconomic stratification by culture, education, and income embedded in occupational class, occupation also captures direct exposure to physical occupational hazards. The contribution of these physical hazards to the observed occupational class mortality gradients has been considered small relative to the socioeconomic dimension (Fox and Adelstein 1978). However, Schrijvers et al (1998) have found that much of gradient of self-rated health (very good or good versus less than good) by occupational class was explained by physical working conditions (Schrijvers et al 1998). They conducted cross-sectional analyses on the baseline data from the Dutch Longitudinal Study on Socioeconomic Health Differences for 6932 employed respondents. Adjusting only for demographic factors, there was an approximately threefold difference in the odds of less than good self-rated health between the lowest and highest occupational classes. Additionally, controlling for physical working conditions reduced this association by approximately 50% for males, and 20% to 35% for females. Thus, for self-rated health in a cross-sectional analysis, physical working conditions explained much of the gradient. Whether this finding will also apply to longitudinal analyses of mortality will have to wait for enough follow-up data to become available.

In addition to physical hazards encountered in the workplace that may mediate the association of occupational class with health, there are also psychosocial hazards. Two models of job stress dominate international research. First, the job strain model posits that independent and synergistic effects of decreasing control and increasing demands cause poorer health (Karasek and Theorell 1989; Schnall and Landsbergis 1994). There is much evidence to support this model of job stress, with many studies reporting significant associations with health (eg, Bobak et al 1998; Bosma et al 1997; Cheng et al 2000; Everson et al 1997; Hemingway and Marmot 1999) particularly for coronary heart disease and the job control dimension, such that less job control is associated with higher rates of coronary heart disease (Hemingway and Marmot 1999). Secondly, an effort-reward model has been developed by Siegrist (Siegrist 1996; Siegrist et al 1990). Components of both models have been shown to predict coronary heart disease among males and females in the Whitehall Study (Bosma et al 1998).

Given that it seems plausible that psychosocial factors may mediate some of the association of socioeconomic position with health (see Figure 2 this report), an obvious question is 'how much does job stress mediate the association of socioeconomic position, in particular occupational class, with health?'. Returning to the cross-sectional analyses of Schrijvers et al cited above in relation to the mediating role of physical hazards between occupational class and health, they also considered the mediating role of job control and demand (Schrijvers et al 1998). Job demands were not important, but adjusting for job control reduced the association of occupational class with less than good self-rated health by approximately 40% for males and by approximately 20% to 30% for females. Together, physical hazards in the work place and job control explained perhaps three-quarters of the association for males, and approximately 40% for females.

The study by Schrijvers et al is limited to a cross-sectional analysis and a self-rated health outcome. What is the possible mediating role of job stress in the association of occupational class with mortality? First, and again from the Erasmus University research group, Bosma et al (1999) assessed the mediating role of perceived control for the association of three socioeconomic factors (occupational level, education, and income) with mortality in the Longitudinal Study of Socioeconomic Health Determinants (Bosma et al 1999). Adjusting for perceived job control, the association of each socioeconomic factor with mortality was reduced by approximately 50%, suggesting that half of the association of socioeconomic position with health may be mediated by psychosocial perceptions of control. Second, Marmot et al (1997) have demonstrated that adjusting for job control and effort-reward imbalance reduces the association of occupational grade with coronary heart disease incidence by approximately half (Marmot et al 1997a). Controlling additionally for traditional risk factors for coronary heart disease, height, and social support essentially eliminates the association of occupational grade with coronary heart disease. Based on previous results from the Whitehall Study that only 25% of the occupational grade gradient for coronary heart disease was explained by traditional risk factors (Marmot et al 1984), it is tantalising to speculate that job stress explains much of the residual association. However, this possibility has been strongly criticised for mistakenly conflating socioeconomic position with job stress. For example, Davey Smith and Harding argue that 'low control over work is virtually synonymous with low socioeconomic position ...', and that job control is a sensitive indicator of socioeconomic position that cannot be changed independently of socioeconomic position (Davey Smith and Harding 1997). In other words, part of the socioeconomic hierarchy is a hierarchy of control. Likewise, Lynch and Kaplan (2000) argue:

> *'For both methodological and conceptual reasons we do not believe that there is much to be gained from statistically partitioning the separate contributions of socioeconomic position and psychosocial working conditions (Davey Smith and Harding 1997; Lynch et al 1997c; Lynch et al 1997d; Marmot et al 1997a). In reality they are intimately related in complex ways that may be trivialised by the crude statistical adjustment of one for the other. 'Explaining' social-level phenomena such as socioeconomic gradients in health cannot be reduced to 'explaining away' these gradients by statistical adjustment for workplace demands and control at the individual-level – the demand-control attributes assigned to individuals are in part a result of the social-level phenomena being explained. (Macintyre 1997).'* (Lynch and Kaplan 2000, p.24).

Marmot et al anticipated these criticisms in their 1997 paper, and made the following observations in their discussion (Marmot et al 1997a) and subsequent correspondence (Marmot et al 1997b):

- While occupational grade and job control were highly correlated in the Whitehall Study, they were not inseparable.
- The association between job stress and coronary heart disease is not simply due to confounding/overlap with socioeconomic position. Two studies that have failed to find an association of job stress with CHD still had a strong association of socioeconomic position with CHD (Hlatky et al 1996; Suadicani et al 1993) arguing against job stress and socioeconomic position necessarily measuring one and the same thing.
- There seems to be some specificity of the association of low job control with CHD, while socioeconomic position is associated with many diseases.
- The results were not isolated to the Whitehall Study. A study (unpublished in 1997) of Czech men also found that a combination of low control and coronary risk factors explained the higher risk of myocardial infarction among less educated men (Bobak et al 1998).
- While the findings of the Whitehall Study were specific to working men, the putative pathophysiological mechanisms mediating the effect of psychosocial stress (eg, autonomic nervous system, hypothalamic-pituitary-axis and allostatic load (Brunner 1997; Brunner and Marmot 1999; McEwen 1998; McEwen and Seeman 1999; Sapolsky et al 1997; Shivley and Clarkson 1994)) are applicable to psychosocial stress in the general population.

The debate regarding the appropriateness, or not, of adjusting occupational class gradients for job stress will undoubtedly continue. Before leaving this debate, it is worth noting that the analyses behind this debate (controlling occupational class for job stress) are subject to the sources of error described in Chapter 5, including the possible unreliability of controlling for intermediate variables (Poole and Kaufman 2000; Robins and Greenland 1992). Perhaps of most concern would be mismeasurement of job stress, or more specifically job control and demand, for two reasons. First, when combined with (inevitable) mismeasurement of the exposure occupational class (or other socioeconomic factor), the observed correlation between the two will likely be diminished. Thus, their effects may appear separable when, if each was perfectly measured, they would in reality be inseparable. Second, and assuming it is plausible to adjust socioeconomic position for job stress, mismeasurement of job stress will probably underestimate the percentage contribution of job stress to the observed socioeconomic gradient.

3.2 Conclusion

Occupational class has been the main measure of socioeconomic position in social epidemiology. The majority of occupational class measures used in research capture socioeconomic ranking akin to a Weberian framework of social class – few measures capture a Marxist framework of social class. Occupational class mortality gradients reported by Pearce and colleagues constitute the major body of evidence on socioeconomic mortality gradients in New Zealand.

Limitations of using occupational class include the difficulty experienced in assigning an occupational class to people outside of the workforce. Indeed, the association of occupational class with mortality when only current occupation is available at baseline may lead to substantial underestimates of the *usual* occupational class mortality gradient.

Occupational class, or more strictly, a manual versus non-manual classification, has been used extensively in cross national comparison studies. Recent evidence suggests that the all-cause mortality gradients are very similar across countries. However, for cause-specific mortality countries have notably different occupational class inequalities.

Occupational class measured at different stages of the life-course has been shown to be independently associated with adult mortality. Interestingly, the independent contribution of occupational class at different stages of the life-course appears to vary by cause of death.

Psychosocial factors (ie, work control and demand) have recently been proposed, and strongly debated, as possible pathways between occupational class and mortality.

4 Income

Income is considered to influence health outcomes by facilitating access to medical care, good housing, good diet, good working conditions, less exposure to noxious environment, and social amenities (Liberatos et al 1988). This materialistic explanation for the effect of income on health has been challenged. The most notable objection to the materialistic effect of income is that in developed countries the majority of people live above some arbitrary poverty line that secures access to the basic necessities of life, yet strong health gradients by income persist. Another objection to the materialistic explanation is that between countries there is a strong association of life expectancy with GDP up to $5000 to $10,000 ppp per capita, but thereafter the relationship is less strong (Wilkinson 1992; Wilkinson 1996). Wilkinson argues that income within countries may be thought of as a marker of relative standing on a socioeconomic hierarchy, up and down which psychosocial comparisons occur (eg, feelings of subordination or dominance, self-efficacy) (Wilkinson 1996; Wilkinson 1999). In addition to the relatively flat association of GDP with health in developed countries, support for this psychosocial explanation is taken from animal (experimental) and human studies that demonstrate pathophysiological effects of psychosocial processes (Brunner and Marmot 1999; McEwen 1998; Sapolsky et al 1997; Shivley and Clarkson 1994). The psychosocial 'exposures' may include changes to social ranking (including experimentally in primate studies), bereavement, amount of social support, and threatened loss of employment (Wilkinson 1999).

However, not all researchers accept such a strictly psychosocial explanation, rather preferring a 'neo-materialistic' explanation for the association of income, and socioeconomic position more generally, with health (Blane et al 1997; Davey Smith et al 1994; Lynch and Kaplan 2000):

> '*The material basis of these [current 20th century socioeconomic inequalities in health] has changed. We need to consider the neo-material conditions that might be relevant to understanding socioeconomic health differences within the context of the historical overall improvement in health. For instance, adequate nutrition in terms of calories is not the same as having a balanced, low-fat diet, rich in fresh fruit, grains and vegetables. Adequate housing is not the same as housing that can protect people from the extremes of heat and cold and overcrowding. Even if the most basic material conditions are satisfied through a low but adequate level of income, each step up the income ladder may bring added neo-material benefits that can produce gains in health .... Furthermore, better neo-material conditions may have immediate and cumulative benefits over the life-course and may also influence the socioeconomic position and health status of future generations. Children who have access to a home computer may be improving the likelihood of later educational success and so influence their subsequent socioeconomic position and health.*' (Lynch and Kaplan 2000, p.25).

Note that the debate between psychosocial and neo-materialistic pathway options is not exclusive to the socioeconomic factor 'income'. For example, they could equally well apply to occupational class. However, the debate seems to have come to the fore with regard to income, probably because of the three main socioeconomic factors (education, occupational class, and income) income is the socioeconomic factor most closely associated with a materialistic explanation. Further, neither are the materialistic and psychosocial explanations mutually exclusive, nor do they rule out health-related behaviours as an explanatory pathway (see Figure 2). Rather, psychosocial and behavioural factors are strongly patterned by material factors (Emmons 2000; Lynch et al 1997a). Given that material, behavioural, and psychosocial factors are probably highly intertwined, why bother about the distinction? One critical reason is the policy implications (Davey Smith et al 1994; Lynch and Kaplan 2000). For example, focusing on the psychosocial correlates implies that the distribution of structural and material factors in society are taken as given, and we merely seek to modify peoples' perceptions and response to that (inequitable) distribution. The advantage of this approach, using job control as one possible example of a psychosocial factor, is summarised by Marmot et al:

> '*Job control is important precisely because it may be amenable to change. Where societies are reluctant to countenance shifts in income distribution, they might consider the re-design of jobs – particularly so if increasing job control is associated with economic as well as health benefits.*' (Marmot et al 1997b)

The disadvantage of focusing on psychosocial (and behavioural) factors without considering change in the underlying material distribution of resources is that it may be unreasonable, and even condescending, to encourage people to change their behaviour and perceptions of the world when resources are not available to facilitate change or there are other more pressing concerns. Graham's (1994) research on smoking among women is illuminating in this regard (Graham 1994). She concluded that low-income women use smoking as a means of coping, and that smoking was an appropriate response to the characteristics of the social environment. Compared to not smoking, the women perceived the advantages of smoking as: reducing stress, providing social

connections, causing disease in the long run (in contrast to the immediate concerns of material deprivation), and being relatively cheap.

In addition to these possible causal pathways from income to mortality, confounding, health selection and mismeasurement may cause an incorrect association between income and mortality to be observed. Health selection has been ruled out as a large component of observed income mortality gradients by studies that measure income well before the mortality follow-up period (Bucher and Raglan 1995; Daly et al in press; McDonough et al 1997). However, without doubt some of the strong associations of income with health in cross-sectional studies, or with mortality soon after measurement of income, will be due to health selection (Backlund et al 1996; Stronks et al 1997).

Confounding is a possible source of error in the observed associations of income and mortality. It may be that income is the outcome of the prior life-course of socioeconomic and other exposures, and (if possible) alteration of income *only* would have no effect on mortality risk (Blakely and Woodward 2000a). Such a counterfactual requirement for assigning causation in social epidemiology is eloquently argued by Kaufman and Cooper (Kaufman and Cooper 1999), but not without critical comment from other social epidemiologists (Krieger and Davey Smith 2000; Muntaner 1999). The critical issue here is 'what is an independent effect of income?'. Is it the effect of a change in income without changing all other possible antecedent or correlated variables? Or is it the effect of a change in income *and* some associated set of socioeconomic factors that immediately determine income (eg, labour force status)? Perhaps one way of resolving this dilemma is to answer according to the possible real-life policy implications. For example, if we are interested in the effect of income redistribution on mortality, but changing nothing else, then we are interested in the income effect independent of all prior socioeconomic factors. However, if we are interested in the effect of changing incomes via employment policies, then we must consider the joint effect of labour force status and income together.

Two final comments on income are warranted before reviewing empirical findings. First, income is usually measured over a reasonably short period of time (eg, one year), and therefore reflects short-term socioeconomic position. Incomes change markedly over the life-course – asset wealth may be a more durable measure of material socioeconomic position (Lynch and Kaplan 2000). The issue is discussed both in the following literature review and under the heading 'Asset wealth: car access and housing tenure' on page 186. Second, income may be a poor measure of socioeconomic position among the retired. This issue will also be addressed.

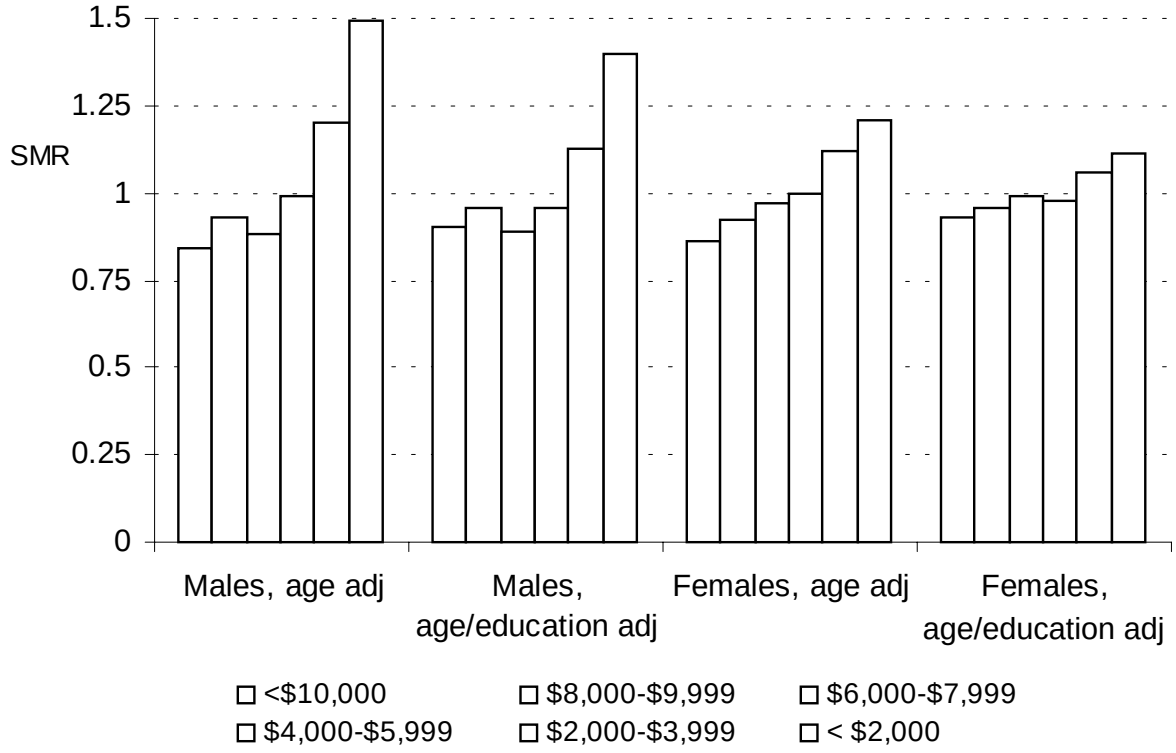
4.1 Findings

Income data is not collected by the census in England and Wales – thus income mortality gradients are not available for the UK OPCS Longitudinal Study. Cross-national comparisons of income mortality gradients are difficult due to varying purchasing power, different taxation regimes, various practices with regard to equivalisation for household size, and varying income data in each country. Thus there are no large cross-national comparison studies to the best of my knowledge – although there is one large cross-national comparison study for the association of income with *self-rated health* (van Doorslaer et al 1997). In this review, focus is on US studies to outline the general association of income with mortality, and both US and European studies to describe some possible mechanisms.

4.1.1 The general association of income with mortality

Hauser and Kitagawa (1973) reported a 80% higher mortality rate among 25–64 year old white men in 1960 with a family income of less than $2000 compared to $10,000 or more, using data from the 1960 Matched Records Study. The comparable difference for females was 40%. The authors acknowledge that with the very short follow-up period (four months) some of the observed association would be due to health selection. Figure 23 shows the SMR results standardisation by education. Perhaps 50% of the observed income mortality gradient was attributable to confounding by education.

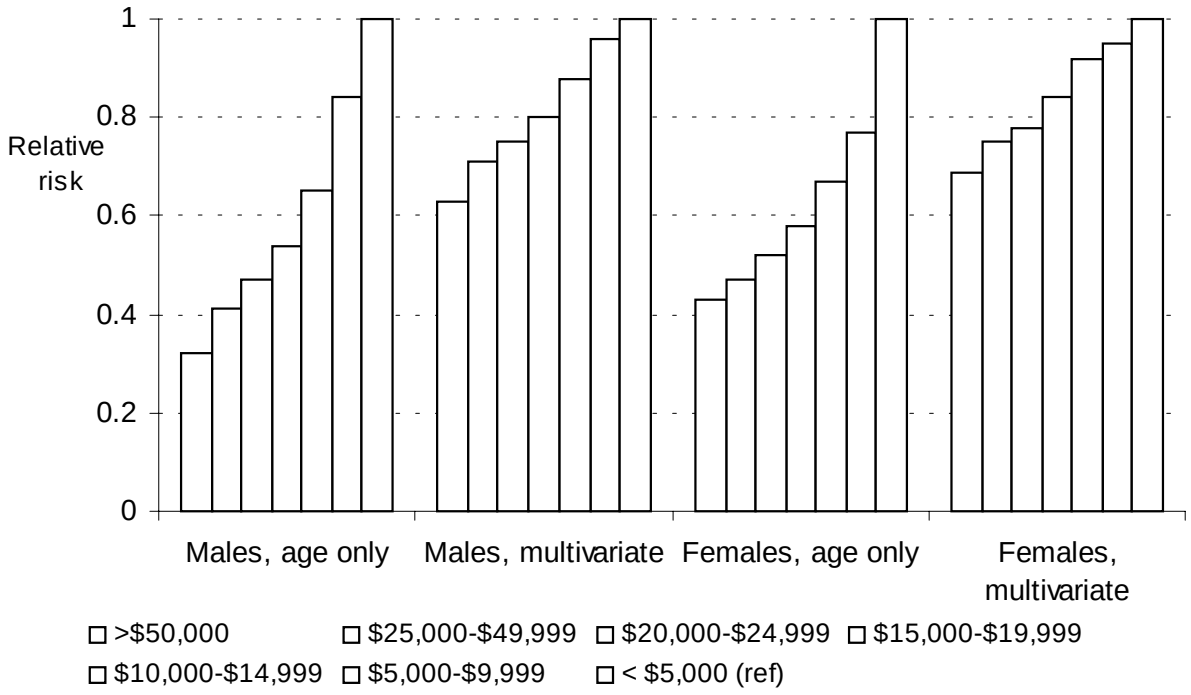
Figure 23: Standardised mortality ratios (SMRs) for 25–64 year old whites by family income in the 1960 Matched Records Study, standardised by age only and age and education



Source: Kitagawa and Hauser 1973

Following on from the 1960 Matched Records Study, the US National Longitudinal Mortality Study (NLMS) analysed the association of income with mortality during the 1980s (Sorlie et al 1995). All individuals were followed up for mortality for at least 2.5 years, and some up to 7.5 years. Adjusting for age only, strong associations of family income with mortality were found for males and females, and 25–44, 45–64, and 65 plus year olds. Figure 24 shows the association for 45–64 year males and females adjusting for age only, and also adjusting for age, race, labour force status, education, marital status, and household size. This latter multivariate adjustment may be considered as controlling for confounders, indicating a strong independent association of income with mortality. (It may also be an over-adjustment if controlling for labour force status introduced bias from differential health selection.)

Figure 24: Relative risk of death among 45–64 year olds by family income in the US National Longitudinal Mortality Study, adjusted for age only and age, race, labour force status, education, marital status, and household size



Source: Sorlie et al 1995

In further analyses of the NLMS data, Backlund et al examined more closely the shape of the income mortality gradient (Backlund et al 1996). As with previous work, they confirmed that the law of diminishing returns applied – the reduction in mortality rates for each extra dollar of family income was less at higher family incomes than lower family incomes. They concluded that the log of family income had a reasonable linear association with the log of the relative risk. For example, if a doubling of income from $5000 to $10,000 reduced the mortality rate by 20% then a doubling of income from $50,000 to $100,000 would also reduce the mortality rate by 20%, despite the absolute differences in income being $5000 versus $50,000. Controlling for covariates (eg, marital status, education) reduced the income mortality gradient, but did not change the underlying shape of the income mortality gradient.

4.1.2 Controlling for labour force status, and health selection

An additional important finding of Backlund et al (1996) was the substantial change in the size of the income mortality gradient after controlling for labour force status. First, Backlund et al estimated the change in the log of the mortality risk for each $1000 of family income within three income bands: low income earners ($0-$5000 category to the $5000-$10,000 category); middle income earners ($5000-$10,000 to $20,000-$25,000); and high income earners ($20,000-$25,000 to over $50,000). That is, the association of income with mortality was allowed to differ within each of these three income bands, and within each income band the log of the relative risk was *linearly* associated with family income. Within separate demographic groups (eg, males aged 25–44), three models were fitted:

- Model 1 – controlling for age, race, and household size
- Model 2 – Model 1, plus controlling for marital status and education
- Model 3 – Model 2, plus controlling for employment status.

Using Model 1 as the baseline, the general finding was that controlling for education and marital status modestly reduced the size of the income effect, but further control for labour force status substantially reduced the income effect size. Table 51 presents the percentage reduction in the relative risk of mortality among 45–64 year olds for Model 2 compared to Model 1, and Model 3 compared to Model 1. For example, among males 45–64 years old with a low family income, controlling for education and marital status reduced the size of the income mortality gradient in the baseline model by 13%. Additionally, controlling for labour force status reduced the gradient by 84%. Except for high income females aged 45–64 years, the marginal reductions in the income mortality gradient from controlling for labour force status shown in Table 51 are all reasonably large.

Table 51: Percentage reduction in the relative risk of mortality for each $1000 of family income due to controlling for education, marital status and employment status, compared to a baseline model controlling only for age, race, and household size

Variables added to model	Family income		
	Low ($0–$10,000)	Middle ($5000–$25,000)	High ($20,000 plus)
Males 45–64 years			
Education, marital status	13%	8%	33%
Above, plus employment status	84%	54%	52%
Females 45–64 years			
Education, marital status	14%	24%	64%
Above, plus employment status	62%	50%	65%

Source: Backlund et al 1996

Akin to the analysis of Backlund et al of income mortality gradients controlling for employment status above, Stronks et al examined the effect of controlling for labour force status on the associations of education, occupational class, and income with *morbidity* (Stronks et al 1997). Stronks et al had to use a proxy variable for income – a composite variable including health insurance, housing tenure and car access. Morbidity was measured in two ways: one or more chronic conditions and self-rated health. They found that the associations of education and occupational class with morbidity altered little after controlling for labour force status, *but* the association of income with morbidity was substantially reduced. Most of this reduction in the income–morbidity association was achieved by simply excluding people with a long-term work disability. (The other employment categories were paid employed, unemployed and looking for pad work, early retired, and house-persons.) Stronks et al (1997) interpreted this as strong evidence that the association of income with morbidity, at least cross-sectionally, is in large part due to health selection – mostly drift health selection following the terminology used in this report. As acknowledged by Backlund et al (1996, p.19), it is possible that the large reduction in the income mortality gradient observed in the NLMS after controlling for labour force status was also due, at least in part, to health selection.

McDonough et al directly tested for drift health selection in a short duration follow-up study such as the NLMS (McDonough et al 1997). Using respondents to the Panel Study of Income Dynamics (PSID) aged 45 years and older, they compared the association of mortality over a five-year period with income measured either: at the commencement of follow-up (ie, one to five years before death), or five years before follow-up commenced (ie, five to 10 years before death). There was essentially no difference between the two measures. If drift health selection were occurring, one would have expected the income–mortality gradient to be stronger for income measured immediately prior to follow-up than for income measured five years prior to follow-up commencement. Thus, this study suggests no drift health selection affecting income–mortality gradients. However, 34% of the person time and 70% of the deaths were for respondents aged 65 years and older – this older population would have dominated the regression model results. Most of this older age-group would have retired from the active labour force, and therefore would not suffer an income drop as a result of poor health preceding death in the same way a younger person in the active labour force would. Unfortunately, McDonough et al (1997) do not report this particular analysis restricted to 45–64 year olds. But in further analyses by the same group of researchers confined to 45–64 year olds in the PSID, little difference was found between:

- the association of post-tax income averaged over five years and 10 years of subsequent mortality
- the association of pre-tax income over one year and 10 years of subsequent mortality (Daly et al in press).

If changing from post-tax to pre-tax income did not make much difference in the income–mortality gradient, this later analyses offers some support against strong health selection effects for mortality over a 10-year period. Unfortunately, small numbers in the PSID prohibit meaningful analyses by cause of death.

Returning to the NLMS, controlling for employment status may adjust for an overstated association of income with mortality due to drift health selection, but (and putting aside differential health selection for the moment) it may also be a form of 'over-control':

> '*The income–mortality gradient is much weaker after adjustment for employment status. However, it should be emphasised that the adjusted model assumes that income has no effect on health (as measured by employment status) prior to baseline. This assumption is probably false as data from both the United States (House et al 1990) and the United Kingdom (Arber 1987) suggest that people with lower social status have a higher incidence of debilitating illness. So while controlling for employment status helps mitigate the effect of reverse causation [ie, drift health selection], it may also cause overadjustment and an underestimate of the true income–mortality gradient.*' (Backlund et al 1996, p.19)

However, this consideration does not allow for bias introduced following control for labour force status due to differential health selection. It is plausible that low-income people are more susceptible to being forced out of employment, or out of the labour force altogether, than high-income people if they become ill. If so, controlling for labour force status could give a false impression of the association of income with mortality. One way to investigate this possibility is to look at changes in mortality risk over time by income strata among just those in the labour force. If the above bias was occurring, one would expect:

- the mortality risk in the low income strata to be low initially compared to their true background risk, as those in poor health at baseline would have been likely to be forced out of the labour force. Over time, the mortality risk would rise to the true background risk
- the mortality risk in the high income strata to only increase slightly over time, as the high income people would have been more likely to retain their jobs when in poor health.

That is, if differential health selection were operating one would expect to see a larger healthy worker effect over time in the lower compared to higher income strata. This type of investigation has not been reported for income to my knowledge, but was used in the NZCMS.

4.1.3 Cumulative and lagged income

A limitation in the measurement of income mortality gradients is mismeasurement error of the income exposure. One form of income mismeasurement error is that most studies measure income for the year before the entry point to the study – and that is it. However, income is dynamic, and varies dramatically from year to year and over the life-course corresponding to predictable (eg, child-rearing, retirement) and unpredictable (eg, unemployment) events (Lynch and Kaplan 2000). If we believe that one's average income over a period longer than a year is more important for health than income in any one given year, then we would expect studies using one-off single year measurements only to underestimate the income mortality gradient. Duncan has reviewed income dynamics in relation to health (Duncan 1996), and reported that:

- the income to needs ratio measured for one year only is essentially not associated with childhood wasting and stunting, but the average income to needs ratio over 10 years is strongly associated with childhood stunting and wasting (Miller and Korenmand 1994)
- occasional family poverty in the first five-years of life has an adverse affect on behaviour and IQ at age five compared to no periods of family poverty, but the adverse effect is even stronger for children in persistent poverty (Duncan et al 1994).

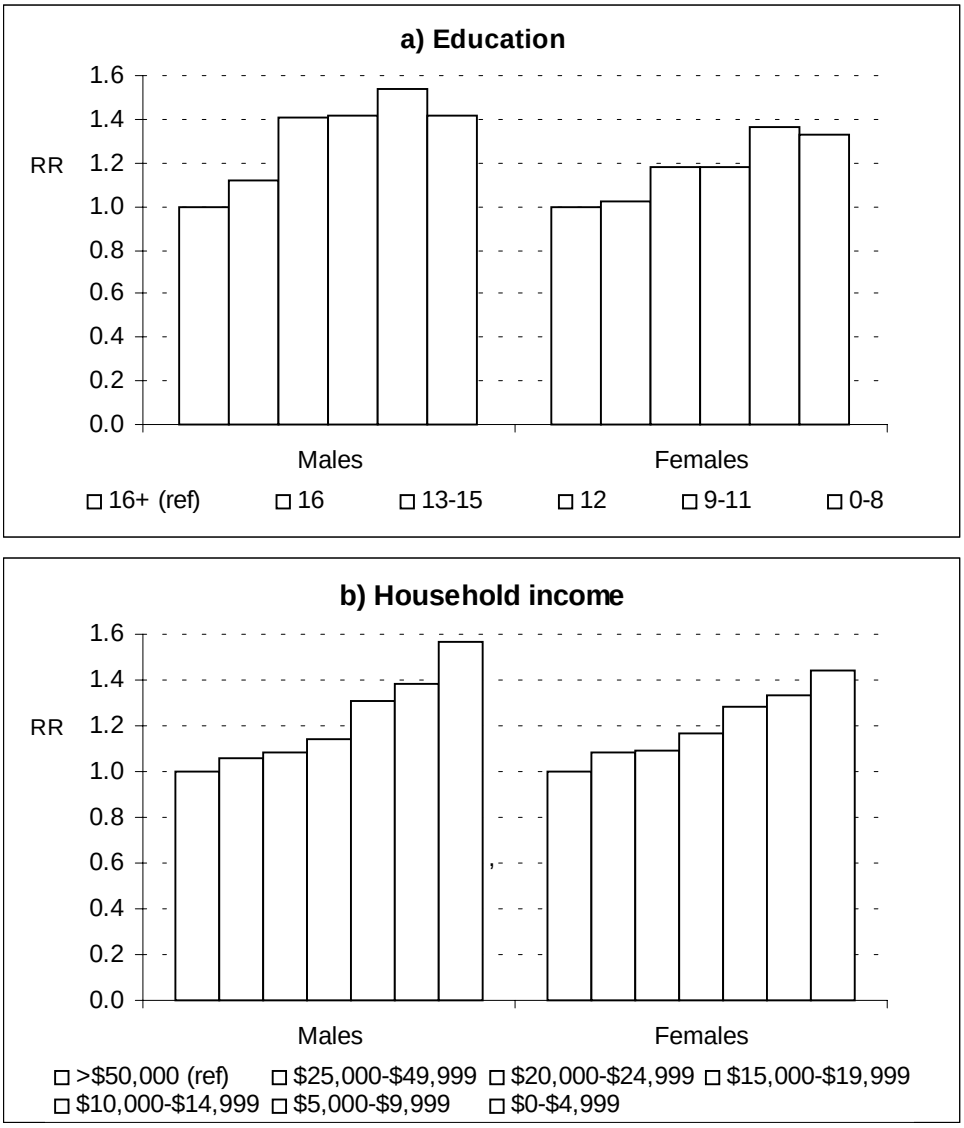
More recent work by McDonough, Duncan and others using the US Panel Study of Income Dynamics has demonstrated an effect of income *dynamics* on mortality among 45–64 year olds (McDonough et al 1997). Those with persistently low income (4–5 years) had a 40% increased odds of death compared to those with low income for 1–3 years. Second, regardless of the average five-year income before commencement of follow-up, one or more greater than 50% drops in income for one year compared to the previous was associated with an increased odds of death compared to the same average income with no drops. However, these results suggesting a stronger effect of persistently low income or sudden decreases in income may be due to confounding. People who experience either of these income dynamics may differ from those people that do not suffer these two forms of income dynamics on a range of socioeconomic and other factors.

4.1.4 Income versus education

Another assessment of the adult income mortality gradient is to examine the independent effect of adult income on mortality independent of other socioeconomic measures – most notably education. Such an examination not only captures the possible independent effect of culture and way of life (education) versus materialistic determinants (income), but also captures socioeconomic position measured in youth (education) and adulthood (income). Unfortunately, disentangling the relative contributions of either of these dimensions is difficult. Presented at the beginning of this Section, the results of both Kitagawa and Hauser (1973) and Sorlie et al (1995) suggest that there is a residual effect of income on mortality controlling for education, although Sorlie et al (1995) also controlled for a range of other variables.

Extending the analyses of the NLMS, Backlund et al specifically examined the relative contributions of education and income (Backlund et al 1999). Their major conclusion was that socioeconomic mortality gradients among males might be primarily a function of income among lower socioeconomic groups, and primarily a function of education among higher socioeconomic groups. Thus for males in Figure 25, there is a mortality protection for the most educated males (16 and 16 plus years of education) compared to little variation between the other educational categories, whereas for income there is a steepening gradient at low incomes. The pattern was less clear-cut for females. This result may in part be due to the choice of comparison categories. An additional factor in interpretation is that the Current Population Survey (the population base for the NLMS) measures income accurately with numerous questions, whereas education is elicited with just the one question. Thus, there may be greater misclassification bias of education than income, causing an underestimate of the relative effect of education. Finally, note that the results shown in Figure 25 from the NLMS control for labour force status (and occupational class for males) in addition to demographic variables (age, race, household size, and marital status). Labour force could be a pathway from education to mortality, whereas it could be a confounder of the income–mortality association. Additionally, controlling for labour force status may bias the analysis due to health selection as discussed above.

Figure 25: Relative risk of death among 25–64 year olds by: a) education controlling for household income; and b) household income controlling for education



Source: Backlund et al 1999

Note: Each model also adjusts for age, race, household size, marital status, employment status, and (for males only) occupation.

Lantz et al also compared the effects of education and income on mortality in the US, using 7.5 years of follow-up data from a study of 3617 adult males and females (Lantz et al 1998). They found that:

- both education and income were associated with health-related behaviours (smoking, alcohol, physical activity, weight) – education more so than income
- considered separately, both education and income were strongly associated with mortality – income more so than education
- considering education and income simultaneously, and controlling for demographic factors only, the income mortality gradient remained strong but the education gradient became negligible – most of the education effect on mortality appeared to be mediated via income

- the rate ratio of mortality for an income of less than $10,000 compared to an income of greater than $30,000 was 3.22 (95% CI 2.01 to 5.16) controlling for demographic factors and education. Controlling additionally for health-related behaviours only reduced this rate ratio to 2.77 (1.74 to 4.42)
- controlling for health status at the beginning of follow-up (ie, controlling for possible health selection) made no substantive difference to their findings
- the pattern of education and income gradients was similar for separate causes of death: tumours, cardiovascular, and other diseases.

In support of the finding of Lantz et al in the first bullet point, Winkleby et al found that education was a stronger predictor of cardiovascular disease risk factors than income, and that in regression models, including both socioeconomic variables education, but not income, was a statistically significant predictor (Winkleby et al 1992). Yet, Lantz et al (1998) found that income was a stronger predictor of mortality than education. Further, they found that little of the income effect was mediated by the measured health-related behaviours of smoking, alcohol, physical activity, and weight. Lantz et al concluded that the education effect was mediated via income, and that policies directed at health-related behaviours had little potential to reduce socioeconomic mortality gradients.

A more thorough investigation of cardiovascular risk factors as mediators of the association of income with all-cause and cardiovascular mortality, and acute myocardial infarction, was undertaken by Lynch et al using the Kuopio Ischaemic Heart Disease Risk Factor Study (Lynch et al 1996). Twenty-three biological (eg, fibrinogen, cholesterol, glucose, blood pressure, BMI), behavioural (eg, smoking, alcohol, physical activity), psychological (eg, depression and hopelessness) and social (eg, marital status, social support) risk factors were available for inclusion in analyses. The excess rates of all-cause mortality, cardiovascular mortality, and acute myocardial infarction in the bottom income quintile compared to the top income quintile decreased by 85%, 118%, and 45%, respectively, after controlling for these 23 cardiovascular disease risk factors. Thus, Lynch et al stated their data showed that the association between income and mortality can be explained by putative mediating factors – as long they are all specified.

4.2 Conclusion

Income is a powerful predictor of mortality. Most research suggests that income remains a strong independent predictor of mortality when controlling for other socioeconomic factors. Controlling specifically for education, both income and education have strong independent associations with mortality – usually stronger for income. However, it is important to remember that income is a likely mediating factor between education and mortality – not just a confounder.

In some studies, controlling for labour force status has been shown to dramatically reduce the association of income with mortality. Whether this reduction demonstrates that the association of income with mortality is affected in large part by health selection is unclear.

There are several methodological issues with the measurement of income–mortality gradients, including health selection, mismeasurement of income, and control of confounders (or prior exposures in the life-course). Regarding mismeasurement, incomes are known to be dynamic over the life-course. Available evidence suggests that the association of income averaged over years with health is probably stronger than that for income in any one given year. Individuals who experience sudden and repeated drops in income over time appear to have a higher mortality risk than those with the same average income, but a more stable income. Regarding the control of confounders, it is unclear exactly what effect on mortality, and with what time lag, a change in income alone (and no change in other socioeconomic factors) would have.

5 Asset wealth: car access and housing tenure

Housing tenure and car access are taken to be measures of material socioeconomic position, particularly as proxies for wealth, income and social class (Fox and Goldblatt 1982; Macintyre et al 1998; Moser et al 1988; Stronks et al 1997). While seemingly crude measures of socioeconomic position, they are remarkably strong predictors of mortality. If research were conducted with more refined, accurate and precise measures of asset wealth the association with mortality would probably be stronger again.

In a study of morbidity outcomes Macintyre et al (1998) tested whether the association of both car access and housing tenure with health was due to being a proxy for either income or self-esteem. When including income and self-esteem in their models, the association of car access and tenure with health was essentially unchanged. Macintyre et al concluded that car access and housing tenure are more than proxies for income and self-esteem. However, and as acknowledged by the authors, more complete and accurate measures of income and self-esteem may have explained more of the observed association of car access and housing tenure with health (Macintyre et al 1998).

Car access is also probably a measure of access to community resources, such as employment opportunities, shops selling healthy food at affordable prices, leisure facilities, and social support networks. It may also reduce exposure to dangers such as mugging, rape or assault (Macintyre et al 1998).

Housing tenure may also capture facets of housing quality (eg, dampness and mould) and overcrowding that affect health (Howden-Chapman et al 1996). Further, both owning your own car and/or house may reflect psychosocial traits (eg, self-efficacy and self-esteem) and confer a sense of accomplishment and security.

The direction of the relationship of car access with health is such that having more cars in the household is better for one's health from a physical and global environment perspective, though increased car ownership is deleterious to the environment and eventually 'feeds back' to adversely affect individual health (McCarthy 1999). Car access being associated with better health at the individual-level is an example of the atomistic fallacy (Diez-Roux 1998) at a population-level higher rates of car usage (all other things being equal) would probably be detrimental to population health. Furthermore, walking and biking to work would probably be better for individual health than commuting by car. Thus the association of car access with better health should be treated cautiously. It is not so much the car access that is important, but that the person in question owns an asset and/or has access to community resources.

5.1 Findings

Car access and housing tenure have been used as measures of socioeconomic position mainly in the UK, and most notably in the OPCS Longitudinal Study (Filakti and Fox 1995; Fox and Goldblatt 1982) and the Whitehall Study (Davey Smith et al 1990).

5.1.1 Car access

During 1981–89 in the OPCS Longitudinal Study the age-standardised rate ratio for all-cause mortality was 1.41 (95% confidence interval 1.29 to 1.54) comparing males with no car access to males with one or more cars in the household. For females, the rate ratio was 1.24 (1.11 to 1.39). The rate ratios were greater among both males and females less than 65 years of age (Filakti and Fox 1995).

Among the males aged 40–64 years in the Whitehall Study I, and followed up for mortality during the 1970s, the age-adjusted relative mortality rate associated with not owning a car was 1.49 (1.4 to 1.7). While occupational grade and car ownership were correlated, a substantial relative risk of 1.28 (1.1 to 1.5) remained after adjusting for occupational grade. For cardiovascular deaths, further adjustment for blood pressure, smoking, cholesterol and glucose intolerance did not greatly affect the association of car access with mortality (Davey Smith et al 1990).

5.1.2 Housing tenure

During 1981–89 in the OPCS Longitudinal Study the age-standardised rate ratio for all-cause mortality was 1.19 (95% confidence interval 1.09 to 1.30) comparing males living in private rental households with males living in owner-occupied households. For males living in local authority rental households the rate ratio was 1.34 (1.22 to 1.44) compared to owner-occupiers. For females, the same rate ratios were 1.19 (1.06 to 1.33) for private rental households and 1.32 (1.19 to 1.47) for local authority rental households. The rate ratios were greater among both males and females less than 65 years of age (Filakti and Fox 1995). Housing tenure in the OPCS Longitudinal Study remained strongly associated with mortality within strata of occupational class (Fox and Goldblatt 1982).

In a Swedish longitudinal study of 32,853 people, the all-cause mortality rate ratio in rental accommodation compared to owner-occupied was 1.56 (1.39 to 1.75) for males and 1.66 (1.44 to 1.92) for females, controlling for age only. Controlling further for ethnicity, marital status, work capacity, occupational class, and educational level, the rate ratios reduced to 1.27 (1.12 to 1.43) and 1.43 (1.22 to 1.66) for males and females, respectively (Sundquist and Johansson 1997). Thus, as in the OPCS Longitudinal Study, housing tenure had a strong and independent association with mortality.

5.2 Conclusion

Car access and housing tenure are powerful and independent predictors of health and mortality. Why is not clear. In the NZCMS they are considered as proxies for asset wealth unless stated otherwise.

Appendix B: Sources of Error Involved in Measuring Socioeconomic Mortality Gradients

Sources of error in epidemiology may be considered as either random or systematic (Rothman and Greenland 1998). Regarding systematic error, these may be categorised as selection bias, confounding, and information bias.

Selection bias occurs 'the relationship between exposure and disease is different for those who participate and those who should be theoretically eligible for study ...' (Rothman and Greenland 1998, p.119). If the eligible population is the population of a country, and census data forms the study-base, then selection bias can only arise if the census data set has to be restricted for some reason (eg, availability of household income).

Confounding occurs when 'the apparent effect of the exposure of interest is distorted because the effect of an extraneous factor is mistaken or mixed with the actual exposure effect ...' (Rothman and Greenland 1998, p.120). The properties of a confounder are that it:

- must be associated with the exposure
- must be associated with the outcome
- must be independently associated with the outcome among the unexposed
- and must *not* be [exclusively] an intermediate variable between the exposure and the outcome.

Given the social structuring of society, and the fact that most socioeconomic factors are correlated, it is difficult to know when it is a appropriate to consider the independent effect of some socioeconomic factor controlling for potential confounders (including other socioeconomic factors) (Kaufman and Cooper 1999).

Information bias (also known as misclassification bias) occurs whenever there are errors measuring exposures, outcomes and covariates. The direction of bias varies depending on whether the measurement error on one variable is correlated with that on another variable. How information bias might affect measurement of socioeconomic mortality gradients is further discussed in subsequent sections of this chapter.

It is neither feasible nor helpful to attempt a comprehensive inventory of all the possible sources of selection bias, confounding and information bias that may affect the measurement of socioeconomic mortality gradients. Indeed, most sources of systematic error will be specific to a given study design and/or a particular study. However, there are three sources of error that arise in the many studies that measure socioeconomic mortality gradients, and warrant specific mention: *confounding*, *health selection* (not 'selection bias'; may also be considered as a special form of confounding), and *misclassification* of socioeconomic exposures, confounders, and the mortality outcome. Health selection has received considerable attention in the literature on socioeconomic determinants of health.

1 Confounding

Consider the association of income with mortality and the potential confounder education. Education and income are correlated, and education is associated with health outcomes. It is also probable that education is independently associated with mortality risk within strata of income (including the 'unexposed' income strata). Finally, for adults at least, education is prior to income in the life-course (see Figure 3 for initial framework or causal diagram used in this report), so education is unlikely to lie on the causal pathway between income and mortality risk. Thus, education is a potential confounder of the association of income with mortality.

Conversely, consider the association of education with mortality, and income as a possible confounder. As income is potentially on the causal pathway between education and mortality, adjusting for income may result in 'over-controlling' of the total effect of education on mortality. However, as education does not fully determine income, some of the effect of income on mortality may not be on the causal pathway from education to mortality, there will be some confounding of the education–mortality association by income. This is a critical (and largely unresolved) problem in modern epidemiology – how to control for a covariate that is in part a confounder, and in part on a causal pathway between the exposure of interest and the outcome. This issue is revisited in Section 5: Epidemiological analysis in the context of causal pathways, confounding, intermediaries, and misclassification bias.

Note that some authors use the term 'indirect selection' to denote confounding by variables that are prior in some causal chain – particularly causal chains over the life-course (Bartley et al 1999; Martikainen and Valkonen 1996; van de Mheen 1998). This usage of the term 'indirect selection' serves to remind us that many variables that satisfy the properties of a confounder are actually prior variables (in part at least) on causal pathways. For example, education is not just spuriously correlated with income. Rather, one's income is in large part determined by education, and hence education is part of the complex of factors that 'select' income.

2 Misclassification of socioeconomic exposures and confounders

Misclassification (or mismeasurement) of independent variables occurs in all studies. Considering socioeconomic variables elicited by a census, misclassification can be thought of at two levels. First, there is misclassification compared to the correct answer for the question as written. For example, personal income data is collected by just one 'tick-box' question on the New Zealand census, to which most people would select a best-guess answer. Second, there is misclassification compared to the underlying socioeconomic factor of interest. For example, we may interpret the results for 'income' as applicable to 'usual income' or 'income accumulated over the life-course'. However, the census questionnaire may only ask about income in the last year.

It is likely that such misclassification of *independent variables* (ie, exposures) is mostly non-differential with respect to the mortality outcome – but it cannot be assured. Such non-differential misclassification bias usually causes a bias to the null in the observed association of each socioeconomic exposure with the mortality outcome (Rothman and Greenland 1998).

Non-differential misclassification bias of *confounders* hampers the ability to control for confounding (Davey Smith and Phillips 1992; Greenland 1980; Liu 1988; Marshall and Hastrup 1996; Phillips and Davey Smith 1992; Phillips et al 1996). For example, if we control the association of income with mortality for the confounder education, and education is non-differentially misclassified, then there will remain residual confounding of the income–mortality association by education. Therefore, controlling for non-differentially misclassified confounders will underestimate the percentage of the association of the socioeconomic exposure with mortality that is explained by that confounding variable. However, the effect measure (eg, risk ratio) for a given socioeconomic exposure, controlled for the confounder(s), could be biased in either direction depending on the *combined effect* of misclassification of *both* the exposure and the confounder(s).

Table 52 shows a hypothetical example of the effect of non-differential misclassification of either/both the exposure (income) or confounder (education) under certain assumptions. First, assume that the mortality risk ratio for high compared to low income (exposure) was 3.0 in the absence of any non-differential misclassification bias, and controlling for the perfectly measured confounder education the risk ratio was 2.0. That is, 50% of the excess association of income with mortality was due to confounding by education. This risk ratio of 2.0 is shown for the top left cell of Table 45 where there was no (0%) misclassification of either exposure or outcome. Second, assume that non-differential misclassification of the exposure resulted in (RR-1) being underestimated by 60% – thus the crude RR is (1 + ([1 – 60%] × [true crude RR – 1]) = (1 + ([40%] × [2]) = 1.8, where the true crude RR is 3.0 as above. Further, assuming that adjusting for the perfectly measured confounder (education) reduced the observed excess risk ratio (crude observed RR – 1) by 50% (as it does for the true crude RR), then the observed adjusted risk ratio is (1 + ([50%] × [observed crude RR – 1]) = 1.4. This risk ratio is shown in the top right cell of Table 52.

Table 52: Hypothetical example of the effect of misclassification of both exposure and confounder, where: the crude RR is 3.0 without misclassification bias of exposure; the adjusted RR is 2.0 without misclassification of either exposure or confounder

% underestimate of confounding of observed (RR-1) due to misclassification of confounder	% underestimate of (RR-1) due to misclassification of exposure			
	0%	20%	40%	60%
0%	2.0	1.8	1.6	1.4
20%	2.2	2.0	1.7	1.5
40%	2.4	2.1	1.8	1.6
60%	2.6	2.3	2.0	1.6

Third, assume there was no misclassification of the exposure, but that misclassification of the confounder caused a 60% underestimate of confounding due to education. Thus, the percentage reduction in the excess risk ratio (RR – 1) for income after controlling for education will be 50% × (1 – 60%) = 20%, where 50% is the amount of the crude exposure–outcome association that is truly due to confounding by education. Consequently, the observed adjusted RR will be (1 + ([100% – 20%] × [crude RR – 1]) = 2.6. This risk ratio is shown in the bottom left cell of Table 52. Finally, assume that non-differential misclassification bias affects both the exposure and the confounder, and that the percentage underestimate due to misclassification of the confounder applied to

the *observed* crude risk ratio. If the crude exposure–outcome association is underestimated by 60% due to misclassification of the exposure, and the percentage underestimate of confounding is 60%, then the observed adjusted risk ratio is now 1.6 (ie, (1 + ([100% – 20%] × [observed crude RR – 1]) = (1 + ([80%] × [1.8 – 1]) = 1.64 ≈ 1.6.

The hypothetical example in Table 52 assumes that the percentage underestimates from misclassification of confounder and exposure act independently of each other, and that the percentage of the *observed* exposure–outcome association explained by the confounder is the same regardless of the amount of exposure misclassification. These assumptions may not hold. However, the point illustrated in Table 52 is that the observed adjusted risk ratio may be either under or overestimated with respect to the true adjusted risk ratio of 2.0 when there is misclassification bias of *both* the exposure and confounder. Put another way, the net effect of misclassification bias of both the exposure and the confounder may be to cause either an under or an overestimate of the adjusted risk ratio compared to the true adjusted risk ratio.

3 Misclassification of the mortality outcome

In the previous section I considered misclassification bias of the exposure and covariates. In this section I consider misclassification bias of the mortality outcome. In many cohort studies this potential bias is minor as the assessment of vital status is accurate. However, in a linked census–mortality study such as the NZCMS the bias may be more substantial. For example, not all deaths may be ascertained, and some ascertained deaths may be incorrect. Misclassification of the outcome was an issue in the NZCMS, *but it was possible to quantify the resultant bias in the cohort analyses* (unlike misclassification bias of the exposures and covariates).

Other census–mortality studies have also had limitations with the completeness and accuracy of the record linkage, for example the US 1960 Matched Records Study (Kitagawa and Hauser 1973), the Canadian study (Houle et al 1996), and the Italian study (Faggiano et al 1995). The percentage of records linked and estimated accuracy (where available) in these other linked census–mortality studies was presented in Section 1.2 of Chapter 1, Table 2).

Table 53 is a two-by-two table of the actual outcome (dead or alive at the end of the follow-up period) cross-classified by the assigned outcome (linked or unlinked) in a linked census–mortality study.

Table 53: Two by two table of link/non-link status by vital status in a linked census–mortality study

		True vital status at the end of follow-up		
		Died	**Alive**	
Output from record linkage	Linked	a	b	a + b
	Unlinked	c	d	c + d
		a + c	b + d	a + b + c + d

Where for each of the four cells in Table 53:

a = true links, or true positives

b = false links, or false positives

c = false non-links, or false negatives

d = true non-links, or true negatives.

This misclassification of the mortality outcome by the record linkage can be characterised with the following familiar terms:

Sensitivity	= Se	= a / (a+c)
Specificity	= Sp	= d / (b+d)
Positive predictive value	= PPV	= a / (a+b)
Negative predictive value	= NPV	= d / (c+d)

(Note the population distributed in Table 53 is *different* to that in Table 10. Table 10 is a two-by-two table of comparison pairs. Thus each census and mortality record is represented many times in the table. Table 53 is a two-by-two table of census records. Correspondingly, the meaning of false links and true non-links, and the specificity and the positive predictive value derived from Table 10 *differ* from that derived from Table 53.)

In most record linkage projects it is not possible to determine the actual vital status for individual census records, except perhaps for a subgroup (eg, Calle and Terrell 1993; Houle et al 1996). However, the number of linked (a+b) and unlinked records (c+b) are known. The sensitivity of the record linkage may be approximated by the percentage of submitted records linked (ie, [a+b]/[number of eligible mortality records]), but will not necessarily give a correct approximation as:

- the numerator (a+b) includes false positive links (b)
- the denominator is not necessarily the number of deaths in the cohort. For example, it may be an overestimate if many of the mortality records were for people that were absent from the country on census night or simply failed to complete the census form. Alternatively, it may be an underestimate if many of the census respondents that actually died in the eligible follow-up period emigrated before death.

It was also possible to estimate the PPV in the NZCMS – the methods to do so are presented in the Technical Report (Blakely et al 1999) and presented briefly in the Methods Chapter of this report. Thus, using these estimated values of the sensitivity and PPV it was possible to estimate each of the cell values in Table 53. Further, it would also have been possible to estimate the sensitivity and specificity by strata of socioeconomic factor (eg, small area deprivation) and covariates (eg, age) in the NZCMS. Having made these estimates, then it would have been possible to correct the observed effect measures (eg, risk differences and risk ratios) for misclassification bias of the mortality outcome using published correction formulas (Brenner and Gefeller 1993; Copeland et al 1977; Green 1983). Appendix B of my PhD thesis (Blakely 2001) (*not* Appendix C of this report) presents a number of illustrative examples of how the varying sensitivity and specificity of a record linkage project may bias the cohort analyses. Of particular note, the risk ratio may be relatively unaffected by misclassification bias of the mortality outcome while the risk difference was notably biased. The supporting material in Appendix B also helps to *understand* the linkage bias in the NZCMS. However, to actually *attempt* a sensitivity analysis of the impact of misclassification bias

of the mortality outcome in the NZCMS using the correction formula presented in Appendix B of the PhD thesis would have been cumbersome, complex and probably inaccurate. Fortunately, it also proved unnecessary – a more direct method of adjusting for linkage bias was used. The method simply used results from analyses that determined the relative risk of a mortality record being to linked to a census record by demographic and socioeconomic factors, and used these relative risks to adjust the odds ratios of mortality observed in the cohort analyses.

4 Health selection

Consider the association of labour force status with mortality. It is probable that people with poor health will exit the active labour force more frequently than people in good health. As poor health is associated with increased mortality, then the mortality risk among those in the non-active labour force will usually be higher than the mortality risk for those in the active labour force. This mortality difference will not be a function of any independent effect of labour force status on mortality, but rather a function of the differential mobility by health status of people between categories of labour force status – the so-called 'healthy-worker effect':

> '*Workers usually exhibit lower overall death rates than the general population, because the severely ill and chronically disabled are ordinarily excluded from employment.*' (Last 1995).

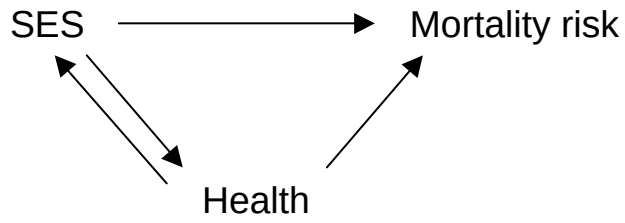
This phenomenon is not just observed with respect to labour force status. For example, people may slide down the income ladder as a result of poor health. Thus, Fox et al (1987) provide a more generalised and formal definition of health-related mobility as:

> '*... the artificial raising or lowering of the average health of people with a particular characteristic associated with the process by which that characteristic is acquired or lost. The mortality of a population with that characteristic is affected by health-related mobility if the health of people acquiring or losing the characteristic differs systematically from others with the characteristic.*' (Fox et al 1987).

This phenomenon has been given various names, eg, health selection bias, health selection, social mobility, direct selection (as opposed to 'indirect selection' mentioned above in Appendix B: Confounding) and health-related mobility. The default term used in this report is 'health selection'. Note that health selection is not a 'selection bias' – all the eligible population may be included in the analysis (meaning there is no selection bias), yet the phenomenon persists. Rather, it is, a special form of confounding as labour force status (a proxy for health status) has an influence on measured socioeconomic position, and labour force status also influences mortality risk. But labour force status (as a proxy for health status) is also an intermediary variable on the causal pathway between socioeconomic position and mortality, thus breaching one of the properties of a (pure) confounder (Rothman and Greenland 1998).

An alternative way to consider health selection is as reverse causation. The causal diagram below represents reverse causation – socioeconomic position not only affects health (arrow from socioeconomic position to health), but health status affects socioeconomic position (arrow from health to socioeconomic position). (Labour force status can be substituted as a proxy for health in the causal diagram.) For example, poor health may cause a decrease in one's income, as well as low income being a determinant of health status. As health status is also causally related to mortality risk

(arrow from health to mortality risk), the observed association of income with mortality risk may be 'biased' by health selection. Extending the income example, health selection may result in an increased mortality risk among those with low income and a corresponding decreased mortality risk among those with high incomes. Thus, assuming an underlying association of increasing income with lower mortality risk, health selection may exaggerate the income mortality gradient.



For the purposes of this report, it is crucial to recognise two variants of health selection, ***drift* health selection** and ***differential* health selection**. *Drift* health selection refers to the above income example – a combination of health (or labour force status as a proxy) as a confounder and an intermediary variable. Here, people drift up and down the income ladder conditional on their health status, and this in turn may cause the income mortality gradient to be overestimated. Note that this 'overestimate' occurs only if current income at the start of follow-up is used as a proxy for long-term, usual, life-time or some other longer-term measure of income. *Differential* health selection occurs when a socioeconomic mortality gradient is assessed among the active labour force only, and bias arises due to exclusion from the active labour force being differential by socioeconomic factor. Classically, differential health selection has been described for occupational class mortality gradients (Martikainen and Valkonen 1999). Here the gradient is underestimated when only *current* occupation is available for the assignation of occupational class. This underestimation is because the lower occupational classes (based on *usual* occupation) are more likely to be forced out of the labour force than the higher occupational classes when in poor health. This differential health selection out of the labour force causes the observed mortality risk/rate among the lower occupational classes (based on *current* occupation) to be underestimated more than the risk/rate among the higher occupational classes. To extend the framework of health selection in terms of classic epidemiological terms, differential health selection represents effect modification of the socioeconomic mortality gradient by labour force status.

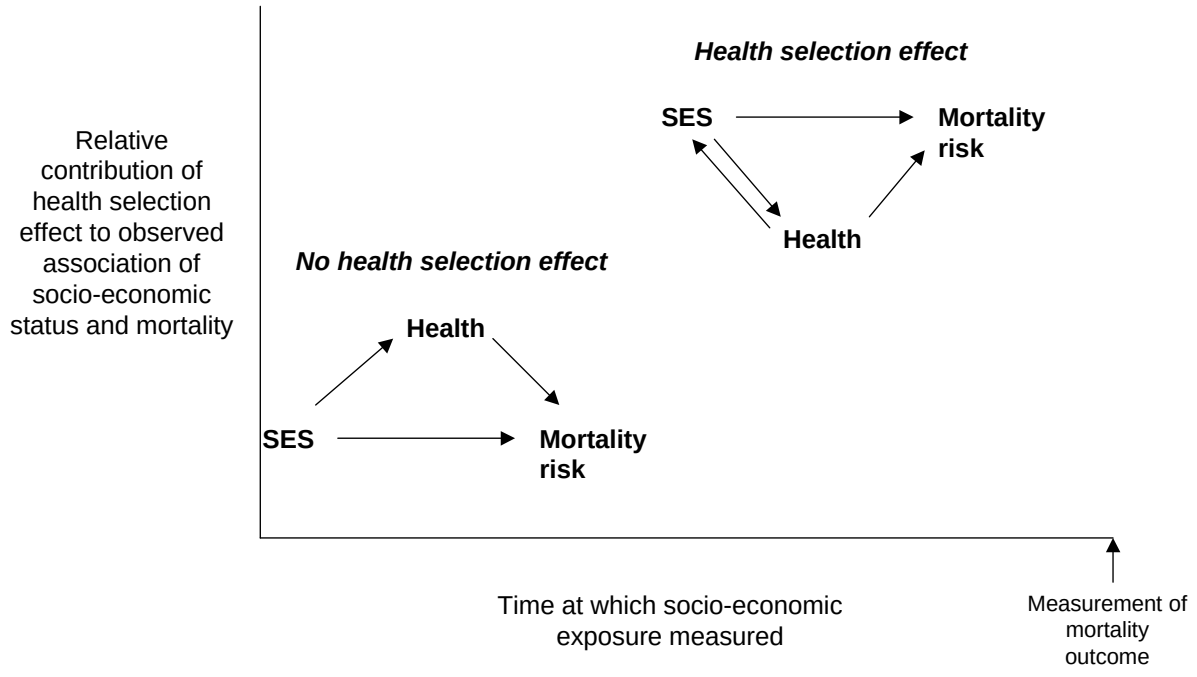
The consideration of drift versus differential health selection will be a recurrent theme in this report.

Health selection may occur over a range of time periods. It may operate in the short term such that a person's ill health precipitating exit from the workforce a year or so before death creates a spurious association between labour force status and mortality. Second, it may operate over an individual's life such that ill (or excellent) health in childhood leads to a lower (or higher) socioeconomic trajectory and, if subsequent death is related to the childhood illness, introduces an element of health selection into the relationship between socioeconomic factors and death (Bartley et al 1999; Kuh et al 1997). Third, it may operate intergenerationally with parental ill (or excellent) health influencing childhood health. For each of these three types of health selection, it is possible to argue that they are either a bias or a causal mechanism – the choice is just a matter of perspective (Blane et al 1993; van de Mheen 1998). For example, from a life-

course perspective, the influence of childhood health status on adult socioeconomic position is of causal interest.

On occasion, short-term health selection is of substantive interest, as opposed to being considered a bias. For example, the New Zealand National Health Committee (with input from J. Mackenbach, Erasmus University, Rotterdam) identified four target areas for intervention to reduce health inequalities (National Health Committee 1998). One of the four areas was the effect of ill health on socioeconomic position (reverse causation). One policy intervention might be adequate income maintenance for people who become ill, thus preventing deterioration in socioeconomic position consequent upon poor health. While acknowledging that short-term health selection may be of direct research and policy interest, the position taken in this report is that health selection over the short-term (ie, a couple of years) is a bias affecting the 'true' underlying association of socioeconomic position with health status. Intragenerational health selection is considered over many years and intergenerational health selection of causal interest, but the issue is of little direct relevance to this report as follow-up in the NZCMS was only for three years. *Unless stated otherwise, 'health selection' in this report refers to that over the short term.*

Figure 26: Possible contribution of health selection to the observed association of socioeconomic position and mortality, by time at which the socioeconomic exposure is measured relative to mortality follow-up



How can bias from short-term health selection (be it differential or drift health selection) be avoided in epidemiological analyses? Two general strategies are available. First, analyses can be restricted to those in good health at the start of follow-up. (Considering health selection as a special form of confounding, this equates to 'restriction' as a method to control for confounding.) However, as much of the 'true' association of socioeconomic position with mortality is mediated by health status, such a restriction may cause an underestimate of the socioeconomic mortality gradient. Second, the analyses can be conducted only for mortality outcomes ascertained sometime after the socioeconomic exposure is measured. For example, income could be measured in 1991, and only deaths

occurring during 1996–98 included in the analyses. The assumption here is that by delaying the ascertainment of mortality outcomes, one has allowed for all the people in poor health in 1991 (and hence also with a low income) to either have died or overcome their poor health. Obviously, there is no point in time when health selection switches on or off. Figure 26 attempts to demonstrate this by illustrating that the shorter the time period between exposure measurement and mortality follow-up, the greater the likely contribution of (short-duration) health selection to the observed association of socioeconomic position and mortality. The actual amount of bias due to health selection varies by time, socioeconomic factor, and cause of death.

Figure 26 suggests that the bias caused by health selection is a smooth linear function directly proportional to time between measure of exposure and mortality outcome. This portrayal is a simplification considering that the short-term health selection varies for acute and chronic illness. For example, Goldblatt and Fox (1979) found that people in hospital on the night of the United Kingdom 1971 census, but not usually resident at that hospital (ie, they had an acute illness), had a rapidly decreasing mortality rate. In the first year of follow-up the mortality rate was ten times higher than the general population's, but fell to a level two to three times that of the general population in the next four years (Goldblatt and Fox 1979). By comparison, people with a hospital as their usual residence (ie, they had a chronic illness) had a mortality rate about three times higher than the general population's in the first year, falling only slowly to about twice the general population's over the next four years. The implication for the NZCMS is that health selection – if any – may 'wash out' within the first year for acute illness, but not so for chronic illness.

A more detailed review of the empirical evidence for health selection with regard to each socioeconomic factor is included in the literature review in Appendix A.

5 Epidemiological analysis in the context of causal pathways, confounding, intermediaries, and misclassification bias

Most studies (Davey Smith et al 1997; Gliksman et al 1995; Vagero and Leon 1994), but not all (Lynch et al 1994), demonstrate that both parental/early childhood and adult socioeconomic position contribute independently to mortality risk. Furthermore, adult socioeconomic position is itself determined in large part by a person's life-history of accumulated social (dis)advantage (Kuh et al 1997). Thus, determining the independent effect of, say, adult income on mortality risk requires controlling for all prior variables on the pathways involving income (Blakely and Woodward 2000a). Kaufman and Cooper (1999) have argued that no matter how many covariates are controlled there will still be residual differences between the exposed and unexposed socioeconomic groups on unmeasured covariates due to the non-random distribution of socioeconomic factors (Kaufman and Cooper 1999). Thus, they argue, the 'counterfactual scenario' that underpins the control of confounding in epidemiology can rarely be realised in social epidemiology, as we cannot guarantee that the exposed and unexposed are exchangeable within strata of the measured covariates. These limitations are not peculiar to social epidemiology (eg, it is unlikely that all confounders of the association of caffeine with pancreatic cancer are controlled in epidemiological analyses). However, given that socioeconomic characteristics are acquired due to common structuring of individuals within society, co-linearity is a particularly notable problem in social epidemiology. Indeed, Kaufman and Cooper argue that it is often inappropriate to consider the 'independent' effects of separate socioeconomic factors. Analyses of the association of one socioeconomic factor with mortality controlling for other socioeconomic factors are

presented in this report. These analyses must be interpreted with caution due to the relatively few measures of adult socioeconomic position available in the NZCMS and wider concerns raised by Kaufman and Cooper. Nevertheless, there is value in presenting these multivariate analyses – particularly in policy terms. For example, it is useful to have some idea whether any excess of suicide deaths among the unemployed remains after controlling for income and education.

The usual approach in epidemiology when considering variables ordered in some causal chain is to conduct a series of analyses to determine the total and residual effects of each variable (be they exposures and/or intermediaries). Specifically, Victora et al (1997) summarise a conceptual framework for the control of confounding by the use of hierarchical analyses for studies that do not have the luxury of repeated longitudinal measures (Victora et al 1997). Figure 3 is the initial framework in the NZCMS and this report. According to this framework (Victora et al 1997):

1 The univariate association of education with mortality in the NZCMS will represent the total effect (*direct* plus *indirect* through other socioeconomic factors) of education on mortality. (Note that 'direct' refers to the arrow directly from education to mortality in Figure 3, but more correctly captures the effect of education through variables other than those shown in Figure 3.)

2 (a) The association of education with mortality adjusted for labour force status, income, and car access is assumed to represent the *direct* effect only of education on mortality.

(b) The difference between the size of the association of education with mortality observed in 1 and 2(a) is assumed to be due to the *indirect* effect of education on mortality (ie, that mediated by labour force status, income, and car access).

3 The univariate association of income with mortality will be confounded by socioeconomic factors that are prior to income in causal pathways. However, the association of income with mortality controlling for the other socioeconomic factors will represent the residual unconfounded effect of income on mortality. That is, we assume counterfactually that the effect of changing an individuals income on mortality risk (but changing nothing else) is estimated by the measured effect of income in an observational study controlling for the measured confounders.

This hierarchical framework explicitly outlined by Victora et al (1997) is implicit in many studies. For example, one commonly cited finding of the Whitehall Study is that only a third of the association of occupational grade (ranking of occupations within the British civil service) with coronary heart disease mortality was 'explained' after adjusting for known cardiovascular risk factors (Marmot et al 1984). This finding is then usually interpreted as suggesting that other factors (eg, unmeasured and/or unknown dietary and lifestyle behaviours, psychosocial factors, or material factors consequent on occupational grade) must be responsible for the remaining two-thirds of the mortality gradient. While this type of analysis and interpretation is common in all branches of epidemiology, it is prone to error. Considering points 1 to 3 above, the sources of error (other than selection bias) might be listed as follows:

1 '*The univariate association of education with mortality represents the total education-mortality effect.*'

- Confounding. There may be uncontrolled confounders that are not on the causal pathway from education to mortality (eg, parental income, early childhood factors, ethnicity).
- Confounding. There may be variables that are both confounders *and* intermediary variables between education and mortality (eg, income, smoking). Not allowing for the confounding *component* of these variables means that the univariate total education–mortality association is confounded. Further, methods to control for just the confounding component of a variable that is both intermediary and confounder are complex, requiring longitudinal data (see Rothman and Greenland, 1998, pp.422–5).
- Information bias. Misclassification of either education or mortality may bias the observed education-mortality association.

2 (a) '*The association of education with mortality after controlling for intermediary variable(s) represents the (direct) effect of education on mortality via pathways not including the intermediary variable(s) (eg, income, labour force status and car access).*'

- Confounding. As with the first point above, there may be uncontrolled confounders that are not on the causal pathway from education to mortality (eg, parental income, early childhood factors, ethnicity).
- Confounding. Controlling for income, labour force status, and car access (if specified and measured perfectly) is often *assumed* to give the direct effect of education on mortality, by both controlling for the confounding by these variables and adjusting further for the component of the education–mortality association that is mediated by these variables. However, the latter assumption is unreliable (Poole and Kaufman 2000; Robins and Greenland 1992). Using a counterfactual model, Robins and Greenland have demonstrated that that even when the total effect of exposure on outcome is completely unconfounded (ie, by both intermediary and other confounding variables), analyses of the exposure-outcome association stratified by a true intermediary variable **does not reliably** give the residual effect of the exposure on outcome via pathways other than the intermediary (ie, stratification) variable (example 1, Robins and Greenland 1992). This finding is counterintuitive, and, simplistically, arises as while the crude exposure–outcome association is unconfounded, *within strata of the exposure the intermediary–outcome association is confounded*. Poole and Kaufman (2000) have returned to this counterfactual problem initially proposed by Robins and Greenland (1992), presenting a social epidemiology example to remind us that simply controlling intermediaries in epidemiological analyses may be an unreliable strategy (Poole and Kaufman 2000). How unreliable this method is remains unclear, and requires further methodological research. However, the counterfactual example used by these authors is noted, while being totally unconfounded by definition for the exposure has a completely arbitrary user-specified distribution of people by respondent type, and it is this arbitrary distribution that causes the lack of reliability of the method.

- Information bias. Misclassification of any of the exposure (education), outcome (death) and intermediary (income) variables may be a source of error. If the sources of error in the above two bullet points were absent, non-differential misclassification bias of income only would usually cause an overestimate of the residual direct effect of education on mortality. This overestimate would arise due to incomplete control of the intermediary–outcome association. Non-differential misclassification bias of education alone, or both education and income (but not mortality), could cause either an under or overestimate of the residual direct effect of education on mortality, depending on the net impact of misclassification of each variable as shown on page 193 (Davey Smith and Phillips 1992; Liu 1988; Phillips and Davey Smith 1992; Phillips et al 1996).

2 *(b)* '*The difference between 1 and 2.(a) represents the (indirect) effect of education on mortality via pathways including the intermediaries (eg, income, labour force status, and car access).*'

The sources of error here are the converse of those for 2(a).

3 '*The association of income with mortality after controlling for confounding variables, say education, labour force status, and car access as in Figure 3, represents the unconfounded effect of income on mortality.*'

This interpretation is as for any epidemiological analysis, and akin to Point 1 above:

- Confounding. There may be uncontrolled confounders that are not on the causal chain from income to mortality (eg, genetic endowment, component of diet not determined by income).
- Information bias. Misclassification of either income or mortality may bias the observed income–mortality association.
- Information bias. Misclassification of the confounding variables adjusted for in the analyses will mean that residual confounding remains.

Appendix C: Sensitivity Analyses of Univariate Results

1 Small area deprivation

The emphasis in this report was on individual-level socioeconomic factors. However, analyses of the association of NZDep91 (a small area measure of socioeconomic position) with mortality provided an invaluable tool for assessing the likely impact of:

- selection bias (the majority of the full cohort had a NZDep91 score)
- misclassification bias of the mortality outcome (NZDep91 scores were available for 90% of the eligible mortality records)
- and health selection effects. (No drift health selection was expected for small area deprivation. Thus, the analyses by NZDep91 set a comparative baseline for later analyses by income where drift health selection was possible.)

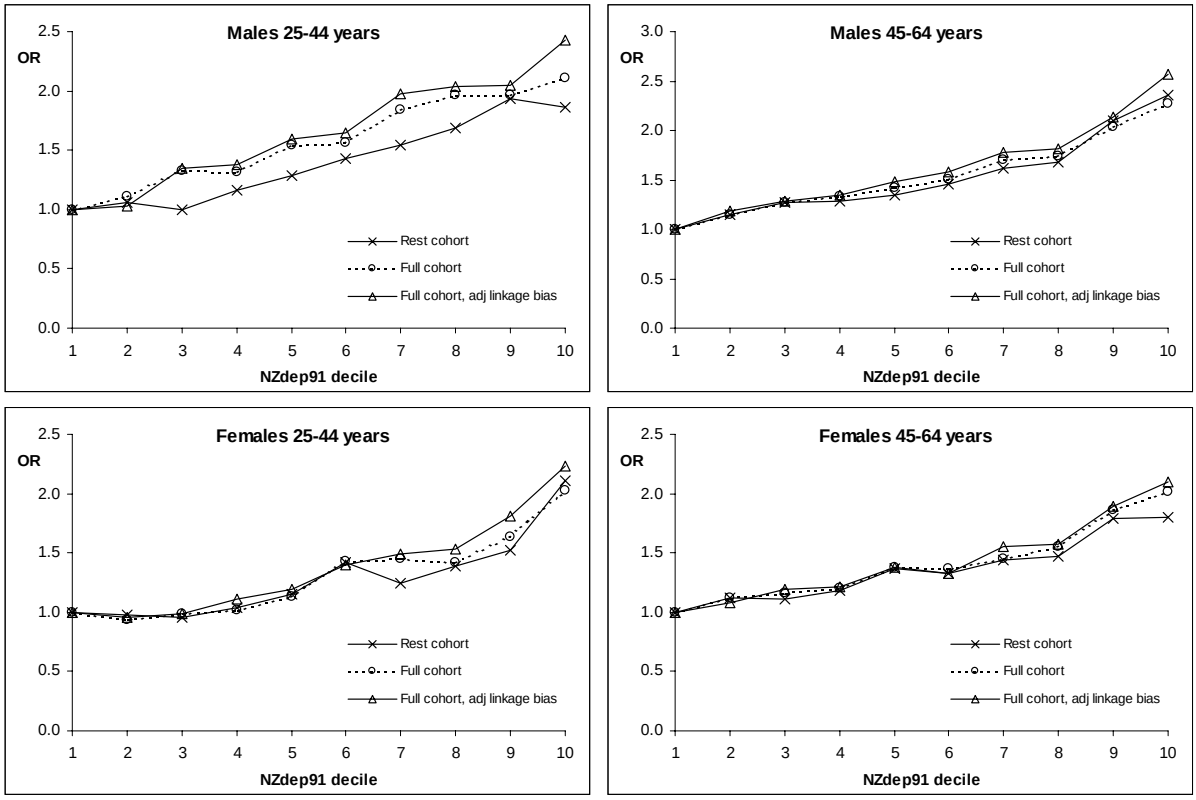
The rationale for these sensitivity analyses was outlined in Section 4.2.2 of Chapter 2: Methods.

1.1 Selection bias

The majority of analyses in this report were conducted upon the restricted cohort. The results from the NZCMS will be generalised to the total New Zealand population – essentially all those people completing the census. Therefore, all that was required to test for selection bias was to compare the association of socioeconomic position with mortality for the 25–64 year olds in the full cohort with that for the remaining 79.5% that remain in the restricted cohort. The reason for conducting these sensitivity analyses of selection bias was *not* primarily to derive the 'true' association of small area deprivation with mortality – that could be done more directly by simply reporting the results for the full census cohort. Rather, the reason was to estimate the magnitude of selection bias for other socioeconomic factors for which there were missing data, most notably equivalised household income.

Of the full census cohort aged 25–64 years (n=1,654,314), 99.0% had an assigned NZDep91 score (n=1,637,523). Of the restricted census cohort (n=1,315,932), 99.9% had an assigned NZDep91 score (n=1,314,852). Thus, the 'full' and 'restricted' cohort results for NZDep91 are highly representative of the true (and slightly larger) full and restricted cohorts.

Figure 27: Comparison of all-cause mortality gradients by NZDep91 deciles between: a) the restricted cohort, b) the full cohort (ie, adjusting for selection bias), and c) the full cohort adjusted for linkage bias



1.1.1 All-cause mortality

Three different all-cause mortality gradients by decile of NZDep91 are presented in Figure 27, for each of the four sex by age groups:

(a) the gradient among the restricted cohort (lines marked with crosses)

(b) the gradient among the full cohort (open circles)

(c) the gradient among the full cohort adjusted for linkage bias (open triangles).

The latter gradient will be considered in the following section on sensitivity analysis for linkage bias. The former two gradients allow an assessment of selection bias – the object of this section.

The restricted and full cohort line graphs for each of the four by sex age groups in Figure 27 demonstrated high concordance for 45–64 year old males, suggesting no substantial selection bias. For 45–64 year old females, the lines only notably diverged at decile 10, where the restricted cohort underestimated the full cohort odds ratio. A likely reason for the variation occurring only among the most deprived deciles is that 70.5% of the full cohort living in decile 10 remained in the restricted cohort, compared to 84.8% of those living in decile 1. Thus, there was a greater possibility of selection bias among the most deprived deciles. Among 25–44 year olds, the lines were unstable with a tendency for the restricted cohort to underestimate the full cohort gradient.

Table 54 attempts to quantify the percentage increase between the restricted and full cohort for the odds ratio comparisons of the least and most deprived deciles. The percentage increase was calculated by first determining the odds ratios for [decile 10 compared to (c.f.) decile 1], [decile 10 c.f. decile 2], [decile 9 c.f. decile 1], and [decile 9 c.f. decile 2], among the restricted and full cohort (and the full cohort adjusted for linkage bias to be described in the subsequent section). The reason for not just reporting the change in the decile 10 compared to decile 1 odds ratio was that it was unstable – particularly with decile 10 often shifting quite markedly. For each of these four odds ratios, in each of the four sex by age groups, the percentage increase in the *excess* odds ratio from the restricted to full cohort was calculated. (As a relative risk or odds ratio of 1.0 is a null finding, the actual relative effect size is given by the relative risk minus 1.0, the so-called 'excess relative risk' or, here, the 'excess odds ratio' (Rothman and Greenland 1998). For example, a change in the odds ratio from 2.0 to 2.1 corresponds to a 10% increase in the excess odds ratio.) Shown in the first row of Table 54 is the average of these four percentage increases in the excess odds ratio for each sex by age group. The results in the first row of Table 54 suggest that the gradient increased when moving from the restricted to the full cohort by 11% for 25–44 year old males and decreased by 7% for 45–64 year old males. Among females the gradient increased by 12% and 19% for 25–44 and 45–64 year olds. Put another way, these percentage increases correspond with 10%, -8%, 11%, and 16% *under*estimates of the 'true' mortality difference due to selection bias for analyses based on the restricted cohort, respectively (eg, for 45–64 year old females 16% = 100 × (1 – [1/1.19])).

Table 54: Percentage increase in the average excess odds ratio† for: a) adjusting for selection bias, b) adjusting for linkage bias, and c) adjusting for both selection and linkage bias

Cohort and adjustment comparison		Males		Females	
		25–44 years	45–64 years	25–44 years	45–64 years
a)	Restricted cohort to full cohort (ie, adjusting for selection bias)	11%	–7%	12%	19%
b)	Full cohort to full cohort adjusted for linkage bias (ie, adjusting for linkage bias)	29%	14%	19%	11%
c)	Restricted to full cohort adjusted for linkage bias (ie, both selection and linkage bias)	45%	6%	34%	33%

† The percentage increase is the average increase for four excess odds ratios – decile 10 compared to decile 1; decile 10 compared to decile 2; decile 9 compared to decile 1; d decile 9 compared to decile 2. As such, it approximates the percentage increase for the quintile 5 compared to quintile 1 comparison.

Box 6: Summarising the effect of selection bias on all-cause mortality gradients by NZDep91

- There was little selection bias across the majority of NZDep91 deciles (Figure 27), for each of the four sex by age groups.
- There was some selection bias affecting the odds ratio comparisons of the most deprived to the least deprived deciles (particularly the decile 10 to decile 1 comparison). Thus, analyses on the restricted cohort *over*estimated the full cohort difference between the most and least deprived by 8% among 45–64 year old males and *under*estimated it by between 10% and 16% for the three remaining sex by age groups.

1.1.2 Cause-specific mortality

Table 55 presents further sensitivity analyses of possible selection bias by quintile of small area deprivation for four broad causes of death (cancer, cardiovascular disease, unintentional injury, and suicide), for 25–64 year olds combined. There was no substantial or consistent selection bias among males for cancer, cardiovascular disease and suicide deaths when considering all the quintiles of deprivation. For male unintentional injury deaths the odds ratio comparing quintile 5 to 1 was 1.67 for the restricted cohort, but only 1.35 for the full cohort – a percentage decrease of 48% in the excess odds ratio. Put another way, the restricted cohort overestimated the injury excess odds ratio by 92% due to selection bias (ie, $100 \times [1.67 - 1.35]/[1.35 - 1.0]$). However, the relative comparisons of the middle three deprivation deciles were not different between the full and restricted cohort. Thus, the apparent selection bias arose due to relative shifts in the mortality risk among the least and most deprived quintiles only. Breaking male unintentional injury into road traffic crash (RTC) and non-RTC deaths (results not presented) demonstrated that two changes between the full and restricted cohort were driving this selection bias:

- compared to the full cohort, the restricted cohort notably underestimated the RTC mortality risk among quintile 1
- compared to the full cohort, the restricted cohort notably overestimated the non-RTC mortality risk among quintile 5.

Among females, the cancer mortality gradient was underestimated for three out of the four non-reference quintile odds ratios in the restricted cohort relative to the full cohort (Table 55). However, the cancer association was not particularly strong in the first place, being an odds ratio of 1.43 and 1.28 for quintile 5 compared to quintile 1 among the full and restricted cohort, respectively. For cardiovascular disease, each of the restricted cohort odds ratios *over*estimated the full cohort odds ratio – but this was entirely due to a change in the relative position of the quintile 1 risk. Thus, there was little selection bias affecting the female cardiovascular disease gradient considered in its entirety. The results for female unintentional injury and suicide should be treated with caution due to smaller numbers, but suggest no substantive selection bias.

Table 55: Comparison of cause-specific odds ratios of mortality by small area deprivation for the restricted cohort versus the full census cohort, ages 25–64 years combined – a test of possible selection bias

	Cohort	Odds ratio (reference group quintile 1)					% change Quintile 5 OR†
		Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	
Males							
Cancer	Restricted	1.00	1.03	1.16	1.21	1.51	
	Full	1.00	1.08	1.21	1.28	1.47	–7%
CVD	Restricted	1.00	1.24	1.27	1.72	2.28	
	Full	1.00	1.27	1.33	1.74	2.21	–6%
Injury	Restricted	1.00	1.16	1.48	1.54	1.67	
	Full	1.00	1.04	1.25	1.36	1.35	–48%
Suicide	Restricted	1.00	1.41	1.01	1.71	2.04	
	Full	1.00	1.45	1.20	1.80	1.97	–7%
Females							
Cancer	Restricted	1.00	1.01	1.17	1.16	1.28	
	Full	1.00	1.05	1.17	1.24	1.43	50%
CVD	Restricted	1.00	1.44	1.65	1.86	2.69	
	Full	1.00	1.34	1.58	1.80	2.55	–8%
Injury	Restricted	1.00*	1.43*	1.69	1.19*	2.26	
	Full	1.00*	1.31*	1.71	1.38*	2.55	23%
Suicide	Restricted	1.00*	1.20*	1.48*	1.48*	2.07*	
	Full	1.00*	1.18*	1.60*	1.31*	2.01*	–6%

Note: Calculations are based on data random rounded to a multiple of three.

† The percentage change is that for the excess odds ratio, ie, the percentage change for [OR minus 1.0].

* Less than 30 deaths in the cell.

Box 7: Summarising the effect of selection bias on cause-specific mortality gradients by NZDep91

- There was notable selection bias for male unintentional injury deaths, such that analyses on the restricted cohort *overestimated* the NZDep91 comparison of the most to least deprived quintiles by 92%.
- There was some selection bias for female cancer deaths, such that analyses on the restricted cohort tended to *underestimate* the NZDep91 gradient by about a third – but as this cancer gradient was small in the first place, the selection bias was not particularly consequential.
- There was no evidence of substantial selection bias for the other broad causes of death by sex.

As with all-cause mortality, the implication of these sensitivity analyses for analyses of cause-specific mortality gradients by other socioeconomic factors (eg, household income) will be presented following sensitivity analyses of selection bias for education (Section 2.1 of this Appendix).

1.2 Linkage bias

1.2.1 All-cause mortality

As documented in Chapter 3 there was some linkage bias by NZDep91 and occupational class, such that inequalities in mortality by socioeconomic factors were *underestimated* due to linkage bias – particularly for the lowest compared to highest socioeconomic groups. Also plotted in Figure 27 are the line graphs for the all-cause mortality gradient among the full cohort adjusted for linkage bias (open triangles), using the linkage bias results from Chapter 3 (Table 20). The shift in this line graph compared to the full cohort line graph (open circles) represents the effect of linkage bias, and compared to the restricted cohort line (crosses) represents the net effect of both selection and linkage bias acting on the restricted cohort. Looking at the three line graphs Figure 27 for each of the four sex by age groups, first note that the *line graphs are generally similar*. However, if expressed in excess risk (or odds) ratio terms and for the extreme comparisons of the most deprived with the least deprived, then the linkage bias (and net effect of linkage and selection biases) becomes more than just trivial. The middle row of Table 54 presents the percentage increase from the full cohort to the full cohort adjusted for linkage bias for the excess odds ratios of deciles 9 and 10 compared to decile 1 and 2. There were 14% and 11% increases for 45–64 year old males and females, respectively, and 29% and 19% increases for 25–44 year old males and females. Considering the net impact of selection bias and linkage bias, the final row of Table 54 demonstrates that they largely offset each other for 45–64 year old males, but compounded each other for the three other sex by age groups. Regarding this latter compounding, the percentage increases was 45% for 25–44 year old males from the restricted cohort to the full cohort adjusted for linkage bias, and 34% and 33% for 25–44 and 45–64 year old females. Put another way, analyses on the restricted cohort *underestimated* the 'true' comparison of most and least deprived deciles due to the cumulative effect of linkage and selection biases by 31% for 25–44 year old males, and 26% and 25% for 25–44 and 45–64 year old females, respectively.

Box 8: Summarising the net impact of selection and linkage biases on all-cause mortality gradients by NZDep91

- The net effect of selection and linkage biases was relatively modest for the mid-decile comparisons, but was more notable comparing the least and most deprived deciles.
- Linkage bias and selection bias tended to off-set each other for males aged 45–64 years.
- Among females aged 25–44 and 45–64 years and males aged 25–44 years the selection and linkage biases compounded each other such that analyses on the restricted cohort *underestimated* the excess odds ratio comparisons of the most and least deprived deciles by about 25%.

1.2.2 Cause-specific mortality

As with the all-cause mortality gradients, there was linkage bias that affected the association of cause-specific mortality with small area deprivation. Using the log-linear regression results for linkage bias by small area deprivation in Table 21 the mortality odds ratios shown above in Table 28 and Table 29 for cancer, cardiovascular disease, injury and suicide were adjusted for linkage bias. Results are shown 212 in Table 56. (Note that the results in Table 56 represent the effect of linkage bias only on the restricted cohort, not the net effect of selection and linkage biases.) There was a considerable percentage increase in the excess odds ratio for quintile 5 compared to quintile 1 for each cause of death among males. Or, put the other way, linkage bias causes an *under*estimation of the mortality gradient for all four causes of death among males. Among females, there was little linkage bias for cancer and cardiovascular disease, and there were too few deaths for a robust sensitivity analysis of injury and suicide deaths.

Table 56: Comparison of cause-specific age and ethnicity adjusted odds ratios of mortality by small area deprivation, with and without adjustment for linkage bias, for 25–64 year old males and females among the restricted cohort

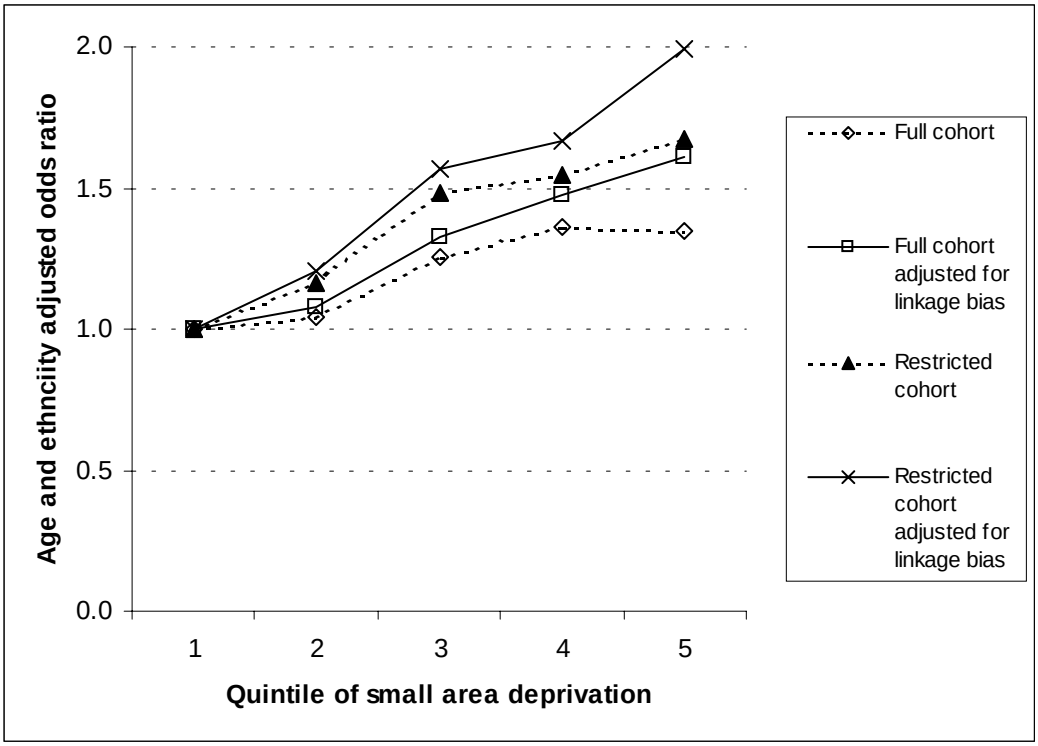
		NZDep91 quintile					% change Quintile 5 OR†
		1	2	3	4	5	
Males							
Cancer	Age/ethnicity adjusted	1.00	1.03	1.16	1.21	1.51	
	Plus linkage bias adjusted	1.00	1.03	1.21	1.22	1.62	23%
CVD	Age/ethnicity adjusted	1.00	1.24	1.27	1.72	2.28	
	Plus linkage bias adjusted	1.00	1.25	1.33	1.79	2.46	14%
Injury	Age/ethnicity adjusted	1.00	1.16	1.48	1.54	1.67	
	Plus linkage bias adjusted	1.00	1.20	1.56	1.67	1.99	48%
Suicide	Age/ethnicity adjusted	1.00	1.41	1.01	1.71	2.04	
	Plus linkage bias adjusted	1.00	1.43	1.37	1.88	2.43	37%
Females							
Cancer	Age/ethnicity adjusted	1.00	1.01	1.17	1.16	1.28	
	Plus linkage bias adjusted	1.00	1.02	1.18	1.19	1.30	4%
CVD	Age/ethnicity adjusted	1.00	1.44	1.65	1.86	2.69	
	Plus linkage bias adjusted	1.00	1.46	1.66	2.01	2.87	11%

Note: Age and ethnicity adjusted odds ratios are taken from Tables 21 and 22. The odds ratios further adjusted for linkage bias are calculated by dividing the age and ethnicity odds ratios by the risk ratio for the linkage bias by NZDep91 shown in Table 14.

† The percentage change is that for the excess odds ratio, ie, the percentage change for [OR minus 1.0].

A substantial selection bias was described for male unintentional injury deaths in the previous section such that the restricted cohort *over*estimated the gradient compared to the full cohort. In Table 56, however, a substantial linkage bias is shown such that the restricted cohort *under*estimated the gradient that would have been observed if there was no selection bias. Thus, these two biases off-set each other as shown in Figure 28. The lower line (open diamonds) represents the odds ratios among the full cohort, ie, unaffected by selection bias. The next line-up (square boxes) represents the full cohort odds ratios adjusted for linkage bias (ie, adjusting for both linkage bias and selection bias simultaneously). Note that this second line is similar to the restricted cohort line (solid triangles). Thus, the simple age and ethnicity adjusted odds ratios among the restricted cohort just so happen to be similar to the odds ratios that would be observed after adjusting for both selection and linkage biases. The top line (crosses) shows the restricted cohort odds ratios adjusted for linkage bias (but not for selection bias) – it overestimates the association of small area deprivation with unintentional injury mortality.

Figure 28: Net effect of adjusting for both selection bias and linkage bias for 25–64 year old male unintentional injury deaths by NZDep91 quintile



For other causes of death among males, the lack of any substantive and consistent selection bias described above meant that linkage bias was not off-set. Thus, the *net* effect of selection bias (Table 55) and linkage bias (Table 56 above) for male cancer, cardiovascular disease and suicide deaths will be broadly similar to that for linkage bias alone shown in Table 56 above.

Among female cardiovascular disease deaths the net effect of selection bias (Table 56) and linkage bias (Table 56 above) was negligible. However, in excess odds ratio terms, the cancer mortality gradient underestimate of about a third due to selection bias was not offset by any linkage bias – but the association of NZDep91 and cancer among females was modest to start with.

Box 9: Summarising the net impact of selection and linkage biases on cause-specific mortality gradients by NZDep91

For analyses on the restricted cohort for *males* aged 25–64 years:

- Linkage and selection biases off-set each other for *injury* deaths in the restricted cohort
- For other causes of death the lack of selection bias meant that the net impact of the two biases was simply approximated by the linkage bias, such that analyses on the restricted cohort *underestimated* the 'true' gradient by:
  - approximately 10% for *cancer* and *cardiovascular disease* deaths
  - approximately 25% for *suicide* deaths.

For analyses on the restricted cohort for *females* aged 25–64 years:

- The selection bias for female *cancer* deaths was neither compounded nor off-set by any linkage bias, such that the net effect was an approximately 30% *underestimate* of the 'true' gradient for analyses on the restricted cohort – but the cancer gradient was modest to start with.
- The net effect of linkage and selection biases for *cardiovascular disease* was negligible.
- The net effect for female *suicide* and *injury* deaths was unable to be robustly determined, but presumably the gradients by non-cancer and non-cardiovascular disease on average tended to be *underestimated* given the approximately 25% underestimate of the all-cause mortality described above in Box 9.

1.3 Health selection

1.3.1 Observed mortality risk over time

All-cause mortality, all labour force categories

Figure 29 below shows the observed mortality risk by six-month period over the three-year follow-up, by quintile of small area deprivation. For 45–64 year old males and females, the plots are roughly parallel. Among 25–44 year olds, the plots are unstable due to small numbers. These parallel lines are consistent with the theoretical expectation of no drift health selection for small area deprivation, and acts as a baseline for mortality plots by income presented later in this report.

Cancer and cardiovascular disease deaths among 45–64 year olds: all labour force categories, and excluding non-active labour force

It is likely that health selection, if acting, varies by cause of death. There were enough deaths by broad level of socioeconomic factor for cancer deaths among 45–64 year olds (both sexes) and cardiovascular disease among males 45–64 year olds to plot mortality risks by six-month period for all labour force categories *and* excluding the non-active labour force.

Figure 29: Mortality risk by six-month period following census night by quintile of small area deprivation, full cohort and all labour force categories

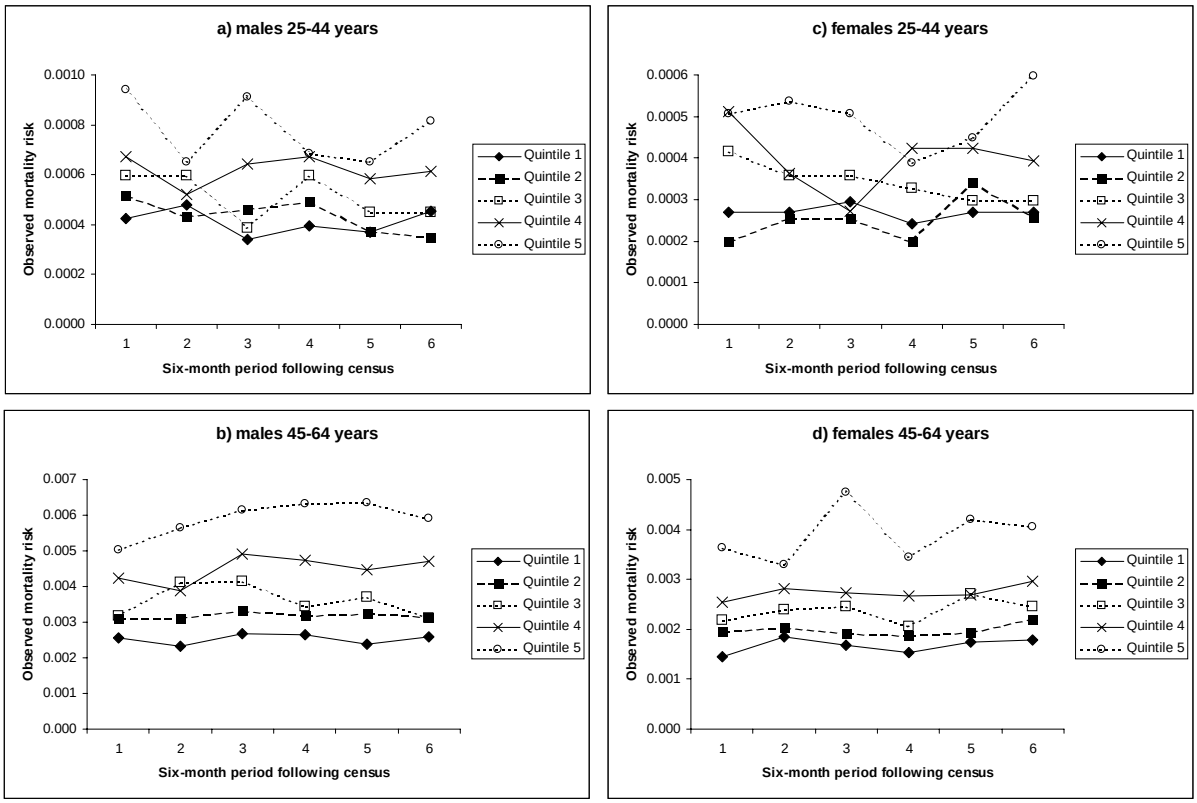
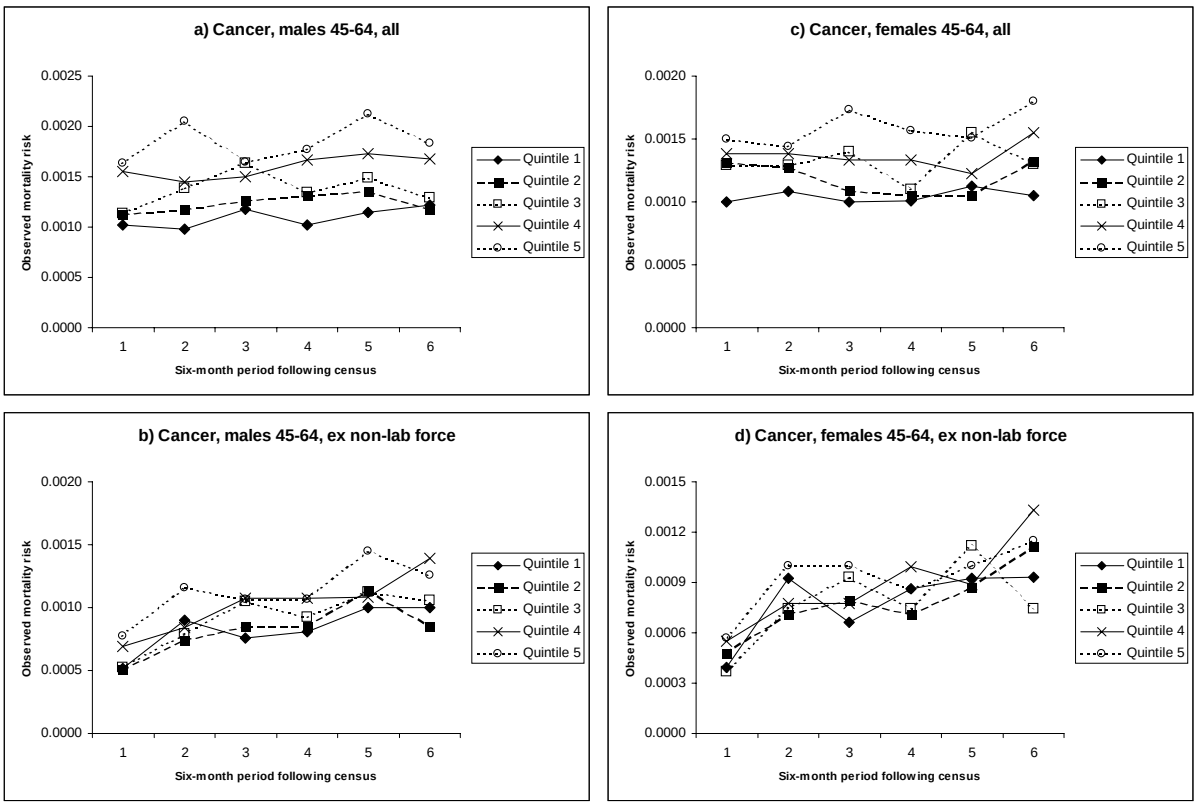


Figure 30 (below) shows the cancer mortality risks over time among 45–64 year olds, both for all labour force categories and excluding the non-active labour force (ie, including only the employed and unemployed). For both males and females the mortality risk lines are approximately horizontal and parallel for all labour force categories combined (ie, Figures a and c). Thus, there is no evidence of *drift* health selection, as would be expected given that one's usual residence (and hence assigned decile of small area deprivation) is probably not substantively influenced by recent poor health. As with the all-cause mortality plots by NZDep91 in Figure 29, Figure 30a and Figure 30c serve as benchmarks for the later plots by income where drift health selection is theoretically plausible.

Figure 30: Mortality risk by six-month period following census night by small area deprivation, for 45–64 year old cancer deaths, for all labour force categories and excluding the non-active labour force



Excluding the non-active labour force, all mortality risk lines by quintile of NZDep91 are sloping upwards (Figure 30b and Figure 30d). Thus, people who die of cancer soon after census night are less likely to be in the active labour force on census night than those that die of cancer up to three years after census night. This is not surprising given that death from cancer is often preceded by a period of deteriorating health. If there was health selection out of the active labour force that was *differential* by level of deprivation (ie, if you had cancer, you were more likely to exit the active work force if you lived in a more deprived small area than if you lived in a less deprived area), then we would expect the mortality risk line for NZDep91 quintile 5 to be steeper than NZDep91 quintile 1. While there is imprecision about each point in Figure 30b and Figure 30d (eg, the 95% confidence intervals for each mortality risk 'point' were approximately plus or minus 0.0002 in Figure 30b), there was not convincing evidence that the slopes differed. (Note that this 'test' is only one test of differential health selection and may lack power. Similar plots by highest qualification and income presented latter in this report were suggestive of differential health selection among cancer deaths. Thus, a 'best' decision about differential health selection in the NZCMS requires a balanced consideration across further tests presented subsequently in this report.)

The pattern of mortality risk plots for 45–64 year old male cardiovascular disease deaths (all labour force categories) was similar to that for cancer. Cardiovascular deaths among 45–64 year olds in the active labour force were too few to allow a robust interpretation.

1.3.2 Excluding sickness beneficiaries

Included in the 1991 census data is whether each individual had been a recipient of a sickness benefit in the preceding 12 months. Excluding these people from the analyses would limit the cohort to a 'healthier' subgroup, reducing the likely effect of *drift* health selection. Assuming that the level of deprivation of where you live is not affected by (recent) poor health, excluding unhealthy people from the analysis should not alter the mortality gradient by NZDep91. However, the sickness benefit is means tested in New Zealand – among unhealthy people only those of lower incomes and socioeconomic position will actually receive a sickness benefit. Therefore, analyses excluding sickness beneficiaries will exclude a greater percentage of the unhealthy people in the deprived areas than the non-deprived areas, which theoretically should *reduce* the observed mortality gradient by stable socioeconomic factors such as NZDep91. In order to use the exclusion of sickness beneficiaries as a test of health selection effects for other socioeconomic factors such as income, we must first quantify the amount of this spurious reduction in the deprivation (and education) mortality gradient due to excluding sickness beneficiaries. For subsequent analyses of income, analyses excluding sickness beneficiaries will be suggestive of health selection effects only if the reduction in the income mortality gradient is substantially more than that observed for the NZDep91 and education mortality gradients.

Table 57 shows the age and ethnicity-adjusted odds ratios by small area deprivation for all cause mortality, for all members of the restricted cohort and following various exclusions. (The interpretation of analyses excluding pre-hospitalised deaths and the non-active labour force follows in points 1.3.3 and 1.4 respectively.) The percentage reduction for the quintile 5 compared to quintile 1 excess odds ratio after excluding sickness beneficiaries ranged between 9% and 18% for the four sex by age groups. The reductions in the mortality gradient for broad causes of death (Figures 31 and 32) were likewise modest only, although greater for male cancer and cardiovascular disease deaths compared to male injury and suicide deaths. The lack of reduction following the exclusion of sickness beneficiaries for injury deaths was probably due to only 4% of injury-decedents having been a sickness beneficiary in the preceding year compared to 10% of all deaths. Nine percent of suicide decedents had received a sickness benefit in the preceding year.

These analyses will be used as a baseline for the investigation of health selection bias affecting the household income mortality gradients.

Table 57: Odds ratios (95% CI) of all cause mortality for 25–64 year olds in the restricted cohort, by quintile of small area deprivation, for various exclusions testing for health selection

Exclusion criteria	Odds ratios (ref group Quintile 1)					% change to null of Quintile 5 OR†
	1	2	3	4	5	
Males						
25–44 year olds						
Nil	1.00	1.05 (0.85–1.30)	1.32 (1.07–1.62)	1.56 (1.28–1.91)	1.85 (1.50–2.27)	
Sickness beneficiaries	1.00	1.00 (0.80–1.25)	1.27 (1.02–1.57)	1.43 (1.16–1.77)	1.72 (1.39–2.14)	14%
Pre-hospitalised deaths	1.00	0.98 (0.77–1.25)	1.14 (0.90–1.45)	1.29 (1.02–1.63)	1.41 (1.11–1.80)	51%
Non-active labour force	1.00	1.06 (0.85–1.33)	1.32 (1.06–1.64)	1.45 (1.16–1.80)	1.73 (1.38–2.17)	14%
45–64 year olds						
Nil	1.00	1.20 (1.09–1.32)	1.31 (1.19–1.45)	1.54 (1.40–1.70)	2.07 (1.88–2.27)	
Sickness beneficiaries	1.00	1.19 (1.08–1.31)	1.29 (1.17–1.43)	1.49 (1.35–1.64)	1.90 (1.72–2.09)	16%
Pre-hospitalised deaths	1.00	1.20 (1.07–1.35)	1.25 (1.11–1.41)	1.47 (1.30–1.65)	1.81 (1.61–2.04)	24%
Non-active labour force	1.00	1.11 (0.98–1.25)	1.22 (1.07–1.39)	1.39 (1.22–1.58)	1.61 (1.40–1.84)	43%
Females						
25–44 year olds						
Nil	1.00	1.01 (0.78–1.31)	1.30 (1.01–1.66)	1.33 (1.03–1.71)	1.81 (1.41–2.32)	
Sickness beneficiaries	1.00	0.97 (0.74–1.27)	1.30 (1.01–1.68)	1.25 (0.96–1.63)	1.67 (1.29–2.16)	18%
Pre-hospitalised deaths	1.00	0.87 (0.60–1.27)	1.66 (1.20–2.31)	1.19 (0.83–1.72)	1.67 (1.17–2.38)	17%
Non-active labour force	1.00	1.08 (0.79–1.48)	1.44 (1.06–1.95)	1.39 (1.01–1.90)	1.55 (1.11–2.17)	32%
45–64 year olds						
Nil	1.00	1.08 (0.96–1.22)	1.27 (1.13–1.43)	1.37 (1.22–1.54)	1.70 (1.51–1.91)	
Sickness beneficiaries	1.00	1.08 (0.95–1.22)	1.27 (1.12–1.42)	1.34 (1.19–1.50)	1.63 (1.45–1.84)	9%
Pre-hospitalised deaths	1.00	1.01 (0.86–1.18)	1.34 (1.15–1.56)	1.28 (1.09–1.49)	1.67 (1.43–1.94)	4%
Non-active labour force	1.00	0.99 (0.82–1.20)	1.03 (0.85–1.25)	1.16 (0.95–1.42)	1.33 (1.07–1.64)	53%

Note: Injury and suicide deaths are not presented for females due to small numbers. For all remaining cells in the table there were at least 30 deaths.

† Percentage change is for the excess odds ratio for quintile 5 compared to quintile 1, compared to the same odds ratio with nil exclusions.

Figure 31: Crude risk ratios of cause-specific mortality for 25–64 year old males in the restricted cohort, by quintile of small area deprivation, for various exclusions testing for possible health selection

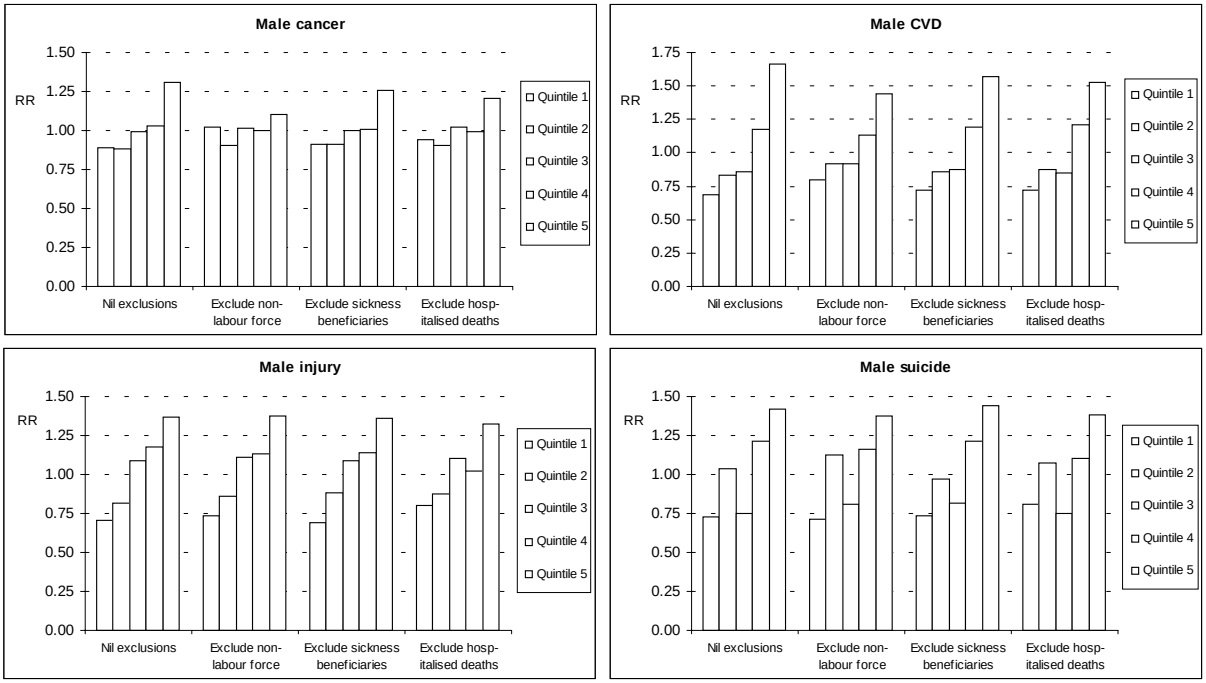
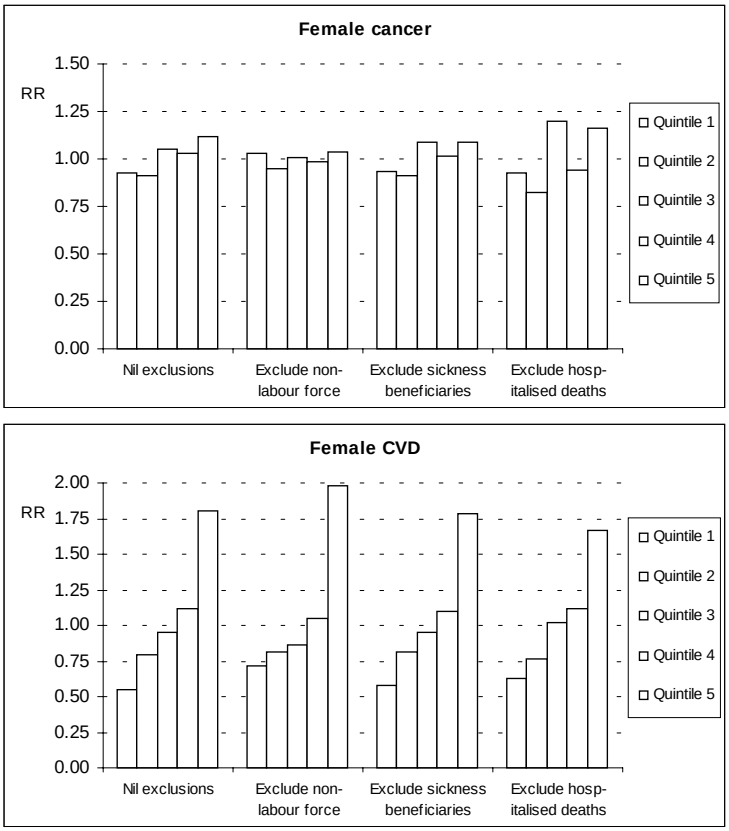


Figure 32: Risk ratios of cancer and cardiovascular mortality for 25–64 year old females in the restricted cohort, by quintile of small area deprivation, for various exclusions testing for possible health selection



1.3.3 Excluding decedents with a hospitalisation event between 1988 and census night

The baseline analyses for excluding pre-hospitalised deaths are also presented in Table 57 for all-cause mortality, and Figures 31 and 32 for broad causes of death. The percentage reductions for the quintile 5 compared to quintile 1 excess odds ratios for all-cause mortality were 51% and 24% for 25–44 and 45–64 year olds, respectively. The percentage reductions for female all-cause mortality were less – 17% and 4% respectively. However, the gradients after excluding pre-hospitalised deaths were somewhat unstable due to the large number of excluded deaths. By cause of death, the largest reduction in the deprivation mortality gradient following the exclusion of pre-hospitalised deaths tended to be for cancer deaths.

Assuming that deprivation mortality gradients are not susceptible to drift health selection, the reduction in these gradients following the exclusion of pre-hospitalised deaths was surprisingly high. For this to occur, the chance of being hospitalised prior to one's death (be it for the subsequent cause of death or an unrelated event) must be associated with deprivation independently of the association of deprivation with mortality. Whatever the reason, for exclusion of sickness beneficiaries to suggest drift health selection affecting the income–mortality gradient, the percentage reduction will have to be substantially greater than that found here for the baseline NZDep91 analysis.

1.4 Excluding the non-active labour force

Controlling socioeconomic mortality gradients for labour force status is difficult to interpret. Reasons include labour force status being a proxy for health status and variables other than health status, and being a proxy for both confounding and intermediary variables (see page 62 for previous discussion).

1.4.1 All-cause mortality

Table 57 shows the odds ratios of mortality by NZDep91 quintile for all labour force categories, and excluding the non-active labour force, using the quintile 5 to quintile 1 odds ratio, excluding the non-labour force results in a 14% and 43% reduction in the excess odds ratio for 25–44 and 45–64 year old males, respectively. Among females the comparable reductions are 32% and 53%, respectively. (Note that the 32% reduction for 25–44 year old females was only apparent for the quintile 5 to quintile 1 odds ratio – there was little change for the quintiles 2 to 4 odds ratios.) As discussed previously in this report, it has been argued in the international literature that a similar reduction for education mortality gradients following exclusion of the non-active labour force was a consequence of differential health selection. However, the mortality risk plots over time above (Section 1.3.1, page 209-210) *do not suggest* health selection out of the active labour force that is *differential* by level of deprivation for cancer deaths or cardiovascular disease deaths. It may be that the 'test' for differential health selection by plotting mortality risks over time lacks power – but it seems unlikely that the mortality risk plots over time would have failed to detect differential health selection that caused up to a 50% reduction in the NZDep91 gradient. Thus the reduction in the NZDep91 mortality gradient following the exclusion of the non-active labour force must also include components of one or more of:

- *Confounding by labour force status* by means other than short-term drift health selection. For example, personality characteristics and behaviours associated with being in the non-labour force may be associated with the relative deprivation of the neighbourhood you live in, and (independently of deprivation) associated with mortality risk.
- *Mediation by labour force status*, or factors it is a proxy for (most notably health status).
  - For example, where you live (and hence small area deprivation) may affect your employment opportunities, such that labour force status acts as an intermediary variable between deprivation and mortality risk (rather than being a confounder as in the above bullet point).
  - Regarding labour force status as a proxy for health status, apart from injury deaths poor health is usually an intermediary variable between socioeconomic factors and mortality. As poor health often results in movement into the non-active labour force, controlling for labour force status may 'over-control' the deprivation-mortality association. (Note that this mechanism is not health selection – poor health is assumed not to influence where you live in the short-term, and mortality risk plots (Section 1.3.1.2) did not suggest health selection out of active labour force that was differential by deprivation.)
- *Effect modification of the deprivation mortality gradient by labour force status* by mechanisms other than differential health selection, such that the gradient is weaker among the active labour force. Because of the strong association of labour force status with mortality risk, any decision about effect modification between labour force status and deprivation was difficult. Analyses (not shown) found that the *relative* deprivation gradient was weaker in the non-labour force than the active labour force, but that the *absolute* gradient was stronger. Thus, it was unclear whether there was truly interaction between labour force status and deprivation.

Putting aside the latter effect modification, one possible way to tease apart the relative contributions of confounding/mediation and differential health selection is to look at the effect of excluding the non-active labour force for specific causes of death – the subject of the next section.

1.4.2 Cause-specific mortality

Regarding the teasing apart of differential health selection from confounding or mediation, one might expect that deaths from cancer (and possibly cardiovascular disease) were vulnerable to differential health selection. Conversely, one might expect that deaths from injury were not associated with a period of poor health prior to death, and therefore not vulnerable to differential health selection. If we also make the assumptions that:

1 Any confounding/mediation of the association of deprivation with mortality by labour force status did not vary greatly by cause of death.

2 Any effect modification of the association of deprivation with mortality by mechanisms other than differential health selection did not vary greatly by cause of death.

3 Drift health selection does not affect the association of deprivation with mortality.

Then:

- If excluding the non-labour force resulted in *equivalent* reductions in the mortality gradients for each of cancer, cardiovascular disease, unintentional injury, and suicide deaths, this would provide evidence *against* a prominent role for differential health selection (and probably point to a major role of confounding/ mediation as an explanation of reduced deprivation-mortality gradients among the active-labour force)
- If excluding the non-labour force resulted in *greater* reductions in the mortality gradients for cancer and cardiovascular disease deaths (compared to unintentional injury and suicide deaths), this would *support* a prominent role for differential health selection as a reason for the reduced gradient among the labour force.

(This test is subject to many assumptions. Perhaps the most important and questionable assumption is number 1 above – the robustness of this assumption will be returned to after presenting the 'test' results.)

Figure 31 (page 213) presents the mortality gradient by NZDep91 quintiles among 25–64 year old males for four broad causes of death: cancer, cardiovascular disease, unintentional injury, and suicide. Figure 32 shows the same graphs for females, but only for cancer and cardiovascular disease – there were too few unintentional injury and suicide deaths among females for a robust determination of the effect of the various exclusions. Both figures use crude data, with floating risk ratios. As the graphs do not use age and ethnicity adjusted odds ratios, the crude risk ratios may be somewhat confounded. Therefore, it is more important to look for patterns than exact differences.

For males in Figure 31, the mortality gradient is notably reduced following exclusion of the non-labour force for cancer and cardiovascular disease (percentage reductions in excess risk ratios comparing quintile 5 to 1 of 83% and 42%, respectively), but little changed for unintentional injury and suicide (percentage reductions of only 7% and 3%). While for females it was not possible to examine the change for unintentional injury and suicide deaths, the patterns in Figure 32 for cancer and cardiovascular disease are consistent with the pattern for males. The female cancer mortality gradient was essentially flattened (albeit small to start with), and the quintile 5 to quintile 1 excess risk ratio for cardiovascular disease reduced by 23%.

There are at least two important problems with this 'test'. First, the size of the cancer mortality gradient by NZDep91 was small to start with, and a small amount of confounding by labour force status that nullified the cancer association may have only marginally reduced the stronger associations of NZDep91 with cardiovascular disease, injury and suicide. Thus, a constant amount of confounding/mediation by labour force across causes of disease may actually have produced the picture above. The interpretation all depends on whether a 'constant' amount of confounding causes an equivalent change in: the percentage excess odds ratio, or the absolute odds ratio.

Second, the assumption that any confounding/mediation of the deprivation-mortality association by labour force status is 'constant' by cause of death is questionable. For example, labour force status is strongly associated with tobacco smoking in New Zealand (Wilson 2000), and the cancer mortality gradient by deprivation was mainly due to lung cancer. Thus, the amount of confounding/mediation of the deprivation–cancer mortality association by labour force status (as a proxy for smoking in this instance) may have been greater for cancer than the other causes of disease.

Given all these reservations about this test for differential health selection, what can we conclude? At best, *these NZDep91 results (for males at least)* ***suggest*** *the possibility of differential health selection for cancer (and cardiovascular disease) – but no stronger conclusion can be made.* If there was differential health selection for cancer deaths (and cardiovascular disease death), then adjusting for labour force status in multivariate analyses may result in an underestimate of the independent association of small area deprivation (and by extrapolation other socioeconomic factors) with mortality. But it is not possible to specify with precision the effect of the relative contributions of differential health selection, effect modification and confounding/mediation by labour force status on the deprivation-mortality gradients. This issue will be pursued further in following sections.

2 Highest qualification

2.1 Selection bias

As with small area deprivation, most of the full cohort had a specified value for highest qualification. Similar analyses to test for a possible selection bias between the full and restricted cohort for all-cause mortality were conducted for highest qualification (see Figure 27 (page 201) for the comparable NZDep91 analysis). The analyses by highest qualification disclosed no substantial selection bias for all-cause mortality. The age and ethnicity-adjusted odds ratio for those with no qualification compared to those with a graduate/ postgraduate degree varied between the full and restricted cohort by less than 1% for each of 25–44 and 45–64 year old males, and by 3% and 7% for 25–44 and 45–64 year old females.

Table 58 below shows the results for cause-specific mortality. Perhaps, the only notable difference between the full cohort and the restricted cohort was for female cancer where the restricted cohort *over*estimated the gradient compared to the full cohort. Note that the comparable NZDep91 analysis (see Table 55, page 204) demonstrated the opposite – the restricted cohort *under*estimated the association relative to the full cohort. Also note that there was a tendency for the male injury odds ratios to be overestimated (ie, further from the null) in the restricted cohort due to selection bias, but by less than that suggested by the NZDep91 sensitivity analyses (Table 55). The results for female unintentional injury and suicide deaths must be treated with caution due to small numbers of deaths.

Table 58: Comparison of cause-specific odds ratios of mortality by highest qualification for the restricted cohort versus the full census cohort, 25–64 year olds combined – a test of possible selection bias

	Cohort	Age and ethnicity adjusted odds ratio				% change to Null tertiary OR†
		Tertiary	Trade	School	Nil	
Males						
Cancer	Restricted	0.74	0.79	0.84	1.00	
	Full	0.74	0.78	0.81	1.00	–1%
CVD	Restricted	0.57	0.80	0.70	1.00	
	Full	0.56	0.79	0.69	1.00	–1%
Injury	Restricted	0.51	0.75	0.79	1.00	
	Full	0.54	0.77	0.85	1.00	8%
Suicide	Restricted	0.56	0.89	1.06	1.00	
	Full	0.52	0.93	1.05	1.00	–8%
Females						
Cancer	Restricted	0.81	0.88	0.85	1.00	
	Full	0.84	0.87	0.85	1.00	13%
CVD	Restricted	0.49	0.54	0.74	1.00	
	Full	0.48	0.52	0.75	1.00	–2%
Injury	Restricted	0.68*	0.92*	0.83	1.00	
	Full	0.62*	0.79*	0.76	1.00	–19%
Suicide	Restricted	1.18*	0.87*	1.17*	1.00	
	Full	1.33	1.17*	1.34	1.00	80%

† The percentage change is that for the tertiary compared to nil excess odds ratio to the null, ie, the percentage change for the absolute value of [OR minus 1.0]. For example, the female cancer tertiary compared to nil OR changes from 0.81 to 0.84 – a change in the excess odds ratio of –0.19 to –0.16 (negative signs indicate preventative odds ratios). Thus the percentage reduction to the null is $0.03/0.19 \times 100 = 16\%$, or 13% if unrounded data are used.

* Less than 30 deaths in the cell.

Box 10: Summarising the effect of selection bias on mortality gradients by NZDep91 (Box 7) and highest qualification

All-cause mortality:

- There was no substantial evidence of selection bias affecting all-cause mortality gradients by highest qualification. This contrasts with the NZDep91 analyses that suggested a modest tendency for the restricted cohort to underestimate the gradient – except among 45–64 year old males where there was a slight overestimate.

Cause-specific mortality:

- The NZDep91 and highest qualification analyses were consistent in suggesting no notable selection bias affecting:
  - male cardiovascular disease, injury and suicide deaths
  - female cardiovascular deaths.
- Both NZDep91 and highest qualification analyses found that analyses on the restricted cohort overestimated the male injury gradient, particularly the NZDep91 analyses. (However, this selection bias would have probably been offset by linkage bias.)
- NZDep91 analyses suggested a selection bias that underestimated the female cancer gradient, but the highest qualification analyses were basically null (or even suggesting the reverse – an overestimation).
- Analyses for female injury and suicide deaths were not robust enough to allow interpretation.

Given the lack of a consistent pattern between the NZDep91 and education selection bias analyses (other than male injury), it seems difficult to reliably predict what the magnitude and direction of any selection bias might be for other socioeconomic factors such as income.

Figure 33: Mortality risk for each six-month period following census night by highest qualification, all labour force categories

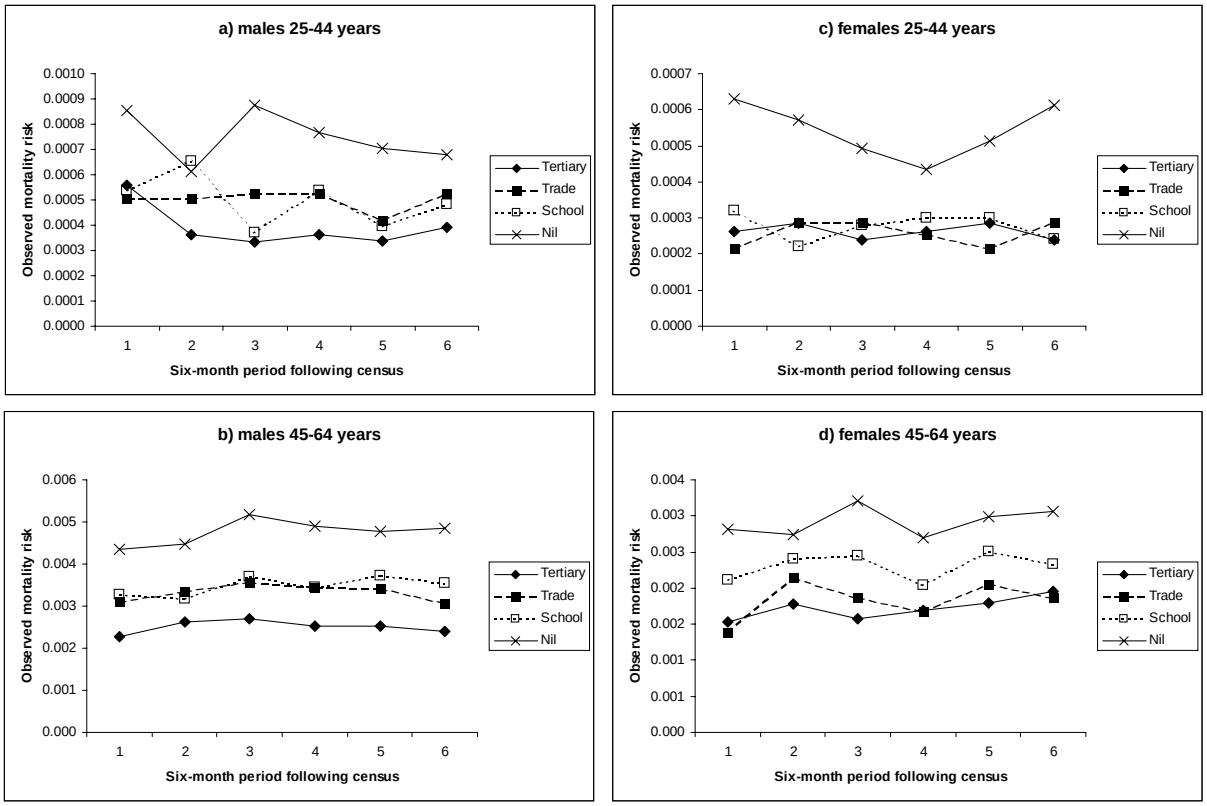
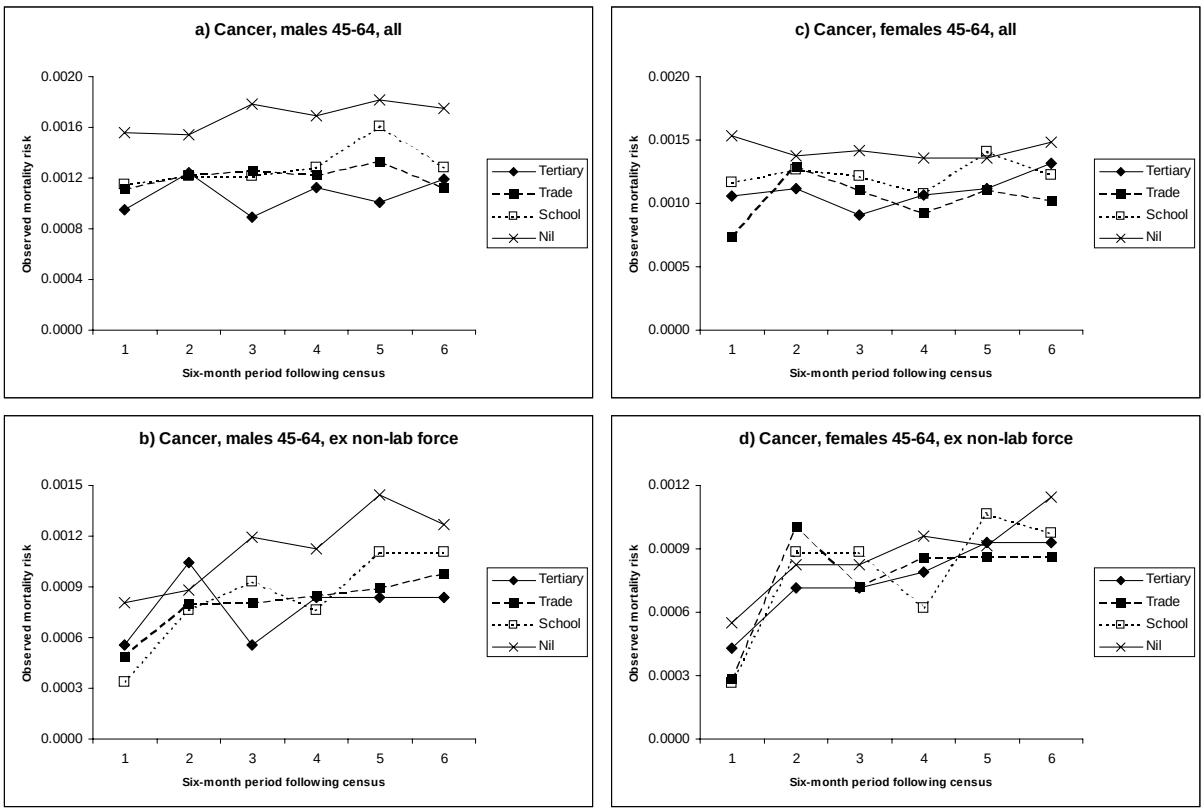


Figure 34: Cancer mortality risk for each six-month period following census night by highest qualification for 45–64 year olds, all labour force categories and excluding the non-active labour force



2.2 Health selection

2.2.1 Observed mortality risk over time

All-cause mortality, all labour force categories

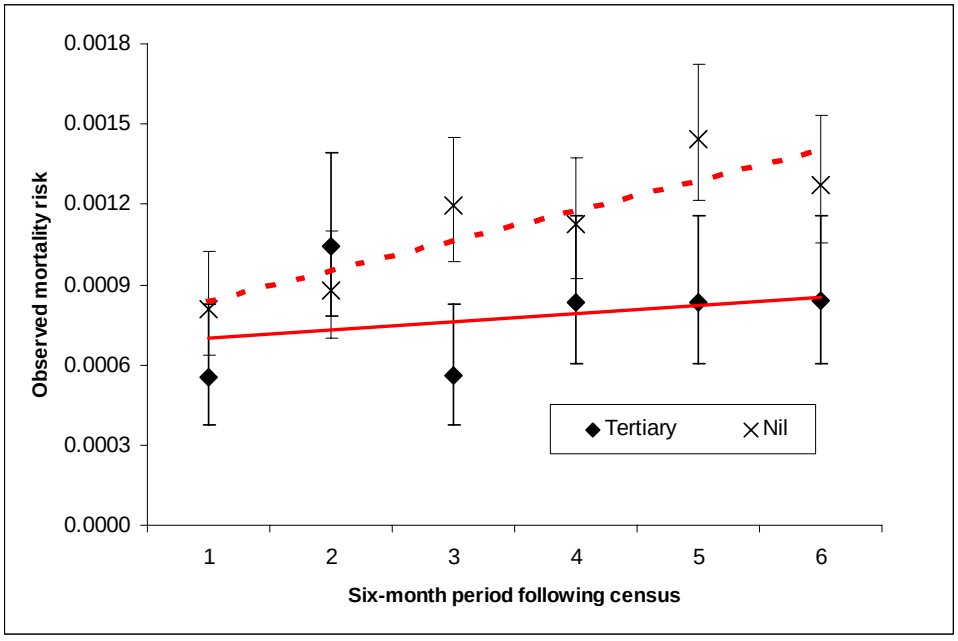
Figure 33 above shows the mortality risk for each six-month period of the three-year follow-up. For later income and labour force analyses using this same method, one would expect different trends in mortality risk by time if there were significant drift health selection effects. However, for education we expect no difference in trends between education categories, because for the vast majority of people aged 25–64 poor health is unlikely to have affected their highest qualification. Figure 33 tests this assumption. While the lines are somewhat erratic for 25–44 year olds due to few deaths, the mostly parallel plots are consistent with no drift health selection by education, and thus also consistent with those shown for small area deprivation (Figure 29, page 209).

Cancer and cardiovascular disease deaths among 45–64 year olds: all labour force categories, and excluding non-active labour force

Figure 34 (above) plots the cancer mortality risks over time for the active labour force only, for 45–64 year olds. There was some evidence of a steeper cancer mortality plot over time for those with nil education compared to those with tertiary education among males (Figure 34b). This difference is demonstrated more clearly in Figure 35, where the mortality risk plots are plotted only for these two educational groups and includes trend-lines (calculated in Microsoft Excel). If the first six-months of deaths were excluded (as they were in the cohort analyses), then the pattern of diverging slopes shown in Figure 35 becomes even more prominent. Thus, there was some suggestion of differential health selection by educational status for cancer mortality among 45–64 year old males. On the other hand:

- the cancer mortality risk plots over time were also rising for the 'school' and 'trade' groups among 45–64 year old males in Figure 34b
- there was no suggestion of differing cancer mortality risk slopes among 45–64 year old females (Figure 34d)
- there was no suggestion of differing cardiovascular mortality risk slopes by education among either males or females aged 45–64 years after excluding the non-active labour (results not shown)
- previously presented mortality risk plots by NZDep91 excluding the non-active labour force status failed to suggest differential health selection for either males and females aged 45–64 years, and for either cancer (Figures 30b, d) or cardiovascular disease.

Figure 35: Cancer mortality risk for each six-month period following census night for 45–64 year old males with tertiary or nil qualifications, excluding the non-active labour force



Note: Error bars are 95% confidence intervals about each mortality risk, calculated according to Rothman and Greenland (1998; pp.240–1).

In light of these bullet points, the apparently flat cancer mortality risk plot for the tertiary qualification group among males aged 45–64 years may just have been an aberrant result. *Thus, it is not possible to point confidently to strong evidence of differential health selection by stable socioeconomic factors (ie, NZDep91 and highest qualification) in the NZCMS.*

(Mortality risk plots by occupational class will be considered later in this appendix, and do suggest some differential health selection – but only in the first six or 12 months.)

2.2.2 Exclusion of sickness beneficiaries

The exclusion of decedents who were sickness beneficiaries caused the tertiary qualification compared to nil qualification excess odds ratio for all-cause mortality to decrease by up to 11% across the four sex by age groups (Table 59 below). The similar percentage reductions for the NZDep91 quintile 5 compared to quintile 1 odds ratio were somewhat greater at 9% to 18% (Table 57, page 212). If *drift* health selection was notably biasing the all-cause mortality gradients by household income in the NZCMS, then excluding sickness beneficiaries should produce a substantially greater reduction in the all-cause mortality gradients for income (presented later in report).

2.2.3 Exclusion of decedents with a hospitalisation event between 1988 and census night

The exclusion of decedents hospitalised between 1988 and census night reduced to the null the tertiary qualification compared to nil qualification excess odds ratio for all-cause mortality by 4% to 21%, across the four sex by age groups (Table 59). The similar percentage reductions for the NZDep91 quintile 5 compared to quintile 1 odds ratio ranged from 4% to 51% (Table 57). As with the sickness beneficiary exclusion above, if *drift* health selection was notably biasing the all-cause mortality gradients by household income in the NZCMS then the later sensitivity analyses for income excluding pre-hospitalised deaths should produce a substantially greater reduction in the all-cause mortality gradients.

Table 59: Odds ratios (95% CI) of all-cause mortality for 25–64 year olds in the restricted cohort, by highest qualification, for various exclusions testing for health selection

Exclusion criteria	Odds ratios (ref group = nil qualifications)				% change to null of tertiary OR †
	Tertiary	Trade	School	Nil	
Males					
25–44 year olds					
Nil	0.57 (0.47–0.69)	0.74 (0.63–0.87)	0.73 (0.61–0.87)	1.00	
Sickness beneficiaries	0.61 (0.50–0.75)	0.78 (0.66–0.92)	0.75 (0.62–0.91)	1.00	10%
Pre-hospitalised deaths	0.66 (0.53–0.82)	0.81 (0.67–0.99)	0.86 (0.70–1.07)	1.00	21%
Non-labour force	0.63 (0.51–0.77)	0.78 (0.65–0.93)	0.76 (0.62–0.93)	1.00	14%
45–64 year olds					
Nil	0.65 (0.59–0.72)	0.80 (0.74–0.85)	0.80 (0.73–0.87)	1.00	
Sickness beneficiaries	0.69 (0.62–0.76)	0.81 (0.75–0.87)	0.82 (0.75–0.90)	1.00	10%
Pre-hospitalised deaths	0.70 (0.62–0.78)	0.80 (0.73–0.87)	0.79 (0.71–0.88)	1.00	13%
Non-labour force	0.73 (0.64–0.83)	0.83 (0.75–0.92)	0.87 (0.77–0.98)	1.00	22%
Females					
25–44 year olds					
Nil	0.56 (0.45–0.69)	0.63 (0.49–0.80)	0.61 (0.50–0.75)	1.00	
Sickness beneficiaries	0.61 (0.49–0.76)	0.65 (0.51–0.84)	0.65 (0.53–0.80)	1.00	11%
Pre-hospitalised deaths	0.58 (0.43–0.78)	0.56 (0.39–0.80)	0.60 (0.45–0.80)	1.00	4%
Non-labour force	0.67 (0.52–0.86)	0.62 (0.45–0.85)	0.61 (0.46–0.80)	1.00	25%
45–64 year olds					
Nil	0.70 (0.63–0.78)	0.79 (0.69–0.90)	0.86 (0.79–0.95)	1.00	
Sickness beneficiaries	0.70 (0.62–0.78)	0.78 (0.68–0.89)	0.86 (0.78–0.94)	1.00	-1%
Pre-hospitalised deaths	0.76 (0.66–0.88)	0.84 (0.70–1.00)	0.96 (0.85–1.08)	1.00	20%
Non-labour force	0.82 (0.69–0.97)	0.90 (0.73–1.11)	1.02 (0.86–1.21)	1.00	40%

Note: Injury and suicide deaths are not presented for females due to small numbers. For all remaining cells in the table there were at least 30 deaths.

† Percentage change is for the excess odds ratio for quintile 5 compared to quintile 1, compared to the same odds ratio with nil exclusions.

2.3 Excluding the non-labour force

2.3.1 All-cause mortality

The reductions in the gradient of all-cause mortality by highest qualification following restriction to the active labour force only (Table 59 above) tended to be less than that for small area deprivation (Table 57, page 212) – but were broadly consistent. The percentage reductions to the null for the tertiary compared to nil qualification odds ratio (and the equivalent reduction for the quintile 5 versus quintile 1 deprivation odds ratio in parentheses taken from Table 57) were 14% (14%) and 22% (43%) for 25–44 and 45–64 year old males respectively, and 25% (32%) and 40% (53%) for 25–44 and 45–64 year old females respectively.

2.3.2 Cause-specific mortality

Regarding cause-specific mortality, the interpretation of the crude risk ratios following exclusions of the non-labour force were problematic due to the strong confounding by age of the association of education with mortality. However, the male results were broadly consistent with those shown in Figure 31 (page 213) for small area deprivation – the reduction in the gradient for cancer and cardiovascular disease deaths following exclusion of the non-labour force was greater than that for suicide deaths, and the gradient actually increased for injury deaths.

3 Household tenure

Only the restricted cohort results are presented for all-cause mortality by housing tenure – a proxy for asset wealth.

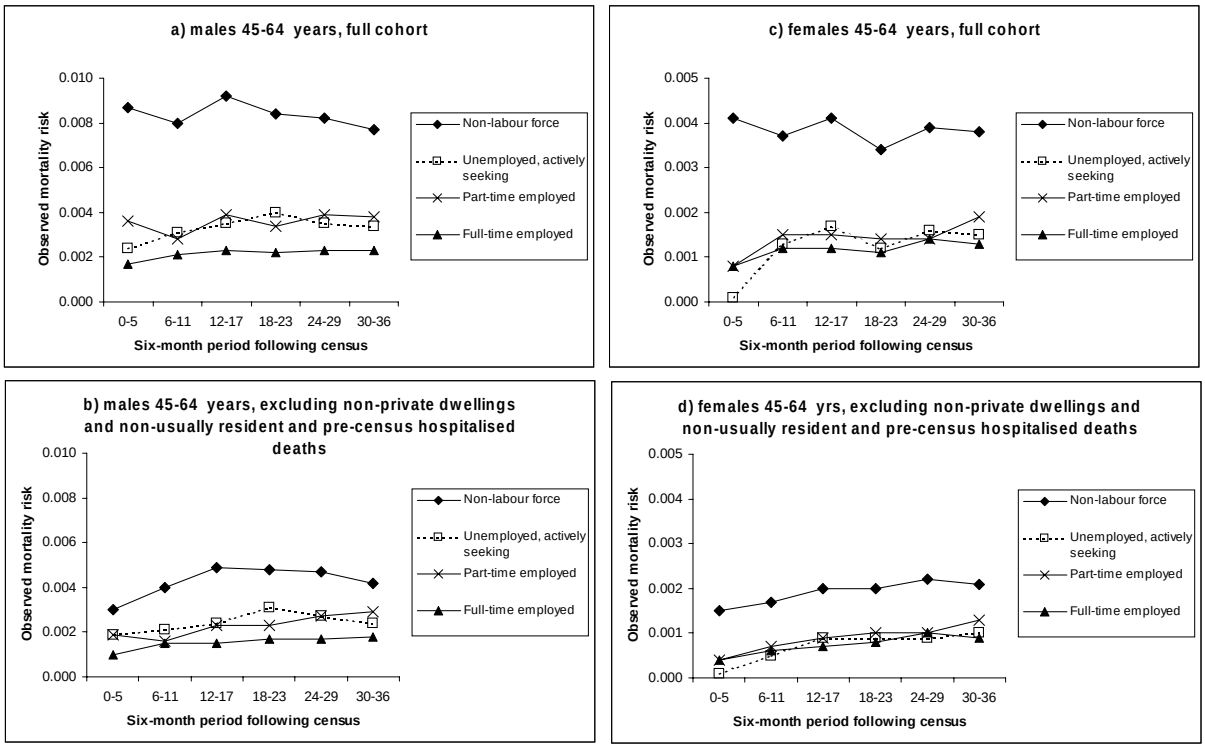
4 Labour force status

There was no suggestion of selection bias for labour force status analyses on the restricted cohort, and a direct assessment of linkage bias was not possible. Thus, the focus for sensitivity analyses was on health selection.

4.1 Health selection

Perhaps the most informative way to assess the impact of health selection on labour force status is to view plots of mortality risk over time by labour force status. Selected plots are presented in Figure 36 (below) for males and females aged 45–64 years. In Figure 36, the lines for the 'seeking or available' category are not shown due to small numbers, and hence unstable plots.

Figure 36: Mortality risk for each six-month period following census night by labour force status for 45–64 year old males and females



Several points are demonstrated or suggested by the plots in Figure 36:

- The mortality risk tended to decrease over time among the 'Retired, homemaker, permanently sick, students, etc' (the majority of the non-labour force) in the full cohort for both males and females aged 45–64 years. This trend was the same among the restricted cohort (not shown), and more pronounced among 25–44 year olds. Regarding the 25–44 year olds, the mortality risk halved from the first to the last six-month period of follow-up among both males and females in the non-labour force aged 25–44 years – a steep reduction in mortality. This diminishing mortality over time is what would be expected with short-term health selection (ie, people who are sick and expected to die soon moving out of the labour force), and was also demonstrated in the OPCS Longitudinal Study (Fox and Goldblatt 1982; Fox et al 1985).
- There was no similar falling mortality risk over time among the unemployed 45–64 year olds in the full cohort. Rather there was a tendency for the mortality risk to increase over time in a similar way to the full-time employed, ie, a healthy worker effect. An inspection of 95% confidence intervals (not shown) for the unemployed mortality risks plotted in Figure 36 suggested that the apparent healthy worker effect was unlikely to be due to chance. (Among the unemployed 25–44 year olds, there were too few deaths to draw any conclusions.) This apparent 'healthy worker effect' among the unemployed strongly argues against the elevated mortality risk among the unemployed being due to health selection. Rather, poor health presumably moves people out of employment to the non-labour force, not to the unemployed. This conjecture is also consistent with the relatively rigorous definition of unemployment in the 1991 census, where the person is required to be both actively seeking work and available for work.

- As in the above bullet point, it is also possible to argue that the part-time employed demonstrate a healthy worker effect among the full cohort, arguing against the elevated mortality of the part-time employed being due to health selection.
- The lower two plots in Figure 36 add further weight to the above conclusions. Excluding sickness beneficiaries and those decedents who were hospitalised most notably reduces the mortality risk among the 'retired permanently sick, students, etc', particularly early in the follow-up. This reduction is supportive of a large health selection effect among the non-labour force. Further, applying the same exclusions to 25–44 year olds dramatically reduced the excess risk of mortality among the non-labour force.
- Among 45–64 year old males the *relatively* larger mortality risk among the unemployed (and part-time employed) compared to the full-time employed is not diminished by excluding people with a previous hospitalisation (plot b). This provides further support to the conclusion based on the full cohort plots that the elevated mortality risk among the unemployed (and part-time employed) was not due to health selection.

4.2 Conclusion

The elevated mortality among the non-labour force is, in large part at least, due to health selection.

Both the unemployed and the part-time employed had elevated mortality risks compared to the full-time employed. While this elevation may be the result of confounding (and will be assessed later in the multivariate analyses), the analyses here strongly suggest that it was *not* due to health selection. Further, there was no apparent selection bias, and linkage bias would probably have caused (if anything) an underestimate of the increased mortality among the non-referent labour force groups.

Finally, the particularly notable healthy worker effect between the first and second six-month periods of follow-up supports the decision in the NZCMS to discard linked deaths for the first six months as a means of mitigating against health selection effects in the cohort analyses.

5 Occupational class

Occupational class analyses presented a challenge due to only current occupation being available on the census. The unlinked occupational class analyses by Pearce and colleagues, however, provide a useful point of comparison (Davis et al 1999a; Kawachi et al 1991; Marshall et al 1993; Pearce and Bethwaite 1997; Pearce et al 1983a; Pearce et al 1983b; Pearce et al 1984; Pearce et al 1985; Pearce and Howard 1986; Pearce et al 1991; Pearce et al 1993).

The main cohort results for occupational class were presented in Chapter 6. The objectives of this section were to:

- to assess the likely impact of health selection on the association of occupational class with mortality observed in the NZCMS
- compare and contrast the results from unlinked analyses (ie, as done by Pearce and colleagues) and results using the NZCMS, with particular attention to identifying sources of numerator–denominator bias that may be affecting unlinked analyses.

Regarding the latter comparability with the unlinked analyses of Pearce and colleagues, this report also:

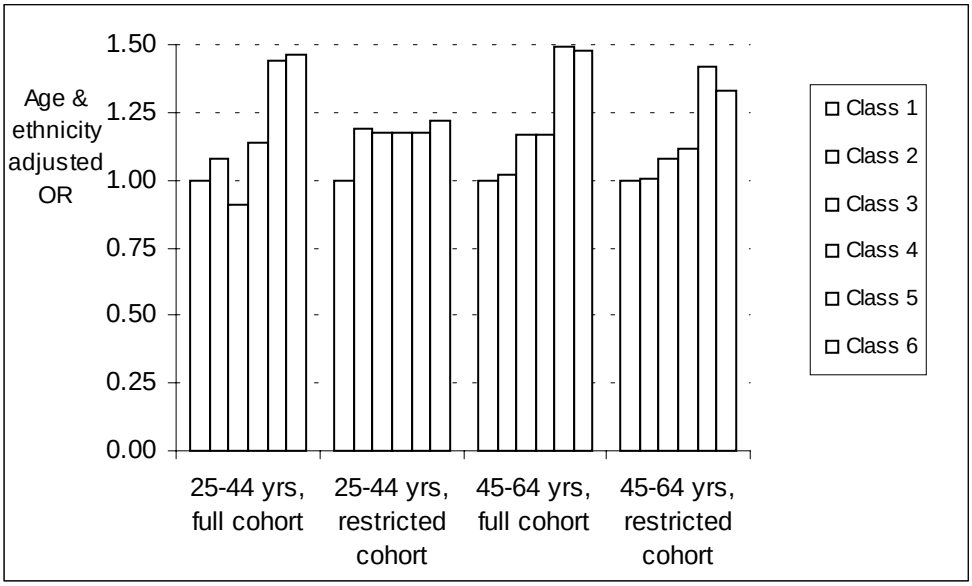
- presents results adjusted for age only, in addition to those adjusted for age and ethnicity, as the unlinked analyses of Pearce and colleagues standardised for age only
- examines how the occupational class assignment among decedents varies between census and mortality data. This variation is best framed as an analysis of numerator–denominator bias between the unlinked and linked analyses, but may also include components of health selection
- conducts analyses using the occupational class based on the death registration form as the exposure (rather than that derived from the census record) among linked deaths, thus imitating an unlinked analysis
- conducts analyses using the occupational class 1 and 2 divisions proposed by Davis et al (Davis et al 1997) to allow direct comparability with the submitted unlinked 1995–97 analyses by Pearce and Sporle (personal communication, 2000).

Due to the substantial problems with occupational class data for females, male data dominates the sensitivity analyses.

5.1 Selection bias for restricted cohort results

Figure 37 presents the occupational class mortality gradients for those in the full cohort with an occupational class. There will be occasion to include occupational class in multivariate analyses with other socioeconomic factors such as education, income, and car access. Thus, it is useful to know the amount of selection bias incurred when assessing the gradient among the restricted cohort. Figure 37 (below) shows the occupational class mortality gradient for males aged 25–44 and 45–64 years, for both the full and restricted cohort. There was a selection bias, such that the gradients were underestimated in the restricted cohort – particularly among 25–44 year olds. The magnitude of selection bias was the same for age-only adjusted odds ratios.

Figure 37: Age and ethnicity adjusted odds ratios of all-cause mortality by NZSEI occupational class among males – a test of selection bias between the full and restricted cohorts



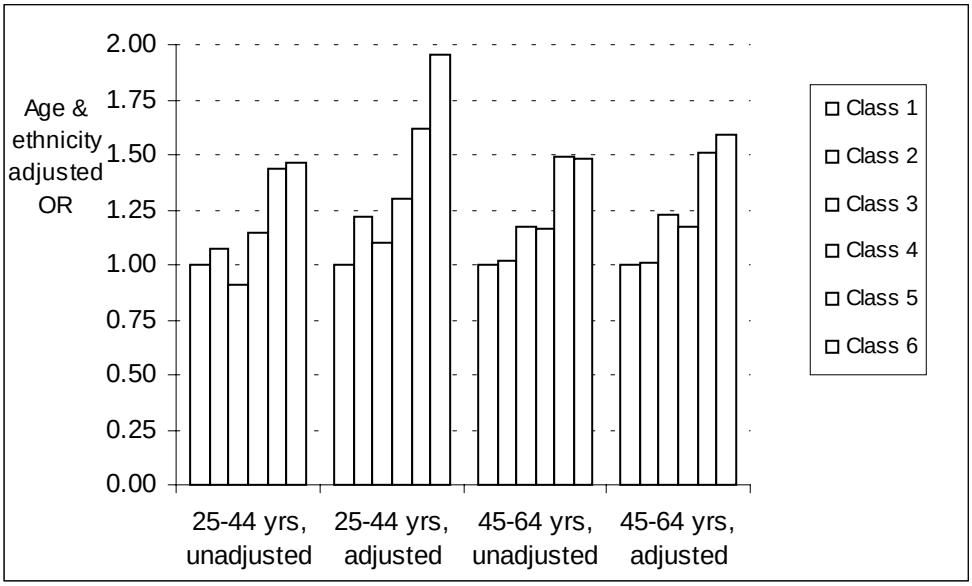
These results for selection bias also add to the accumulating evidence on selection bias in this report. For all-mortality, there was no selection bias of note by highest education for either 25–44 or 45–64 year old males (Section 2.1 of this appendix), but there was selection bias by NZDep91 such that the restricted cohort modestly underestimated the gradient among 25–44 year olds and modestly overestimated it among 45–64 year olds. Adding the occupational class analyses to this accumulating evidence, there is now evidence of selection bias causing:

- an underestimate of the 25–44 year old male all-cause mortality gradient for analyses on the restricted cohort for two out of three socioeconomic factors tested
- an underestimate (occupational class), no effect (education), and an overestimate (NZDep91) of the 45–64 year old male all-cause mortality gradient for analyses on the restricted cohort.

5.2 Linkage bias

Results present in Chapter 3 demonstrated linkage bias by occupational class among males – particularly among 25–44 year olds, and for occupational class 6. Using the risk ratios reported in Table 23, the occupational class mortality gradients were adjusted for linkage bias. The gradients increased, particularly for 25–44 year olds males such that after adjustment for linkage bias there was a near doubling of the odds of death between occupational class 6 and class 1 (Figure 38 below).

Figure 38: Age and ethnicity adjusted odds ratios of all-cause mortality by NZSEI occupational class among males in the full cohort– the effect of adjusting for linkage bias



5.3 Numerator–denominator bias

Unlinked occupational class analyses, such as those by Pearce and colleagues, may be prone to numerator–denominator bias whereby the occupation is collected differently between mortality and census data. In New Zealand there are likely to be differences between census and mortality data due to the occupation on the census being *current* and that on the death registration form being *usual*. The objective of this section is to assess these likely numerator–denominator biases.

Tables 60 and 61 below present a summarised cross-classification of occupational class for the mortality and census data, for male and female decedents respectively. The most informative results in Tables 60 and 61 are the final columns that present the ratio of the number of census to mortality records for each class. For males with any assigned occupational class the overall ratio was 0.56, ie, there were 44% fewer male decedents with an occupational class on census data than on mortality data. The fact that the ratios are relatively constant by occupational class (with the exception of occupational class 4 to be discussed subsequently) means that there should be at most only moderate numerator–denominator biases affecting unlinked occupational class analyses. However, the comparability of these ratios by occupational class does not rule out systematic biases affecting both the unlinked and linked analyses in the same direction. For example, differential health selection out of the labour force may cause an underestimate of lower occupational class deaths in census data and a tendency to 'promote the dead' would also tend to cause an underestimate of the lower occupational class deaths on mortality data.

Table 60: Cross-classification of mortality by census occupational class for 5844 male 25–64 year old deaths during the second and third year of follow-up

Occupational class	Census data	Number (%) by census data that:			Mortality data	Number (%) by mortality data that:			Census total to mortality total ratio
		Had same class on *mortality* data	Had different class on *mortality* data	Had no occupation on *mortality* data		Had same class on *census* data	Had different class on *census* data	Had no occupation on *census* data	
Class 1	240	165 (69%)	69 (29%)	6 (3%)	387	168 (43%)	75 (19%)	147 (38%)	0.62
Class 2	300	150 (50%)	126 (42%)	24 (8%)	462	150 (32%)	126 (27%)	189 (41%)	0.65
Class 3	498	279 (56%)	189 (38%)	33 (7%)	861	276 (32%)	216 (25%)	369 (43%)	0.58
Class 4	531	333 (63%)	162 (31%)	36 (7%)	1182	333 (28%)	177 (15%)	672 (57%)	0.45
Class 5	627	369 (59%)	180 (29%)	72 (11%)	1137	372 (33%)	159 (14%)	609 (54%)	0.55
Class 6	252	93 (37%)	111 (44%)	51 (20%)	393	93 (24%)	60 (15%)	243 (62%)	0.64
Farmers	306	261 (85%)	33 (11%)	12 (4%)	507	261 (51%)	57 (11%)	189 (37%)	0.60
Subtotal with occupation	2754	1650 (60%)	867 (31%)	234 (8%)	4932	1650 (33%)	867 (18%)	2415 (49%)	0.56
No occupation	3090	NA	NA	678 (22%)	912	NA	NA	678 (74%)	3.39

Regarding occupational class 4 for males, the census to mortality ratio of 0.45 is substantially lower than the other ratios. This discrepancy suggests that there was a tendency to overestimate the number of deaths in occupational class 4 on mortality data compared to census data.

For female deaths (Table 61), the census to mortality ratio increased with lower occupational class. *Thus, unlinked analyses using only female mortality data will find a shallower (or even inverse) occupational class mortality gradient compared to linked census–mortality data.* This discrepancy between the census and mortality data is not due to differential health selection – that should cause the ratios to decrease, not increase, with lower class. A possible explanation is that females of lower occupational class tend to have frequent part-time or casual work that is detected cross-sectionally by the census, but tends not to be entered as an 'usual occupation' on the death registration form. It is interesting to note that slightly more female deaths had an occupational class according to census data (1158) compared to mortality data (1016) – for male deaths many more had a mortality record occupation (4932) than a census record occupation (2754).

Table 61: Cross-classification of mortality by census occupational class for 3798 female 25–64 year old deaths during the second and third year of follow-up

Occupational class	Census data	Number (%) by census data that:			Mortality data	Number (%) by mortality data that:			Census total to mortality total ratio
		had same class on *mortality* data	had different class on *mortality* data	had no occupation on *mortality* data		had same class on *census* data	had different class on *census* data	had no occupation on *census* data	
Class 1	63	27 (43%)	24 (38%)	12 (19%)	54	24 (44%)	12 (22%)	15 (28%)	1.17
Class 2	141	72 (51%)	27 (19%)	39 (28%)	211	72 (34%)	57 (27%)	81 (38%)	0.67
Class 3	177	57 (32%)	48 (27%)	72 (41%)	190	60 (32%)	72 (38%)	54 (28%)	0.93
Class 4	264	96 (36%)	45 (17%)	123 (47%)	227	96 (42%)	33 (15%)	96 (42%)	1.16
Class 5	219	63 (29%)	48 (22%)	111 (51%)	162	63 (39%)	27 (17%)	69 (43%)	1.35
Class 6	216	54 (25%)	27 (13%)	132 (61%)	134	51 (38%)	12 (9%)	63 (47%)	1.61
Farmers	78	15 (19%)	9 (12%)	57 (73%)	40	15 (38%)	6 (15%)	18 (45%)	1.95
Subtotal with occupation	1158	387 (33%)	225 (19%)	549 (47%)	1016	384 (38%)	225 (22%)	402 (40%)	1.14
No occupation	2640	NA	NA	2232 (85%)	2782	NA	NA	2235 (80%)	0.95

Another useful piece of information that can be gleaned from Tables 60 and 61 above was that for both males and females there was much greater variation between census and mortality data for the assignation of a lower occupational class. For example, a greater percentage of male decedents assigned to occupational class 6 on the mortality data had no occupation on census data (62%, second to last column of Table 60) than did male decedents assigned to occupational class 1 on mortality data (38%). Likewise, for assignment of male census occupational class a similar trend was evident for having no occupation on mortality data (20% and 3% for occupational class 6 and 1, respectively). Further, the same trends were evident for females (Table 61). The reason for these trends was that of the decedents with no occupation on census data (mortality data) but with an occupation on mortality data (census data), the latter occupation was more likely to be for a lower occupational class (data not shown). These trends suggest that *both* census and mortality data may be underestimating the number of lower occupational class deaths more so than higher occupational class deaths. Whether this means that both unlinked (eg, those by Pearce and colleagues) and linked analyses (eg, those in the NZCMS) are likely to underestimate the occupational class mortality gradient, however, is uncertain. It is likely that entry and exit (or cycling) into and out of employment among the lower occupational classes may be more common than among the higher occupational classes. Thus, there may be a similar tendency among the census denominator to underestimate the number of people in the lower occupational classes, causing no overall bias in the occupational mortality gradients.

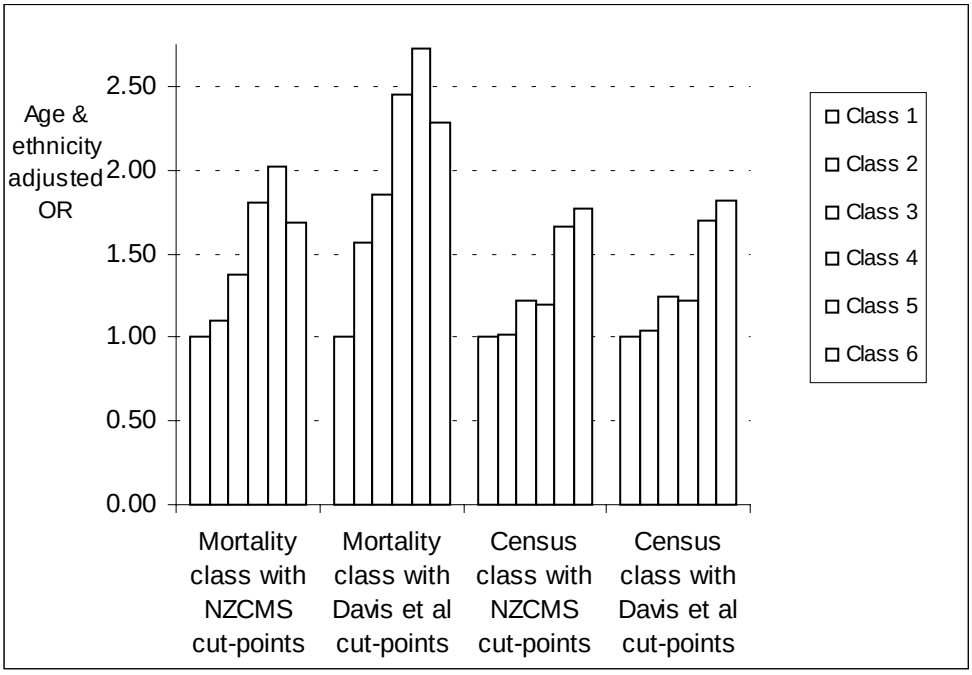
One possible numerator–denominator bias suggested in the literature for unlinked analyses is a tendency to 'promote the dead' whereby a decedent's occupational class is categorised as higher than their actual class. It was possible to test this directly in the NZCMS. Shown in Tables 60 and 61 are the numbers of decedents that changed occupational class between census and mortality data. Of the 774 male decedents with an assigned occupational class of 1 to 6 on mortality data, and who were assigned to another class (ie, 1 to 6) on census data, 357 were assigned a higher occupational class on census data and 417 were assigned a lower occupational class on census data. Thus, there was a tendency for higher occupational class on the mortality data consistent with the 'promoting the dead' hypothesis, but the magnitude of the bias was small. The same numbers for females were 126 and 87, respectively, again consistent with a (modest) tendency to promote the dead.

There was a specific numerator–denominator bias apparent for the NZSEI occupational classes. The NZCMS uses a different cut-off between occupational classes 1 and 2 than that originally proposed by Davis et al (Davis et al 1997), for reasons described in Chapter 2 of this report. The occupational class mortality gradients presented below in Figure 39 demonstrate the effect of using the NZCMS cut-points versus the Davis et al cut-points, for 45–64 year old males. (Analyses for 25–44 year olds found the same conclusions as follow.) The first set of columns in Figure 39 present the occupational class mortality gradient using the NZCMS cut-points, where occupational class for the deaths was taken from the mortality data – thus this analysis imitates an unlinked analysis. The second set of columns uses the cut-point between occupational classes 1 and 2 proposed by Davis et al (1997), but still uses the mortality data occupational class for decedents.

There is a dramatic difference between the gradient in the first and second set of columns, due to the mortality risk among the Davis et al class 1 being substantially lower than that in the NZCMS class 1. Correspondingly, using the Davis et al cut-points, and occupational class 1 as the reference category, the gradient appears much steeper. If the mortality risk among the Davis et al class 1 was *truly* much less than the NZCMS class 1, then one would expect a similar change in the gradient using the census assigned occupational classes for the entire cohort-base. The third and fourth set of columns in Figure 39 present the gradient using the NZCMS and Davis et al cut-points and census assigned occupational class – there is virtually no difference in the gradient between the third and fourth set of columns.

Thus, this report concludes that there is a substantial numerator–denominator bias on the mortality data *within* the 70 to 90 range of NZSEI scores. It appears that the coding of occupation on mortality data is assigning many decedents to an occupation with a NZSEI score between 70 and 75, when in fact the census coding would have assigned them to an occupation with a NZSEI score between 75 and 90. Put another way, the mortality gradient depicted in the second set of columns of Figure 39 is affected by a numerator–denominator bias that causes a substantial underestimate of the mortality risk in occupational class 1. An initial inspection of the draft results for 1995–97 by Pearce, Davis and Sporle show a gradient for 1995-97 by NZSEI occupational class that is remarkably similar to that shown in the second set of columns of Figure 39 below. Thus, it seems likely that the apparent numerator denominator bias for mortality data compared to census data about the cut-point between NZSEI classes 1 and 2 also affects 1995-97 mortality data.

Figure 39: Odds ratios of all-cause mortality by NZSEI occupational class among 45–64 year old males, using NZCMS versus Davis et al cut-points, and mortality versus census data occupational class for decedents



Note: All analyses use the full cohort, age-only adjusted odds ratios – but additionally adjusted for linkage bias. Occupational class 1 is the references group.

Putting aside the apparent numerator–denominator bias affecting class 1 aside, there are two other useful findings demonstrated in Figure 39. First, and as suggested by the ratios in the final column of Table 60 (page 230), the NZSEI occupational class 4 mortality risk appears to be overestimated by mortality data. Second, using the census data, occupational class 6 has a higher mortality risk (after adjustment for linkage bias) than occupational class 5. Using mortality data, or an unlinked analysis, the mortality risk is apparently lower in class 6.

Summarising, there were apparent numerator–denominator biases:

- biasing the mortality risk for different NZSEI score cut-points between classes 1 and 2
- causing a relative overestimate of the mortality risk among class 4 using mortality data
- and causing a relative underestimate of the mortality risk for class 6 using mortality data; or conversely, a relative overestimate of the mortality risk for class 5 using mortality data.

Taking all these into account, however, it is important to note that there is broad agreement between the unlinked and linked analyses (eg, the first and third set of columns in Figure 39 are broadly comparable).

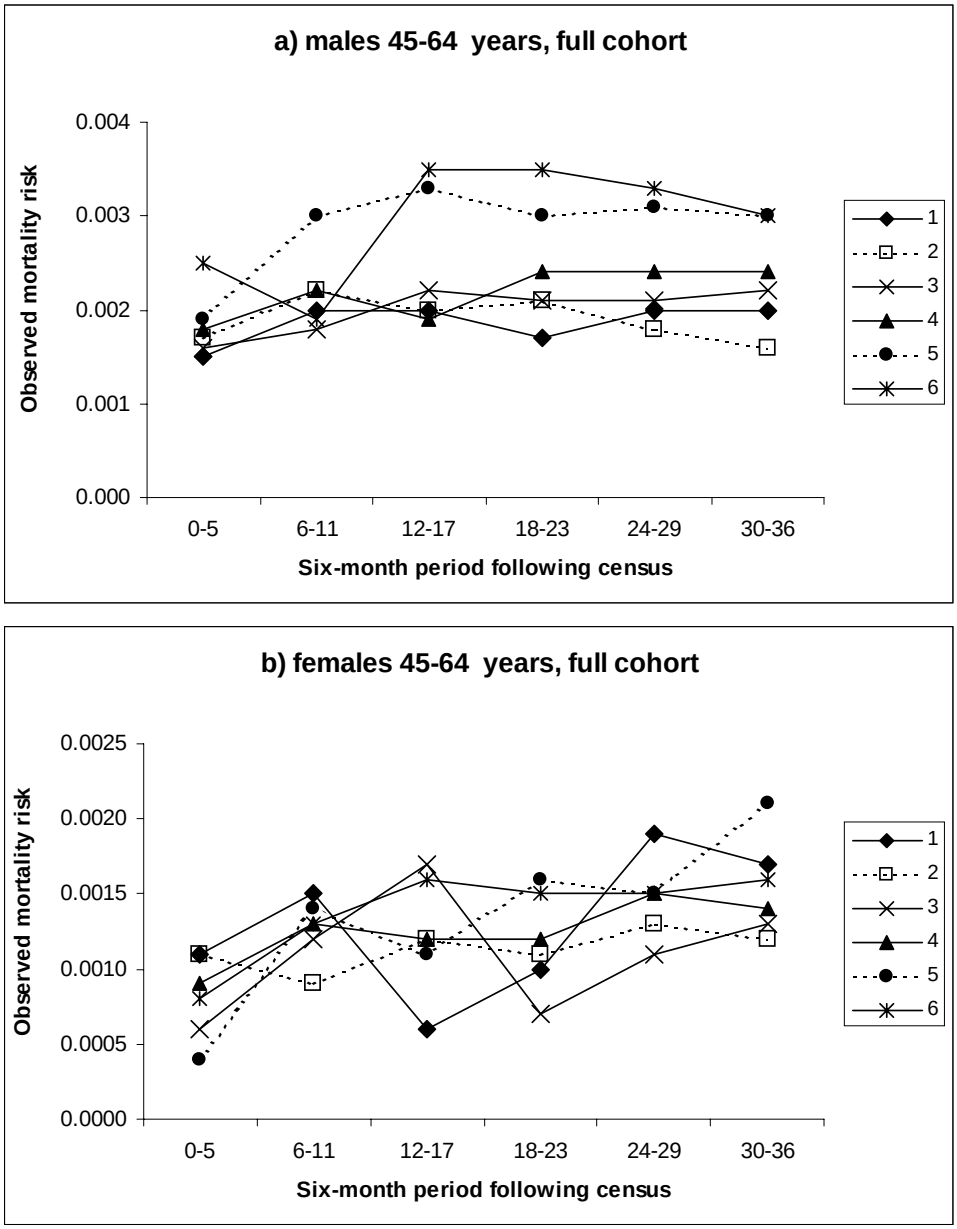
5.4 Health selection

Sensitivity analyses, excluding sickness beneficiaries and decedents with a hospitalisation between 1988 and census night, are not presented for occupational class mortality gradients. As the occupational class mortality gradients were already restricted to those in the labour force (a possible form of differential health selection itself), such sensitivity analyses were difficult to interpret and probably unreliable. Rather, plots of mortality risk over time have been relied on to investigate differential health selection.

5.4.1 Observed mortality risk over time

Plots of mortality risk over time by current occupational class allow the investigation of two forms of health selection that may act over the short term: *differential* health selection out of the labour force, and short-term *drift* health selection up and down occupational classes. If present, these two forms of health selection will give opposite patterns of mortality risks plotted over time – a divergence for differential health selection, and a convergence for short-term health selection down occupational classes. Strictly speaking, plots of mortality risk over time by occupational class allow comment only on the net effect of these two forms of health selection. However, previous research suggests that drift health selection for occupational classes is negligible (Fox et al 1985; Power et al 1996; van de Mheen et al 1999). Accordingly, the patterns in mortality risk over time in terms of differential health selection have been interpreted.

Figure 40: Mortality risk for each six-month period following census night by NZSEI occupational class for 45–64 year old males and females



The observed mortality risks by time period after census night are shown in Figure 40 above for males and females aged 45–64 in the full cohort. Unfortunately, the point estimates are too unstable for females (due to small numbers) to allow interpretation. However, for males there was little difference between the mortality risks in the first six months, consistent with differential health selection that would be more likely to force lower occupational class people with poor health out of the work force than higher occupational classes. Over time, the mortality risks diverge to the expected higher mortality for occupational classes 5 and 6. Of note, the occupational class 6 mortality risk did not rise till the second year of follow-up. While this delayed rise may just be random variation, it may also be a longer apparent differential health selection among the lowest occupational class. (As such, it was one of the contributing reasons to discarding all census respondents dying in the first year for the cohort analyses of occupational class.) Thus, there was evidence of differential health selection affecting the occupational class mortality gradient – but not beyond one year of follow-up, and

particularly in the first six-months of follow-up. It is unclear whether the mortality risks would have diverged further over a period of follow-up longer than three years, although there is some suggestion of stability in the last three six-month periods of follow-up.

If the pattern in Figure 40 is assumed to reflect differential health selection, then what are the implications? First, as occupational class cohort analyses discarded deaths in the first year, and assuming that further divergence would not occur with a longer follow-up, then the cohort analyses of occupational class mortality gradients in this report might be relatively free of differential health selection. However, based on the international literature and the possibility that the plots for all-cause mortality may hide some further divergence by cause of death beyond three years of follow-up, it is more prudent to conclude that some differential health selection may be affecting the occupational class analyses in the NZCMS – but probably not greatly. Second, the pattern does add some support to the suggestion of differential health selection for cancer deaths among 45–64 year old males by highest qualification (Figure 35). But, the support it offers is of limited practical importance as the differential health selection by occupational class appeared to be mostly in the first six months, and all cohort analyses in the NZCMS discard deaths in the first six-months.

5.4.2 Adjustment of the *current* occupational class mortality gradient to approximate the *usual* occupational class mortality gradient

One method proposed by Kunst and colleagues to adjust occupational mortality gradients for differential health selection was to assess the change in the educational mortality gradient between that for all people and that among just the labour force, and assume that this same difference would apply to occupational class mortality gradients (Kunst et al 1998b) (see discussion in Appendix A, page 169). However, as discussed in Appendix A (and elsewhere in this report), such a method may adjust not only for differential health selection, but also for confounding/mediation of the association of *usual/last* occupational class with mortality by *current* labour force status. This possibility is supported by the pattern in Figure 40 above of notable differential health selection in the first year, but little evidence thereafter – although a three-year follow-up limits this conclusion. Nevertheless, using the association of small area deprivation and education with mortality including and excluding the non-labour force presented previously in this report (crude data), it was possible to adjust the current occupational class mortality gradients to approximate the (unobserved) usual occupational class mortality gradients.

The adjustment was conducted as follows:

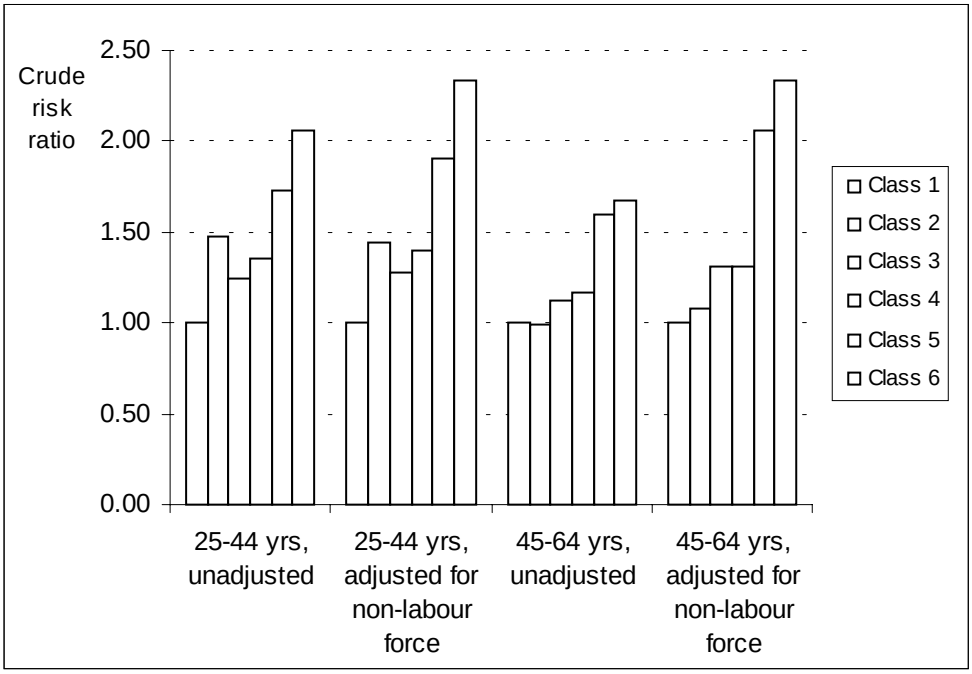
- Approximating the percentage distribution of the population by occupational class, the following small area deprivation and education categories were assumed to be equivalent to occupational classes 1 to 6:
  - [NZDep91 decile 1] and [graduate, postgraduate] for occupational class 1
  - [deciles 2 and 3] and [undergraduate, technical, teaching] for class 2
  - [deciles 4 and 5] and [trade certificate, other tertiary] for class 3
  - [deciles 6 and 7] and [10-12 years of school] for class 4
  - [deciles 8 and 9] and [nil qualification] for class 5
  - [decile 10] and [nil qualification] for class 6.

- The average change in the mortality risk ratio for the deprivation and highest qualification gradient after excluding the non-labour force was determined, where decile 1 and 'graduate, postgraduate' highest qualification were the reference categories.
- Doing so, the ratios to be multiplied into the *current* occupational class odds ratios for 25–44 year old males were 1.00, 0.98, 1.02, 1.03, 1.10, and 1.13 and for 45–64 year old males were 1.00, 1.09, 1.18, 1.13, 1.29, and 1.40, for occupational classes 1 to 6 respectively.

Thus, for example, the mortality risk ratio among males aged 45–64 years in occupational class 6 compared to occupational class 1 (where current occupation only is available) should be multiplied by 1.40 to estimate the risk ratio that would have been observed if usual occupation was available.

Figure 41 below presents the crude risk ratios of mortality by occupational class for males aged 25–44 and 45–64 years in the first and third set of columns. The second and fourth set of columns present the occupational class mortality gradient adjusted for exclusion of the non-labour force, for 25–44 and 45–64 year olds respectively, using the above estimated adjustment ratios. The gradient changes modestly for 25–44 year olds, but increases notably for 45–64 year olds.

Figure 41: Crude risk ratios of all-cause mortality by NZSEI occupational class among 25–44 and 45–64 year old males, before and after adjustment for labour force status



Note: The 'unadjusted' crude risk ratios were adjusted for linkage bias – but not labour force participation.

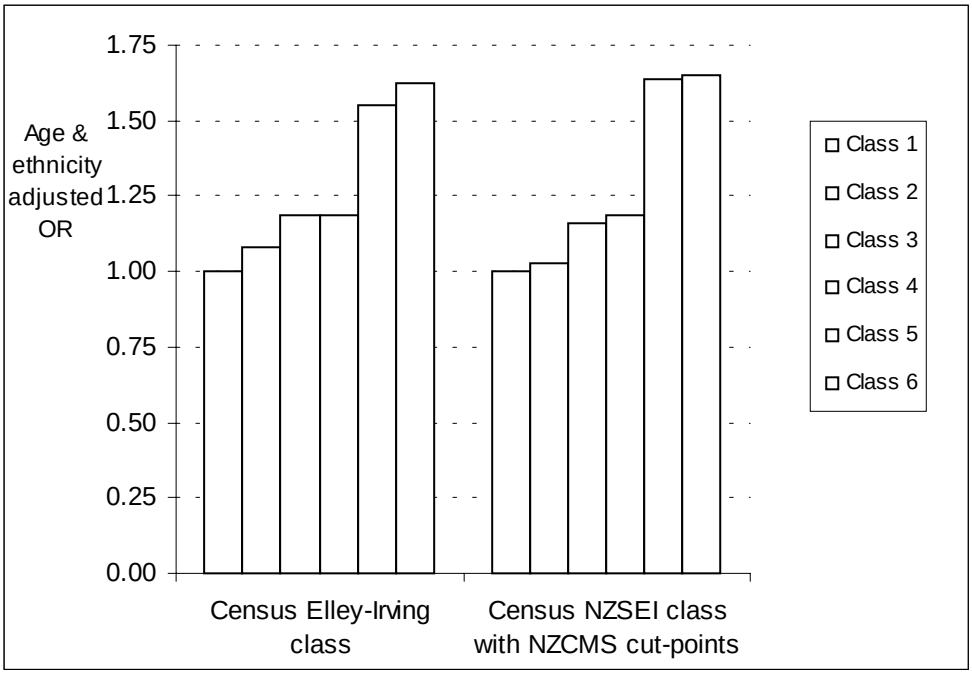
The method used above to adjust for non-labour force status is approximate only, and uses crude data only. The underlying assumption was that the gradient changes by deprivation and education that occur when the non-labour force were excluded was the same as that for occupational class – a questionable assumption. However, it does at least suggest that:

- the use of *current* occupation probably results in a substantial underestimate of the *usual* occupational class mortality gradient for 45–64 year old males, but less so for 25–44 year old males
- adjusting for the exclusion of the non-labour force results in a similar gradient for 25–44 and 45–64 year olds.

Kunst and colleagues have proposed another method to adjust occupational mortality gradients based on *current* occupation to approximate those for *usual* occupation (Kunst 1997; Kunst et al 1996; Kunst et al 1998c). This method draws on survey data external to the census-cohort study that give estimates of the proportion of each occupational class (based on *usual* occupation) that were in the labour force at the time of the census. Such a sensitivity analysis may be conducted in New Zealand using Household Labour Force Survey, but was beyond the scope of this report.

5.5 Comparison of NZSEI and Elley-Irving occupational mortality gradients

Figure 42: Odds ratios of all-cause mortality by NZSEI and Elley-Irving occupational class among 45–64 year old males



Note: The odds ratios are age-adjusted only, and not adjusted for linkage bias (linkage bias was only estimated for NZSEI classes.) Analyses are on the full census cohort, and use the census occupation codes.

The published analyses of occupational class mortality gradients in New Zealand by Pearce and colleagues have used Elley-Irving occupational classes. This classification system requires NZSCO68 codes. Fortunately, both NZSCO68 and NZSCO90 codes were available for 1991 census data, allowing a direct comparison of the NZSEI and Elley-Irving occupational class mortality gradients. Figure 42 presents the gradient for the two different occupational class schemes – the gradients are almost identical.

This comparability between the NZSEI and Elley-Irving gradients means that Pearce and colleagues should be able to extend their times series analysis of occupational class mortality gradients despite having to switch from Elley-Irving to NZSEI classes. However, the above comparability was for the census occupation codes, not the death registration form codes. Unfortunately, comparability on the latter could not be assessed due to the absence of both NZSCO68 and NZSCO90 codes for the 1991–94 time-period. Nevertheless, the two occupational class systems should also be comparable for death registration data and the resultant unlinked analyses.

5.6 Conclusion

The analyses of the association of occupational class with mortality in the NZCMS were limited by the availability of current occupation only. However, it was reasonable to conclude that:

- there was an occupational class mortality gradient in the expected direction for both males and females aged 25–64 years, although only weakly for females
- compared to a male occupational mortality gradients by *usual* occupation, the gradients observed in the NZCMS by *current* occupation were probably an underestimate – particularly for 45–64 year old males
- it was unclear whether the shallower gradient for *current* occupational class compared to *usual* occupational class was due to health selection or confounding/mediation by the range of variables that labour force status may be a proxy for. The plots of mortality risk over time for current occupational class suggested that differential health selection had largely worn-off after the first year of follow-up
- the occupational class mortality gradient by *current* occupation observed in the NZCMS was steeper among 25–44 year old males compared to 45–64 year old males. However, if *usual* occupation data were available there would probably have been little difference in the gradient between the two age-groups.

The above analyses by occupational class also provide important comparative information for unlinked analyses such as those by Pearce and colleagues:

- The linked and unlinked analyses were broadly comparable for *males*.
- Unlinked analyses for females would grossly underestimate the female occupational class mortality gradient due to numerator–denominator biases.
- Despite the broad agreement of the male gradients, there appeared to be several specific numerator–denominator biases for male death registration form occupational class compared to census occupational class:
  - using the cut-off between occupational class 1 and 2 recommended by Davis et al (1997) resulted in a notable underestimate of the class 1 mortality risk according to death registration form data – a cut-off NZSEI score of 70 between occupational classes 1 and 2 is recommended for unlinked analyses

  - occupational class 4 deaths appeared to be overestimated by death registration form data relative to census data
  - compared to census data, occupational class 6 deaths appeared to be underestimated relative to occupational class 5 deaths by death registration form data. Adjusting for this numerator–denominator bias would probably make the class 6 mortality risk higher (rather than lower) than the class 5 mortality risk in unlinked analyses.
- The occupational class mortality gradients appeared similar for the Elley Irving and NZSEI scales.

6 Equivalised household income

6.1 Likely impact of selection bias and linkage bias

It was not possible to directly estimate selection and linkage bias in the NZCMS for equivalised household income. Therefore, only 'likely estimates' are possible.

6.1.1 All-cause mortality

Selection bias for analyses on the restricted cohort has been measured previously in this report for analyses by small area deprivation, highest qualification, and occupational class (Section 1.1, 2.1 and 5.1 of Appendix C). Linkage bias has been measured for small area deprivation and occupational class (Sections 3.3.1 and 3.3.2 of Chapter 3, respectively). On the basis of these sensitivity analyses, it seems plausible to expect that the net impact of selection and linkage biases was an underestimation of the odds ratio comparisons for the lowest compared to the highest equivalised household income groups of approximately:

- 20% for 25–44 year old males and 25–44 and 45–64 year old females
- and 10% for males aged 45–64 years.

This underestimation would be mainly a consequence of linkage bias. Comparisons of middle-income households with high-income households would probably not be underestimated by as much – *the biases mainly operated at the extremes*. Accordingly, the univariate all-cause mortality odds ratios comparing the lowest (<$10,000) and second lowest ($10–$14,999) equivalised household income groups compared with the highest household income group (≥$70,000) for all New Zealanders might have been in the range of:

- 2.15 to 2.55 for males aged 25–44 and 45–64 years (compared with 1.92 to 2.37 observed, calculated from Table 48)
- 1.75 to 2.05 for females aged 25–44 and 45–64 years (compared with 1.55 to 1.85 observed, calculated from Table 48).

6.1.2 Cause-specific mortality

While the sum of linkage and selection bias affecting cause-specific income mortality gradients must equal that expected above for all-cause mortality, it was not possible to estimate reliably the contribution by each cause of death. However, based on the conclusions for the net effect of both biases by NZDep91 (Box 9, page 208), selection biases for education and NZDep91 (Box 10, page 219), and (unstable) estimates of

linkage bias by occupational class for males (Table 24, page 70), it might be reasonable to conclude that:

- there was probably a modest underestimate of the *male cancer* and *cardiovascular disease* gradients
- there was probably little or no net bias affecting the *male injury* gradient
- there was probably an underestimation of the *male suicide* mortality gradient by household income, such that the odds ratio comparing <$20,000 to >$50,000 was greater than 2.5 (observed was 2.27)
- there was probably a sizeable (say 30%) underestimation of the *female* cancer gradient, but it was relatively modest to start with (Table 41, page 95)
- there was probably little or no net bias affecting the *female cardiovascular disease* gradient
- the direction and magnitude of any net bias of the *female injury* and *suicide* gradients was uncertain.

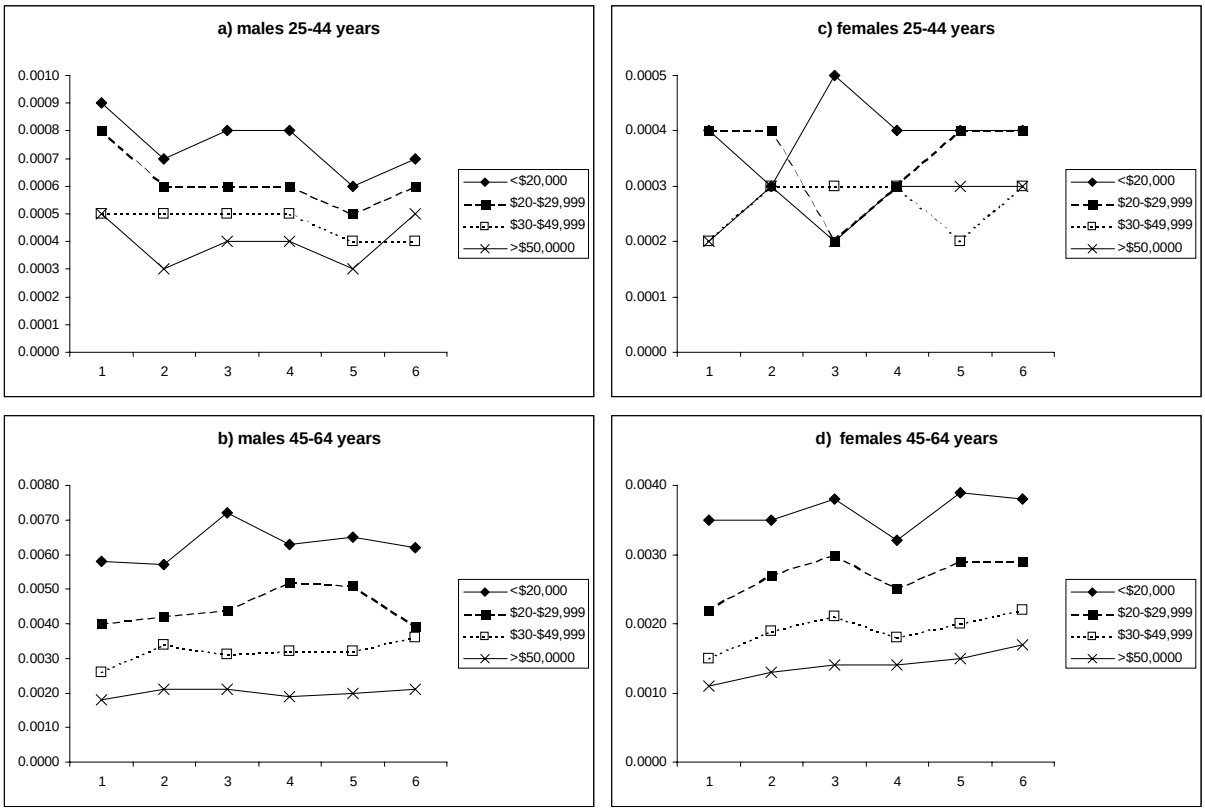
6.2 Health selection

Disentangling the possible impact of health selection on income gradients proved challenging. Most notably, plots of the mortality risk over time suggested no short-term health selection, but the association of income with mortality excluding the non-labour force greatly reduced the income gradient. The results presented in the following sections attempt to present a logical and stepwise series of analyses investigating possible health selection.

Before considering the sensitivity analyses, useful information that may assist the disentangling of health selection includes:

- the income exposure in the NZCMS was equivalised *household* income. Therefore, one would expect the effect of *drift* health selection consequent on one person's ill-health to be mitigated by the income of other members of the household. This would be particularly so for females, where, for 1991–94 and 45–64 year olds at least, the main income earner would often be a male partner.
- the household income was that for the 12 months prior to census night. This 'exposure ascertainment' period, plus the exclusion of deaths in the first six months, should mitigate against short-duration health selection effects.

Figure 43: Mortality risk for each six-month period following census night by four-levels of household equivalised income, 25–64 year olds

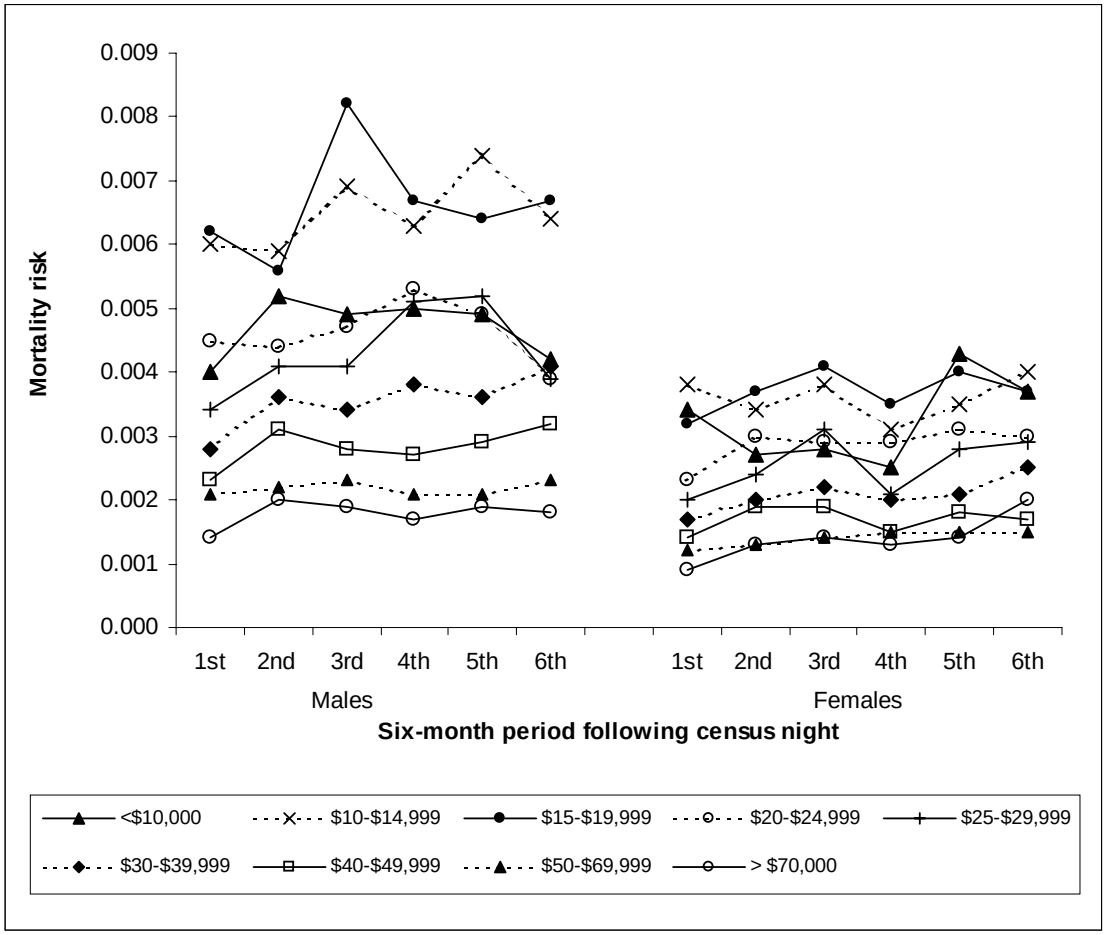


6.2.1 Observed mortality risk over time

All-cause mortality, all labour force categories

Figure 43 shows the plots of mortality risk by time period for each level of equivalised household income, for males and females. Unfortunately, the point estimates for females aged 25–44 are too unstable (due to small numbers) to allow a meaningful interpretation. If *drift* health selection over the short-term was occurring, then one would expect the lines in Figure 43 to converge over time. That is, the low-income people would have high mortality initially due to poor health, and this would fall over time as the unhealthy died or improved in health. Conversely, the mortality risk among the high-income people would be low initially and increase over time as people became unwell, (dropped their income), and died.

Figure 44: Mortality risk for each six-month period following census night by household equivalised income for 45–64 year old males and females



Although not compelling overall, there is some evidence of drift health selection in Figure 43. First, it might be argued that the overall pattern is of the mortality risk lines converging over time among 25–44 year old males (Figure 43a). However, small numbers and SNZ privacy requirements make this conclusion tenuous. For example, 95% CIs are about ±0.00015 (ie, 15-30% of the observed mortality risk at each point in time). Second, for the middle two income categories among 45–64 year old males (Figure 43b) there was convergence during the last six-month period. Figure 44 demonstrates that this was also the case for finer strata of equivalised household income. However, both Figures 43b and 44 demonstrate that apart from this last six-month period the overall pattern among 45–64 year old males is one of approximately parallel mortality risk lines. Third, the mortality risk among 45–64 year old females with a high income (≥$50,000) rose approximately 50% over time (Figure 43d), from a risk of 0.0011 (95% CI 0.0009 to 0.0014) in the first six-month period to 0.0017 (95% CI 0.0014 to 0.0020) in the last six-month period. Among low-income people (<$20,000) the rise was only 10%, from 0.0035 (95% CI 0.0030 to 0.0039) in the first six-month period to 0.0038 (95% CI 0.0033 to 0.0042) in the last six-month period. Thus there was some evidence of convergence in mortality risk by equivalised household income over time among 45–64 year old females, consistent with some drift health selection.

As well as interpreting the income mortality risk plots in isolation, they should also be interpreted relative to the 'base-line' plots by NZDep91 (Figure 29, page 209) and highest qualification (Figure 33, page 219). Such a comparison suggests that the possible convergence of plots by income among 25–44 year olds was not seen in the baseline analyses. The pattern of plots for 45–64 year old males was not notably different by socioeconomic factors. Among females aged 45–64, there was a rising mortality risk over time among those with tertiary education, warning against over-interpretation of the rising mortality risk among high income females in Figure 43d.

Two additional problems with the mortality risk plots limit their interpretation with regard to drift health selection. First, it may be that over a longer period of follow-up the mortality risk plots would have more convincingly demonstrated drift health selection. Second, drift health selection will only apply to the component of the income–mortality association that is not due to confounding by other factors. The relative height of the lines in the above plots over time is confounded by factors such as age and education, and it is likely that these factors will exaggerate the distance between the low-income and high-income plots. Thus, for drift health selection to explain the unconfounded association of income with mortality in a short-duration study such as the NZCMS, does not require that the line-plots fully converge. Rather, they would only have to 'converge' to a point where the remaining differences in the height of the lines were due to confounding of the association of income with mortality by factors other than health status. Therefore, the plots for 25-45 year old males and 45–64 year old females (Figures 43a, d) might be more suggestive of drift health selection than they first appear.

Cancer and cardiovascular disease deaths among 45–64 year olds: all labour force categories, and excluding non-active labour force

Death from cancer is usually preceded by a period of poor health. Therefore, if drift health selection was truly biasing the income–mortality gradients, we would expect to a health selection pattern more clearly among cancer deaths. Figures 45a and 45c below show the cancer mortality risk plots over time for 45–64 year old males and females in all labour force categories. As with all-cause mortality, there was not compelling evidence of drift health selection among males, but some suggestion of drift health selection among females. Certainly, the pattern of slopes was not notably different from those for all-cause mortality, which would have been expected if there was drift health selection. Furthermore, other than some suggestion of convergence for 45–64 year old females, the pattern of *slopes* for the cancer mortality plots was not notably different from that for the baseline socioeconomic factors where drift health selection should not theoretically be operating – small area deprivation (Figure 30) and education (Figure 34).

As with the cancer plots, the cardiovascular mortality risk plots for 45–64 year old males were also not particularly suggestive of drift health selection (not shown).

Figure 45: Cancer mortality risk for each six-month period following census night by equivalised household income for 45–64 year olds, all labour force categories and excluding the non-active labour force

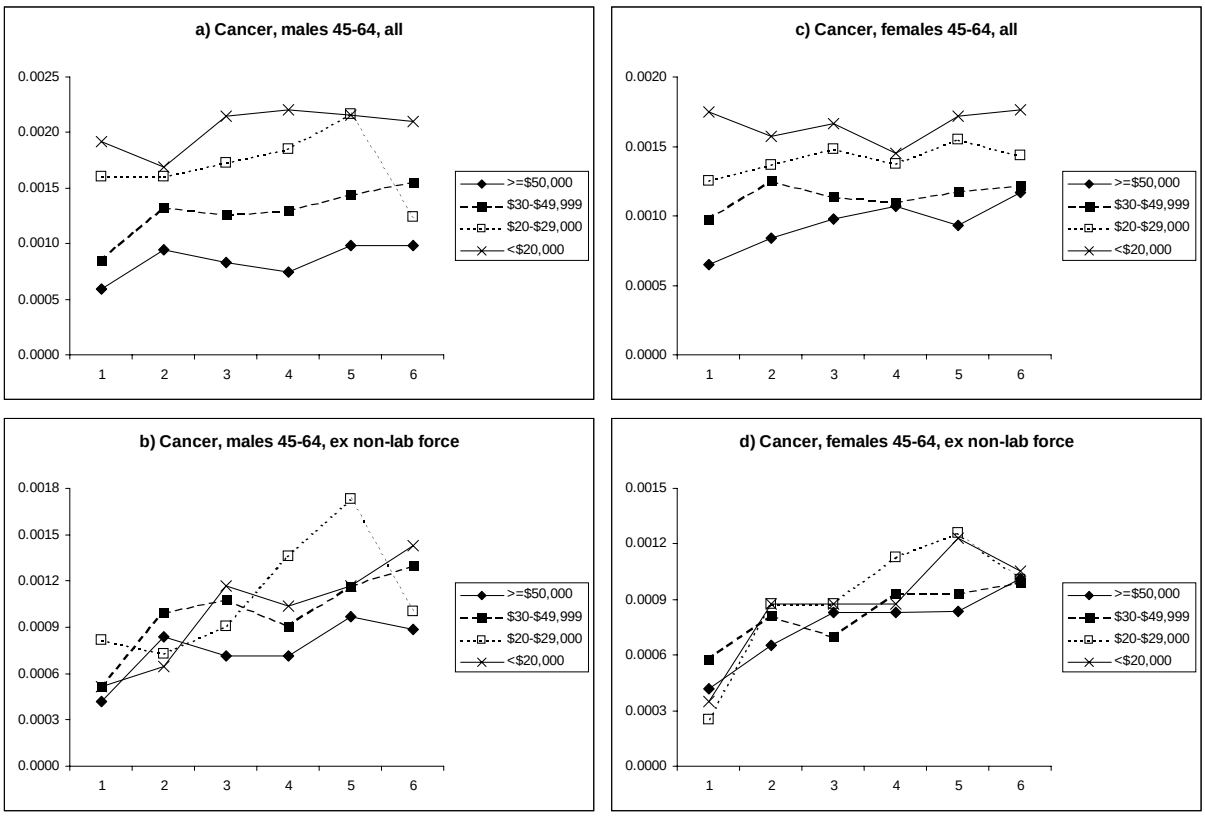
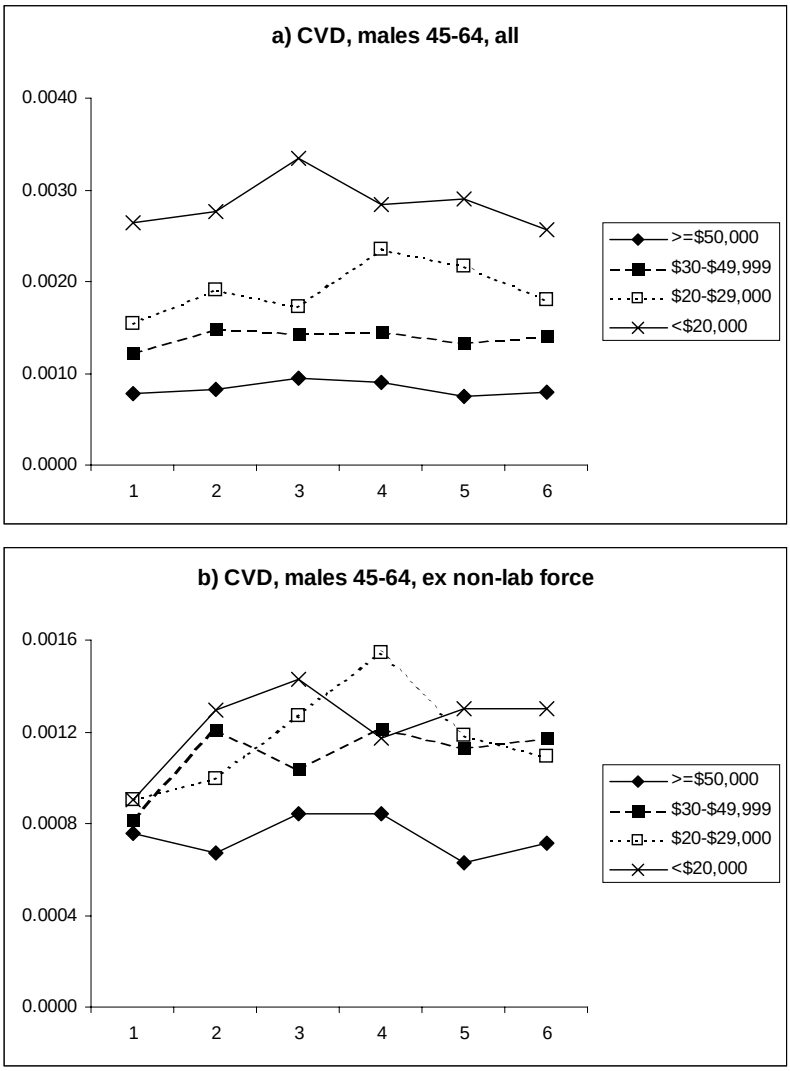
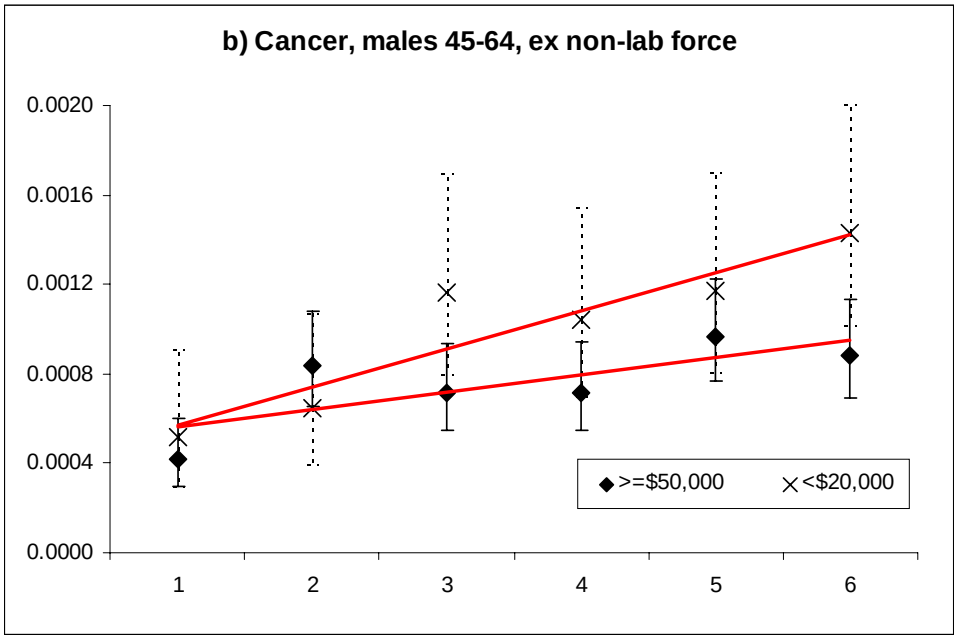


Figure 46: Cardiovascular disease mortality risk for each six-month period following census night by equivalised household income for 45–64 year old males, all labour force categories and excluding the non-active labour force



Mortality risk plots, excluding the non-active labour force for 45–64 year old male and female cancer deaths, are shown in Figures 45b and 45d above, respectively, and for 45–64 year old male cardiovascular disease deaths in Figure 46b. Among 45–64 year old male cancer deaths, there was some evidence of steeper slopes for the low-income strata compared to the high-income strata. Figure 47 presents the plots for just those with an equivalised household income of greater than or equal to $50,000, and those less than $20,000, to more clearly demonstrate the differing slopes. (If deaths in the first six months were deleted from Figure 47 (consistent with the cohort analyses) the difference in slopes became more convincing.) There was little evidence of differing slopes for 45–64 year old female cancer deaths and for 45–64 year old male cardiovascular disease deaths. Thus differential health selection by income was only supported for 45–64 year old male cancer deaths. Previous plots in this report also suggested differential health selection for 45–64 year old male cancer deaths for education (Figures 34b and 35), but not for deprivation (Figure 30b).

Figure 47: Cancer mortality risk for each six-month period following census night for 45–64 year old males with high and low equivalised household income, excluding the non-active labour force



6.2.2 Excluding sickness beneficiaries

The association of equivalised household income with all-cause mortality before and after, excluding sickness beneficiaries, is shown in Table 62. The percentage reduction to the null for the excess odds ratio after excluding sickness beneficiaries, comparing people with an equivalised household income over $50,000 to less than $20,000, ranged between 11% to 32% for the four sex by age groups. The percentage reductions among 45–64 year olds were similar to those for small area deprivation (Table 57) and highest qualification (Table 59). However, the percentage reductions among 25–44 year old males (32%) and females (27%) were greater than those for small area deprivation (14% and 18%) and highest qualification (10% and 11%). Thus, these results are suggestive of some *drift* health selection affecting the 25–44 year old income–mortality gradients. An alternative explanation is that equivalised household income is more highly correlated with receipt of a sickness benefit among 25–44 year olds than either small area deprivation or highest qualification.

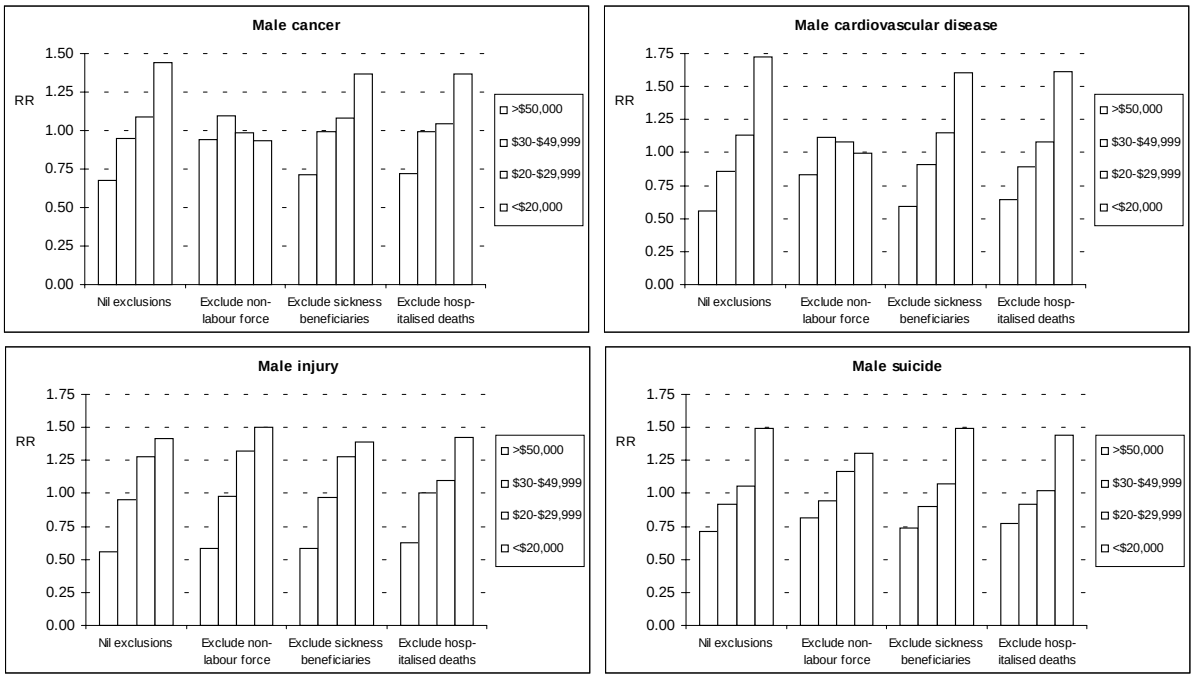
Table 62: Odds ratios of all-cause mortality for 25–64 year olds in the restricted cohort, by household income, for various exclusions testing for health selection

Exclusion criteria	Odds ratios (ref group = <$20,000)				% change to null of ≥$50 OR †
	≥$50,000	$30–$49,999	$20–$29,999	<$20,000	
Males					
25–44 year olds					
Nil	0.65 (0.55–0.78)	0.81 (0.68–0.96)	0.94 (0.79–1.10)	1.00	
Sickness beneficiaries	0.76 (0.63–0.92)	0.94 (0.78–1.13)	1.06 (0.89–1.27)	1.00	32%
Pre-hospitalised deaths	0.74 (0.60–0.91)	0.95 (0.77–1.16)	0.92 (0.75–1.12)	1.00	24%
Non-labour force	0.81 (0.66–0.98)	0.95 (0.78–1.15)	1.05 (0.87–1.27)	1.00	44%
45–64 year olds					
Nil	0.54 (0.49–0.59)	0.77 (0.70–0.83)	0.89 (0.83–0.96)	1.00	
Sickness beneficiaries	0.61 (0.56–0.67)	0.86 (0.79–0.94)	0.92 (0.85–1.00)	1.00	17%
Pre-hospitalised deaths	0.61 (0.55–0.68)	0.81 (0.73–0.90)	0.88 (0.80–0.97)	1.00	16%
Non-labour force	0.75 (0.67–0.84)	1.05 (0.93–1.17)	1.01 (0.90–1.14)	1.00	47%
Females					
25–44 year olds					
Nil	0.70 (0.56–0.86)	0.71 (0.57–0.88)	0.91 (0.74–1.12)	1.00	
Sickness beneficiaries	0.78 (0.62–0.97)	0.78 (0.62–0.98)	0.90 (0.72–1.12)	1.00	27%
Pre-hospitalised deaths	1.00 (0.75–1.33)	0.94 (0.69–1.27)	0.84 (0.61–1.15)	1.00	100%
Non-labour force	0.89 (0.69–1.16)	0.83 (0.63–1.10)	0.96 (0.72–1.28)	1.00	65%
45–64 year olds					
Nil	0.66 (0.59–0.74)	0.77 (0.69–0.86)	0.95 (0.87–1.04)	1.00	
Sickness beneficiaries	0.70 (0.63–0.78)	0.80 (0.72–0.90)	0.97 (0.89–1.07)	1.00	11%
Pre-hospitalised deaths	0.76 (0.66–0.87)	0.79 (0.68–0.91)	0.89 (0.78–1.00)	1.00	29%
Non-labour force	0.83 (0.70–0.99)	0.88 (0.73–1.06)	0.97 (0.80–1.18)	1.00	51%

Note: Injury and suicide deaths are not presented for females due to small numbers. For all remaining cells in the table there were at least 30 deaths.

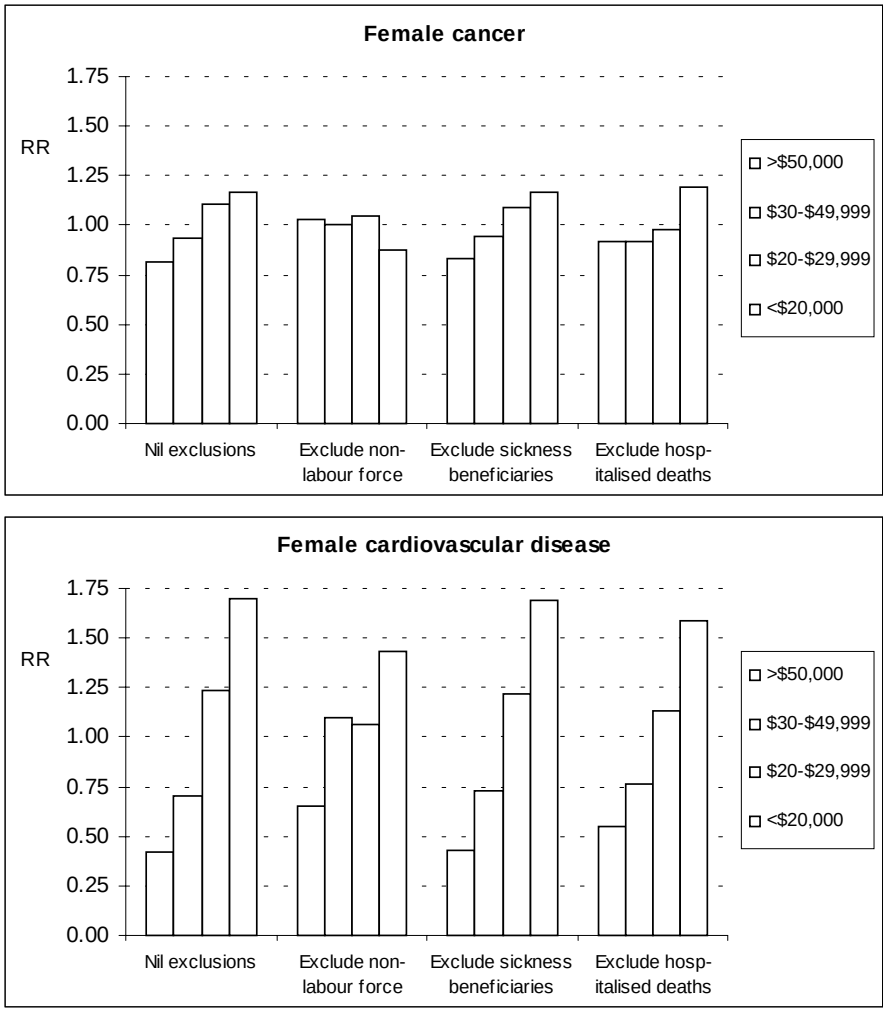
† Percentage change is for the excess odds ratio for quintile 5 compared to quintile 1, compared to the same odds ratio with nil exclusions.

Figure 48: Risk ratios of cause-specific mortality for 25–64 year old males in the restricted cohort, by household income, for various exclusions testing for possible health selection



By cause of death for 25–64 year olds combined (Figures 48 and 49), there was little evidence of any substantive diminution of the income–mortality gradient following the exclusion of sickness beneficiaries for any of the examined causes of death – except perhaps for male cancer and cardiovascular disease. However, this was the same pattern as that for small area deprivation (Figure 31), failing to strongly suggest drift health selection as a bias affecting income mortality gradients.

Figure 49: Risk ratios of cancer and cardiovascular disease mortality for 25–64 year old females in the restricted cohort, by household income, for various exclusions testing for possible health selection



6.2.3 Excluding decedents with a hospitalisation event between 1988 and census night

The association of equivalised household income with all-cause mortality before and after excluding deaths hospitalised between 1988 and census night is also shown in Table 62. The percentage reduction to the null for the excess odds ratio, comparing people with an equivalised household income over $50,000 to less than $20,000, ranged between 16% to 100% for the four sex by age groups. Only the reduction to the null for the 25–44 year old females (100%) was substantially greater than the comparable reductions for small area deprivation (17%; Table 57) and highest qualification (4%; Table 59). However, the number of deaths among 25–44 year olds (n=651), particularly after exclusion of pre-hospitalised deaths (n=324), was the smallest of all four sex by age groups. Consequently, the 95% confidence intervals about the odds ratios for 25–44 year old females after exclusion of pre-hospitalised deaths are wide (Table 62). Thus, as with the exclusion of sickness beneficiaries, there was not convincing evidence of *drift* health selection affecting the all-cause mortality gradients by income for the sensitivity analyses excluding pre-hospitalised deaths.

By cause of death for 25–64 year olds combined (Figures 48 and 49 above), the pattern for the exclusion of pre-hospitalised deaths was very similar to that for the exclusion of sickness beneficiaries described in the previous section. It was also very similar to that for pre-hospitalised deaths in the NZDep91 analyses (Figure 31, page 213).

6.3 Excluding the non-active labour force

Income is highly dependent on labour force status – to have a high household income, at least one person in the household needs a high-paying job. This 'necessary' relationship between labour force status and NZDep91 or education does not exist. However, if income were truly associated with mortality, then we would still expect to see a strong association among the active labour force where there is large variation in incomes. Put in epidemiological terms there is a strong association between labour force status and income, but it is not an exact concordance. Given the strong association of labour force status with mortality risk, then the income–mortality association should be more prone to confounding by labour force status than either the education–mortality or deprivation–mortality gradients. If present, this confounding will be disclosed when restricting the analyses to the active labour force – presumably for all specific causes of death associated with income.

On the other hand, income is theoretically the socioeconomic factor in the NZCMS most likely to be affected by drift health selection. As labour force status is a proxy for, among other things, health status, large reductions in the income–mortality association when restricting analyses to the active labour force would also be suggestive of health selection – particularly if larger reductions were evident for causes of death preceded by poor health.

6.3.1 All-cause mortality

For each sex by age group, the income mortality gradient was approximately halved following exclusion of the non-active labour force. The reductions to the null for the ≥$50,000 compared to <$20,000 odds ratio for all-cause mortality following exclusion of the non-labour force (Table 62 above) were:

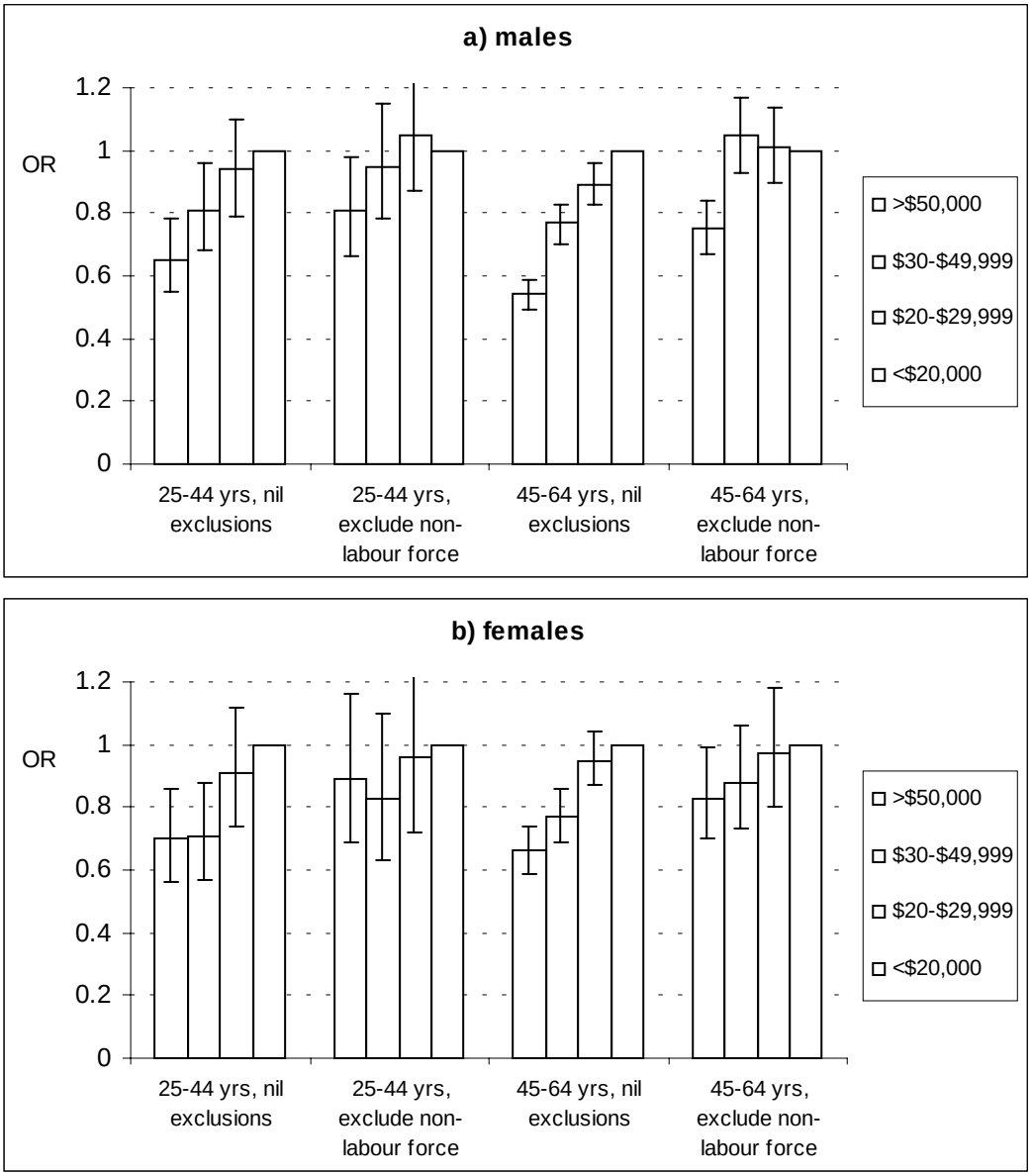
- 44% among 25–44 year old males, compared to 14% for the quintile 5 compared to quintile 1 deprivation (Table 57) and 14% for the tertiary qualification compared to nil qualification (Table 59) reductions to the null following exclusion of the non-labour force, respectively
- 47% among 45–64 year old males, compared to 43% and 22% for deprivation and highest qualification
- 65% among 25–44 year old females, compared to 32% and 25% for deprivation and highest qualification
- 51% among 45–64 year old females, compared to 53% and 40% for deprivation and highest qualification.

Figure 50 (below) plots the income mortality gradient for each sex by age group, before and after excluding the non-labour force, using the odds ratios and 95% confidence intervals in Table 62, page 247. For males, the figure highlights that after excluding the non-labour force only the highest income category (≥$50,000 equivalised household income) had a notable and statistically significantly lower mortality risk than the poorest household income category (<$20,000). Moreover, there was essentially little difference between the three income categories up to $50,000. For females, a suggestion of a gradient remains after excluding the non-labour force, but the 95% confidence intervals nearly all included 1.0. For both males and females, the patterns shown in Figure 50 may be affected by linkage (and selection) bias, such that a small gradient might have remained after excluding the non-labour force if there had been no linkage bias. However, on balance, linkage bias would have been unlikely to alter the substantive interpretation that excluding the non-labour force dramatically reduces the income mortality gradient – more so than for NZDep91 and highest qualification.

6.3.2 Cause-specific mortality

The effect of excluding the non-labour force on the income mortality gradient varied markedly by cause of death. The crude floating risk ratios for 25–64 year olds combined for all labour force categories and excluding the non-active labour force are shown in Figure 48, page 248 for males and Figure 49, page 249 for females. As with the comparable figure using crude risk ratios for various exclusions for small area deprivation (Figure 31), there is some confounding by age and ethnicity. For example, a stronger association of income and cancer is depicted by the crude risk ratios in Figure 48 for males than in the age and ethnicity-adjusted odds ratios in Table 40. While this confounding means that the gradients shown in Figure 48 and Figure 49 are not exactly correct, it does not invalidate a comparison of the relative changes in the gradient for the various exclusions.

Figure 50: Odds ratios of all-cause mortality for 25–44 and 45–64 year old males and females by equivalised household income for the restricted cohort with no exclusions and excluding the non-labour force



Note: Error bars are 95% confidence intervals.

The main finding in Figure 48 for males was that the exclusion of the non-labour force dramatically reduced the gradient for cancer and cardiovascular disease, moderately reduced the gradient for suicide deaths, but did not alter the unintentional injury gradient. This pattern is consistent with that observed for the small area deprivation gradients after excluding the non-labour force (Figure 31), although the reductions of the cancer and cardiovascular disease gradients for income were greater than for deprivation.

6.4 Conclusion

There was a strong univariate association of equivalised household income with all-cause mortality, and most specific causes of mortality. These associations were probably not greatly affected by selection bias, and somewhat underestimated due to linkage bias (see Section 6.1).

The two major difficulties for the income analyses were determining: (a) whether health selection affected the income–mortality gradients, and (b) what was causing the income–mortality gradients to decrease dramatically following the exclusion of the non-active labour force.

Theoretically, there were two possible types of health selection: *drift* health selection and *differential* health selection. Further, we would only expect health selection to operate for causes of death where a period of poor health is common before death (eg, cancer and cardiovascular disease). Finally, the health selection is framed as being a bias over the short-term (ie, a couple of years). The mortality risk plots over time in this chapter found some evidence of *differential* health selection for the association of income with cancer among males, but not for the association of income with other causes of death. Regarding *drift* health selection, there was some occasional evidence from the mortality risk plots and exclusions of sickness beneficiaries and pre-hospitalised deaths, but it was patchy, inconsistent, and not usually notably different from the 'baseline' NZDep91 and education analyses. Thus, the tests of health selection in this chapter, and compared to baseline analyses in previous chapters, did not strongly suggest health selection *over the short term*.

How can this conclusion that the income–mortality association was little affected by health selection be reconciled with the larger reductions in the income–mortality gradient following exclusion of the non-active labour force than for NZDep91 and highest qualification? In short, not easily.

There are five possible ways that excluding the non-active labour force might decrease the income–mortality association:

1 *Drift health selection, whereby labour force status is a proxy for health status.* While the above sensitivity analyses suggested little drift health selection by income in the short-term, it may be that:

   - the 'tests' for short-term drift health selection used were too crude
   - drift health selection was acting over a longer period than the three years observable in the NZCMS
   - while short-term drift health selection of the income–mortality association was modest, it was enough in combination with the reasons listed below to drive the income–mortality association further to the null that the associations of other socioeconomic factors with mortality.

2 *Differential health selection, whereby labour force status is a proxy for health status.* This may also have been one reason, but certainly not the major reason as discussed above. Also, differential health selection, if present, should apply to all socioeconomic factors – not just income.

3 *Confounding by labour force status* (by means other than short-term drift health selection). Such confounding was undoubtedly occurring, and for good reason, given the strong correlation between labour force status and income. But if it was the major reason for the reduction of the income-gradient, more notable reductions for the injury (and suicide) gradients – not just the cancer and cardiovascular disease gradients – would have been expected.

4 *As a proxy for health status – an intermediary variable between income and mortality.* This mechanism is different from drift health selection, where health status actually influences the socioeconomic factor of interest (ie, reverse causation). Rather, this possible mechanism is common to all socioeconomic factors, and involves over-controlling the association of a socioeconomic factor with mortality by the inclusion of a proxy for an intermediary variable (ie, health status). It is not clear why this mechanism would reduce the income gradient more than, say, the education gradient.

5 *Effect modification of the income–mortality gradient by labour force status* by mechanisms other than differential health selection, such that the gradient is weaker among the active labour force. It is not clear why this mechanism should be more important for income than other socioeconomic factors.

On balance, and putting aside the fifth reason, it seems likely that a *combination* of the first four reasons explains why the income gradient diminishes dramatically when excluding the non-active labour force. There was a similar pattern of varying reduction in the cause-specific mortality gradients by NZDep91 (Section 1.4) – although not as marked as that for income. Thus, a moderate variation in the mix of the first four reasons between NZDep91 and income may have been enough to make the difference. It is interesting to speculate that a modest amount of drift health selection for income, either over a period longer than three years or simply not reliably detected by the sensitivity analyses in this report, may be enough to make the difference between the NZDep91 and income analyses. However, it is impossible to be more precise in drawing conclusions.

Glossary

General terms

Differential health selection

(See glossary definition of **health selection** first.)

Health selection that occurs when a socioeconomic mortality gradient is assessed among the active labour force only, due to exclusion from the active labour force that is differential by that socioeconomic factor.

Classically, differential health selection has been described for occupational class mortality gradients. Here the gradient is underestimated when only *current* occupation is available for the assignation of occupational class. This underestimation is because the lower occupational classes (based on *usual* occupation) are more likely to be forced out of the labour force than the higher occupational classes when sick, causing the observed mortality risk/rate among the lower occupational classes (based on *current* occupation) to be underestimated.

Drift health selection

(See glossary definition of **health selection** first.)

Health selection that occurs when people drift down the socioeconomic ladder consequent on their health status. For example, poor health may both lower one's (current) income and be associated with an increased risk of death, causing an overestimate of the association of (usual) income with mortality.

Health selection

'... the artificial raising or lowering of the average health of people with a particular characteristic associated with the process by which that characteristic is acquired or lost. The mortality of a population with that characteristic is affected by health-related mobility if the health of people acquiring or losing the characteristic differs systematically from others with the characteristic.' (Fox et al 1987).

In this report, it is crucial to consider two types:

- **Drift health selection** (see glossary definition)
- **Differential health selection** (see glossary definition).

It is also useful to consider health selection (either of the two types above) as occurring over the short or long term. With three years of follow-up in the NZCMS, it was only possible to investigate short-term health selection.

Linkage bias

The biases by demographic and socioeconomic factors in the proportion of mortality records linked to a census record.

Multivariate analysis

In this report, regression analyses that include more than one socioeconomic factor as independent variables.

Univariate analysis

'It is increasingly common in the medical literature to use the term *univariate analysis* to refer to analyses which examine only a single explanatory variable's relationship to a [single] response [outcome] variable.' (Armitage and Colton 1998, p.4663). In this report, age and ethnicity is treated as fundamental demographic variables that must be adjusted for. Therefore, 'univariate analysis' in this report refers to the association of one socioeconomic factor (eg, income) with mortality, controlling for age and ethnicity. (All analyses were conducted separately by sex.) (See **multivariate analysis**.)

Glossary of record linkage terms

Agreement frequency ratio or **agreement odds**

The agreement frequency ratio is the odds of a particular matching variable agreeing among **links** versus agreeing among **non-links**, that is the *m* probability divided by the *u* probability.

Agreement weight

In record linkage using **Automatch®**, the weight is the logarithm (base two) of the **agreement frequency ratio**.

Automatch®

Commercial probabilistic record linkage software used in the NZCMS.

Blocking

A procedure used in record linkage to reduce the number of possible comparisons. That is, the records on both files are divided into blocks (eg, area of residence), and record linkage is conducted within these blocks only.

Blocking variable

Variable used to 'block' files in record linkage. In the NZCMS, the blocking variables were geocodes: [meshblock] and [census area units].

Combined frequency ratio or **combined odds**

For any given **comparison pair**, the product of the **agreement frequency ratios** for each **matching variable** that agrees and the **disagreement frequency ratios** for each matching variable that disagrees. The combined frequency ratio constitutes the information on which the overall 'betting odds' in favour of (or against) a correct match are based, and hence pairs categorised as links or non-links.

Combined weight

For any given **comparison pair**, the sum of the **agreement weights** for each **matching variable** that agrees and the **disagreement weights** for each matching variable that disagrees. The combined weight constitutes the information on the relative weight in favour of (or against) a correct match, and hence pairs are categorised as links or non-links.

Disagreement frequency ratio or **disagreement odds**

The disagreement frequency ratio is the odds of a particular matching variable disagreeing among **links** versus disagreeing **non-links**, that is [1 minus the *m* probability] divided by [1 minus the *u* probability].

Disagreement weight — In record linkage using **Automatch®**, the weight is the logarithm (base two) of the **disagreement frequency ratio**.

False link or **false positive** — As conceptualised in the **linkage bias** analysis in the NZCMS, those census respondents who did not die but were linked.

False non-link or **false negative** — As conceptualised in the **linkage bias** analysis in the NZCMS, those census respondents who did die but were not linked.

Files — The sets of records to be compared in the record linkage – in the NZCMS the mortality and census files.

Frequency ratio — (See 'agreement frequency ratio' and 'disagreement frequency ratio'.)

Linkage bias — As operationalised in the NZCMS, the *misclassification bias of the mortality outcome due to the record linkage*. Taking all census records as the population, one can create a two-by-two table of **links** (equivalent to the 'diagnostic test' for vital status) by actual vital status. It is then possible to consider the **sensitivity**, **specificity**, **positive predictive value** and **negative predictive value** of the record linkage for vital status.

Links — Those comparison **pairs** that are categorised during the record linkage as *probably* including the same person's mortality and census record. As people die only once, each mortality record can only be linked to one census record. In most record linkage projects (including the NZCMS) it is impossible to determine which links are **matches** or **non-matches**, although it is assumed that the majority (hopefully the vast majority) of links are matches.

***m* probability** — The probability that a **matching variable** agrees given that the comparison pair being examined is categorised as a **link**. It may be global (eg, one common *m* probability for all values of [year of birth]) or value specific (eg, different *m* probabilities for each value of [year of birth]).

Match — Following Newcombe (1988), a **pair** of mortality and census records that are for the same person (ie, the pair is 'correct'). It is usually impossible to verify in a record linkage project whether any pair is indeed a match – rather, the relative probability of a pair being a match is estimated.

Matching variable — Variable that is common to both the mortality and census files, and hence available for comparing pairs of records during the record linkage. In the NZCMS, matching variables were [day of birth], [month of birth], [year of birth], [sex], [ethnicity], [country of birth].

Match-run — The full series of **passes** that make up the record linkage.

Non-links — Those comparison **pairs** that are categorised during the record linkage as *probably not* including the same person's mortality and census record.

Non-match	Following Newcombe (1988), a **pair** of mortality and census records that are *not* for the same person (ie, the pair is 'incorrect').
Pair or **comparison pair**	Any comparison of a pair of **records** from different **files**. The theoretical number of pairs is the product of the number of records on the two files (eg, if there were 100 mortality records and 10,000 census records, the number of possible pairs is 1 million). **Blocking** dramatically reduces the number of possible pairs.
Pass	A given specification of matching variables, blocking variable(s), mortality and census records to be processed, *m* and *u* probabilities, and other parameters set by the operator of **Automatch®**. A sequence of passes (up to eight in Automatch®) makes up a **match-run**.
Positive predictive value	As conceptualised in the **linkage bias** analysis in the NZCMS, the proportion of the linked census records who did die.
Records	In the NZCMS, the separate mortality events on the mortality **file** and the separate census entries of the census file.
Sensitivity	As conceptualised in the **linkage bias** analysis in the NZCMS, the proportion of the census cohort who did die and who were linked during the record linkage.
Specificity	As conceptualised in the **linkage bias** analysis in the NZCMS, the proportion of the census cohort who did not die and who were not linked during the record linkage.
True link or **true positive**	As conceptualised in the **linkage bias** analysis in the NZCMS, those census respondents who died and who were linked.
True non-link or **true negative**	As conceptualised in the **linkage bias** analysis in the NZCMS, those census respondents who did not die and who were not linked.
***u* probability**	The probability that a **matching variable** agrees given that the comparison pair being examined is categorised as a **non-link** (ie, the probability that variables agree purely by chance among non-links). It may be global (eg, one common *u* probability for all values of [year of birth]) or value specific (eg, different *u* probabilities for each value of [year of birth]).
Weight	See **agreement weight** and **disagreement weight**.