

Socio-economic position and mortality risks: a sibling approach

Analysis using the New Zealand Longitudinal Census (NZLC)

By

Fui Swen Kuh



A dissertation submitted in partial fulfilment of the requirements for the degree of

Bachelor of Science (Honours) in Statistics

Supervised by Dr. Barry Milne and Professor Alan Lee

The University of Auckland, 2015

ISBN 978-0-473-42503-6 (online)

“The quiet statisticians have changed our world; not by discovering new facts or technical developments, but by changing the ways that we reason, experiment and form our opinions.”

–Ian Hacking

Abstract

The use of sibling analysis in epidemiological studies can be a powerful tool to account for unmeasured within-family confounders like common childhood environment, genetic makeup and cultural background (Donovan & Susser, 2011). In this study, we perform a sibling analysis investigating the effect of life-course socio-economic influences on mortality. An individual cohort was also analysed to assess the robustness of the results from our sibling analysis. The study was done on two cohorts — 51,267 siblings from the sibling cohort and 170,697 individuals from the individual cohort that were identified from the 1981–2006 New Zealand Longitudinal Census (NZLC) linked data records, which are also linked with mortality records of 2006–2011 from the New Zealand Census Mortality Study (NZCMS).

Three series of Cox regression model were fitted to the two cohorts adjusting for age and gender, among other covariates to estimate the mortality risks for individuals who have different socio-economic positions.

The results show that for both siblings and unrelated individuals, there are increased mortality risks for those who have lower socio-economic positions in midlife. Significant results were obtained for both siblings and unrelated individuals for those who were in lowest household income quintile at least *twice or more* compared to those who were never in lowest household income quintile between the years 1981–2006. No significant results were found for siblings when comparing those who were in lowest household income quintile *once* to those who were *never* in lowest household income quintile, suggesting the significant effects found between unrelated individuals may be confounded by early childhood background.

Disclaimer: Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the author, not Statistics New Zealand.



Acknowledgements

Firstly, I would like to express my gratitude to my supervisors **Dr. Barry Milne** and **Professor Alan Lee** for the constant guidance and support throughout the year, humour and croissants.

I am also grateful to my family; **mum** and **dad** for their unlimited care and the support in pursuing my passion in Statistics, sisters, **Leely** and **Chin** for the yummy meals so I could spend more time on course work and also putting together the dissertation.

I would also like to thank:

Professor Brian D’Onofrio from Indiana University for guidance on how to do sibling fixed effects analyses.

The team at COMPASS; **Peter, Jessica, Martin, Roy, Kevin** and **Nichola** for the conducive environment to do all my research work in and for being my go-to team whenever I have questions; invitations to tea and baby shower party; and not forgetting the team behind the ‘closed curtains’ – **Liza**, for the free-flow supply of tea, advice, help and support in both course work and dissertation; and **Justin**, for working as a team together in all SAS and non-SAS discussions.

Statistics New Zealand, for providing the data and especially the Microdata Access Team – **Amily, Natalie** and **Dariusz** for the speedy checks of all of my Data Lab outputs.

Statistics students in 303S-299 especially **Natalia** and **Banu** for help with proofreading, and all other fellow students for all the stimulating discussions, chats, good times, and late nights that added the fun to the challenging and daunting Honours year; and

My friends whom I should be spending more time with but nevertheless time spent together escaping from studies are still precious. Also, shout out to **Chris** and **Kam** for all the mathematical and non-mathematical help and encouragement, tea breaks and chats.

Last but not least, friends that I met at ANU to be able to share our academic research experiences and my supervisors over the summer, **Anton** and **Grace** for all their kind comments, supportive words and giving me the opportunity to have first taste of statistical research.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Advantages and disadvantages of sibling analysis.....	2
2	Data Preparation	3
2.1	Data Description	3
2.1.1	New Zealand longitudinal census (NZLC).....	3
2.1.2	New Zealand Census Mortality Study (NZCMS).....	4
2.2	Measures.....	4
2.2.1	Cohorts.....	4
2.2.2	Variables	7
2.2.3	Outcome	10
2.2.4	Bias weights.....	10
2.3	Data set for analyses	11
3	Methods	12
3.1	Cox Regression	12
3.1.1	Cox regression in SAS.....	14
3.2	Cox Regression with Fixed Effects.....	15
3.2.1	Cox Regression with Fixed Effects in SAS	16
4	Results	18
4.1	Individual cohort.....	18
4.2	Sibling cohort without fixed effects	22
4.3	Sibling cohort with fixed effects	26
4.4	Summaries	31
4.5	Diagnostics.....	32
5	Discussion	34
5.1	Strengths and limitations.....	35
5.2	Further work.....	35
6	Bibliography.....	37

7	Appendix.....	39
7.1	Appendix A.....	39
7.2	Appendix B.....	40
	Model Specification	40
7.3	Appendix C.....	41
7.4	Appendix D.....	42
	Reading in data	42
7.5	Appendix E.....	58
	Model.....	58
	Survival curves.....	62
7.6	Appendix F.....	69
	Diagnostics in R.....	69

List of Tables

Table 1. SEP predictors, definition, variable names and its categories.....	8
Table 2. Grouped covariates, its description, variable name and categories.....	9
Table 3. Definition, variable name and range values of bias weights	11
Table 4. Hazard ratios of individuals in individual cohort relative to individual who were never in lowest household income quintile	19
Table 5. Hazard ratios of individuals in individual cohort relative to individual who were never in lowest NZDep.....	21
Table 6. Hazard ratios of individuals in sibling cohort relative to individual who were never in lowest household income quintile	23
Table 7. Hazard ratios of individuals in individual cohort relative to individual who were never in lowest NZDep.....	25
Table 8. Hazard ratios of a sibling relative to sibling who was never in lowest household income quintile	28
Table 9. Hazard ratios of low_inc when compared to sibling/individual who was never in lowest household income quintile from three series of models on two cohorts...31	
Table 10. Hazard ratios of low_nzdep when compared to sibling/individual who was never in lowest NZDep from three series of models on two cohorts	31

List of Figures

Figure 1. The identification of individual and sibling cohort from 1981 Census data.....	5
Figure 2. Age distribution of individual cohort.....	6
Figure 3. Age distribution of sibling cohort.....	6
Figure 4. Number of times proportion of individual cohort in lowest household income quintile.....	8
Figure 5. Number of times proportion of sibling cohort in lowest household income quintile.	8
Figure 6. Timeline of 1981-2006 longitudinal records of individuals from NZLC and 2006-2011 mortality records from NZCMS	10
Figure 7. Merging dataset from different sources ready for analyses	11
Figure 8. Example of censoring in our data – most individuals are censored at 61 months where the dotted lines are the individual’s true life span.....	13

Figure 9. Hazard ratios of an individual (in individual cohort) in lowest household income quintile relative to individual who were never in lowest household income quintile after adjusting for groups of covariates sequentially	18
Figure 11. Survival curves of an individual from the individual cohort for all categories of low_inc with mean values for all other covariates in the model	20
Figure 12. Hazard ratios of individuals in individual cohort in lowest NZDep adjusting for groups of covariates sequentially	21
Figure 13. Hazard ratios of siblings with robust variance estimation in lowest household income quintile relative to sibling who was never in the lowest household income quintile adjusting for groups of covariates sequentially	22
Figure 14. Survival curves of siblings for all categories of low_inc and mean values for all other covariates in the model	23
Figure 15. Hazard ratios of individuals in the sibling cohort with robust variance estimation in lowest NZDep adjusting for groups of covariates	24
Figure 16. Hazard ratios of siblings in lowest household income quintile adjusting for groups of covariates.....	26
Figure 17. Log-hazard ratios of siblings in lowest household income quintile adjusting for groups of covariates.....	27
Figure 18. Hazard ratios of a sibling relative to sibling who was never in lowest NZDep adjusting for groups of covariates sequentially	30
Figure 19. Schoenfeld residuals plot for low_inc against months with fitted loess	32
Figure 20. Deviance residuals plot against linear predictors included in model with fitted loess.....	33

Abbreviations

NZ	New Zealand
SEP	Socio-economic position
SNZ	Statistics New Zealand
NZLC	New Zealand Longitudinal Census
NZCMS	New Zealand Census Mortality Study
NZDep	New Zealand Deprivation Index
NZSEI	NZ Socio-Economic Index
HR	Hazard Ratio
CI	Confidence Interval

Chapter 1

Introduction

Whether we like it or not, death is inevitable — it will happen to every one of us regardless of our social, cultural and economic background; but why might some fare worse than others in terms of mortality risk? Like many other countries, mortality inequalities exist in New Zealand (NZ). The inequalities can be driven by many factors — biological, lifestyle, socio-economic, societal or environmental factors (AIHW, 2014). Socio-economic position (SEP) is commonly defined as the economic and sociological measure of a person's social position in relation to others, and is usually measured in terms of a person's income, education, occupation or a combination of any of those. It is shocking to note that mortality rates have been decreasing overall in NZ while 'relative socio-economic inequalities in mortality increased or plateaued' (Blakely, 2007). Among all these factors, socio-economic position throughout the life course is especially known to affect our mortality outcomes (Kuh & Shlomo, 2004). Life course SEP looks at the processes throughout a person's life time, as opposed to their SEP at any one time.

1.1 Motivation

Results from research done in other countries like Norway and Sweden (Næss et al., 2004; Padyab et al., 2013) point us towards the importance of life-course socio-economic experiences on mortality risk. The effect of socio-economic positions on mortality has been studied proximally in New Zealand but not over the life course. The life-course approach to investigate socio-economic influences on mortality would provide us with a better understanding of whether the experience of disadvantaged socio-economic circumstances across the life-course more strongly relates to mortality than the experience of disadvantaged socio-economic circumstances at any one time, in NZ. We include the use of sibling as comparison to account for possible within-family confounding variables in the life course approach in our study.

The recent development of the New Zealand Longitudinal Census (NZLC) data which links individuals across the 1981–2006 NZ censuses, and also linkage to mortality records in the New Zealand Census Mortality Study (NZCMS) make it possible for us to assess life-course socio-economic factors against mortality. Similar work using the life-course approach has been done in other countries, but not many have used siblings as comparisons, especially not yet in New Zealand.

1.2 Advantages and disadvantages of sibling analysis

We are interested in looking at the mortality risk for siblings who have different SEP. When running statistical analyses, we want to control for confounders wherever possible. Comparisons of two individuals on their socio-economic positions and mortality risk will be confounded by factors like family background, genetic makeup and parental characteristics. Sibling analysis allows us to partly control for confounders like biological and early-life environmental factors that otherwise could not be accounted for between two individuals. These characteristics make sibling-comparison design common in epidemiology studies (Frisell et al., 2012). Twin studies can allow us to have stronger control of genetic makeup but the results are arguably less generalizable to the public than sibling design (Class et al., 2014).

However, there are certain caveats that are tied with the use of sibling design in our study. The results from using sibling comparisons could be complicated by ‘uncontrolled confounders that vary between siblings, making the results hard to interpret’ (Sudan et al., 2014). The differences between siblings in family context like birth order, gestational age, birth weights, maternal and paternal age are potential confounders that need to be accounted for. Due to the limitations of the data, we can only control for birth order in our analyses.

Chapter 2

Data Preparation

This chapter will briefly describe the two data sources used in this study: the longitudinal census dataset from the New Zealand Longitudinal Census (NZLC) and linked individual mortality with census records from the New Zealand Census Mortality Study (NZCMS). It will also outline how we utilised the data sources to identify the two cohorts for analyses — henceforth referred to as the *individual cohort* and *sibling cohort*. The individual cohort contains individuals from private households who can be linked from 1981–2006. The sibling cohort is a subset of the former cohort, containing siblings where more than one person identify themselves as son or daughter to the occupier. The descriptions of the two cohorts will be included in detail in the following. The chapter also describes how the key variables are extracted to investigate whether discordance in socio-economic position in later life explains different mortality risks.

2.1 Data Description

2.1.1 New Zealand longitudinal census (NZLC)

The NZLC links 1981, 1986, 1991, 1996, 2001, and 2006 censuses into five linked pairs. Pairs of censuses were linked ‘backwards’ (e.g. individuals in 2006 were linked to those in 2001) for all individuals except for those who:

- had not been born at the time of the previous census,
- were living overseas at the time of the previous census, and
- had not filled in a census form and a substitute form was created.

This results in approximately 70% of records at one census having a deterministic link (using key variables like sex and date of birth, with area unit as blocking variable) with records at the previous census. Any records that failed to be deterministically linked are eligible for probabilistic linking. Probabilistic linking

uses a weight to represent the likelihood of two records belonging to the same person and links with weights that are above the cut-off weight are considered as a true link (Statistics New Zealand, 2014).

2.1.2 New Zealand Census Mortality Study (NZCMS)

The New Zealand Census Mortality Study (NZCMS) records consist of data from the censuses that also have been anonymously and probabilistically linked to mortality records for the 3-years immediately after the 1981, 1986, 1991, and 1996 censuses and 5-years immediately after the 2001 and 2006 census (Fawcett et al., 2002; Tan et al., 2012). The records also contain cause of death information obtained from New Zealand Health Information Service mortality data.

Utilising data from the NZLC and NZCMS allows us to construct a prospective cohort study which spans across the 1981–2006 censuses in which exposures at these censuses will be associated with mortality up to 2011.

2.2 Measures

There were four parts to retrieve the appropriate data for the study:

1. Extracting individual and sibling cohorts as our samples
2. Extract main predictor variables and covariates
3. Linked mortality information from NZCMS
4. Linkage bias weights for the two cohorts

2.2.1 Cohorts

2.2.1.1 Sibling and Individual Cohort from 1981 Census Records

The 1981 Census contains about 2.88 million records of individuals in New Zealand. As shown in Figure 1, the *individual cohort* contains individuals from private dwellings in 1981 and a subset of those that are linked across 1981–2006 censuses were identified ($n = 170,697$). The individual cohort is analysed to assess the robustness of the results from our sibling analysis. 82.5% of the cohort consists of individuals aged less than 20 (see Figure 2). The *sibling cohort* was formed by further restricting the individual cohort to families that contained more than one

individual who identified themselves as son or daughter to the occupier (n = 51,330). Families that contain siblings with difference in age greater than 20 years were removed (n = 51,267). The cut-off value was chosen as it is reasonable to assume that siblings with such a large age difference may not have a shared childhood or family background, which defeats the point of using siblings in the analysis. 88.5% of our sibling cohort is made up of individuals that are aged 20 years and below (see Figure 3).

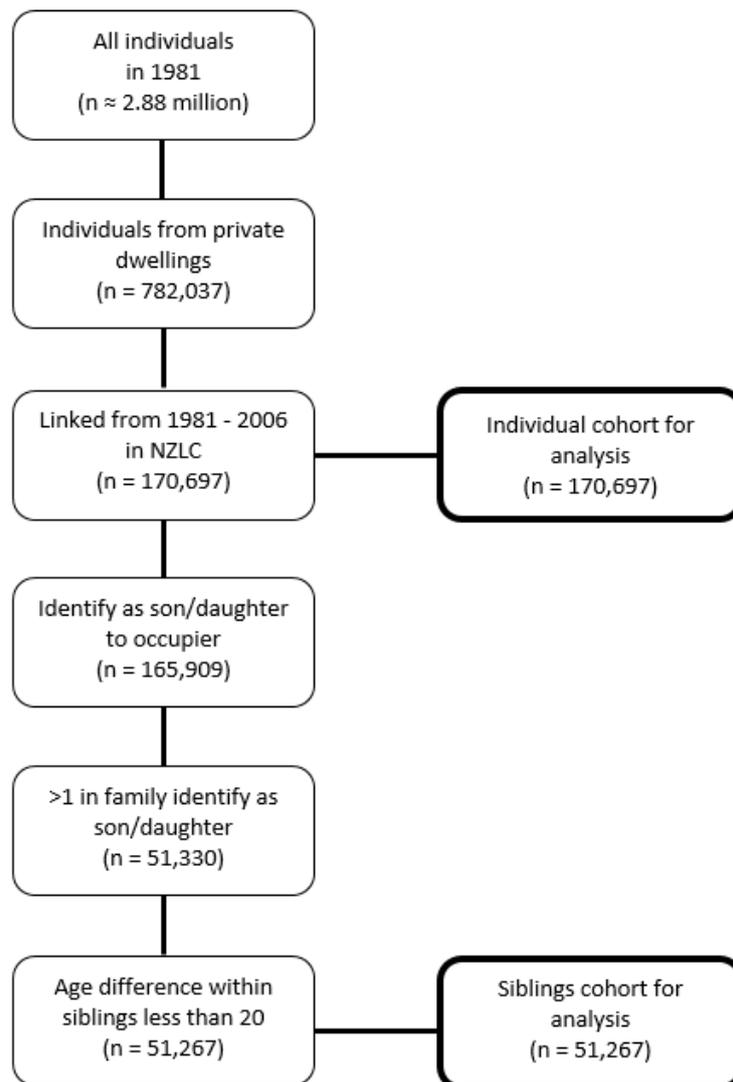


Figure 1. The identification of individual and sibling cohort from 1981 Census data

Note: Counts in this paper have been randomly rounded to base 3 in accordance with SNZ protocol.

Age distribution of individual cohort

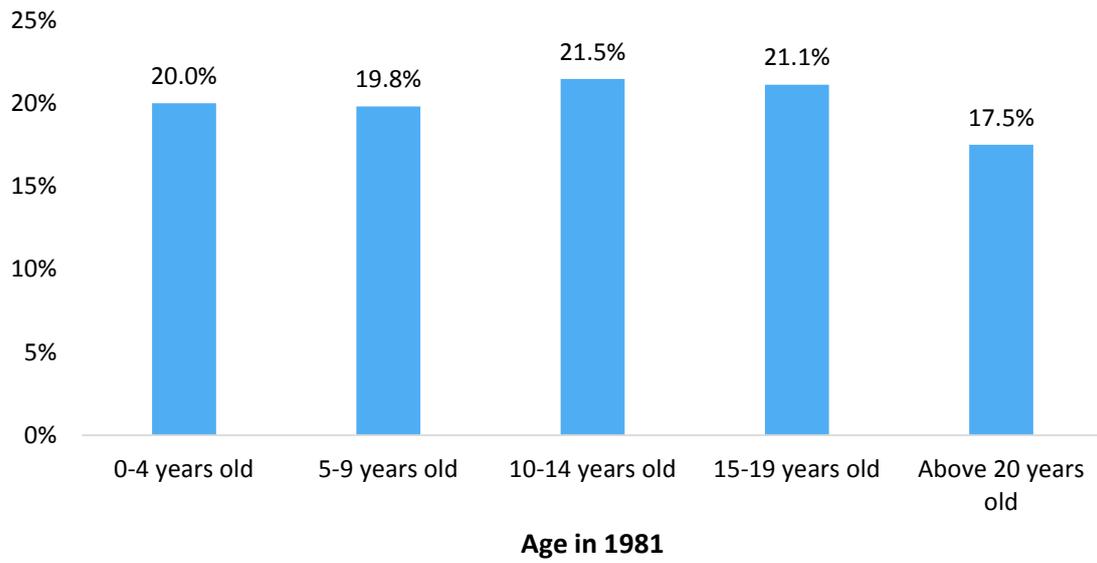


Figure 2. Age distribution of individual cohort

Age distribution for sibling cohort

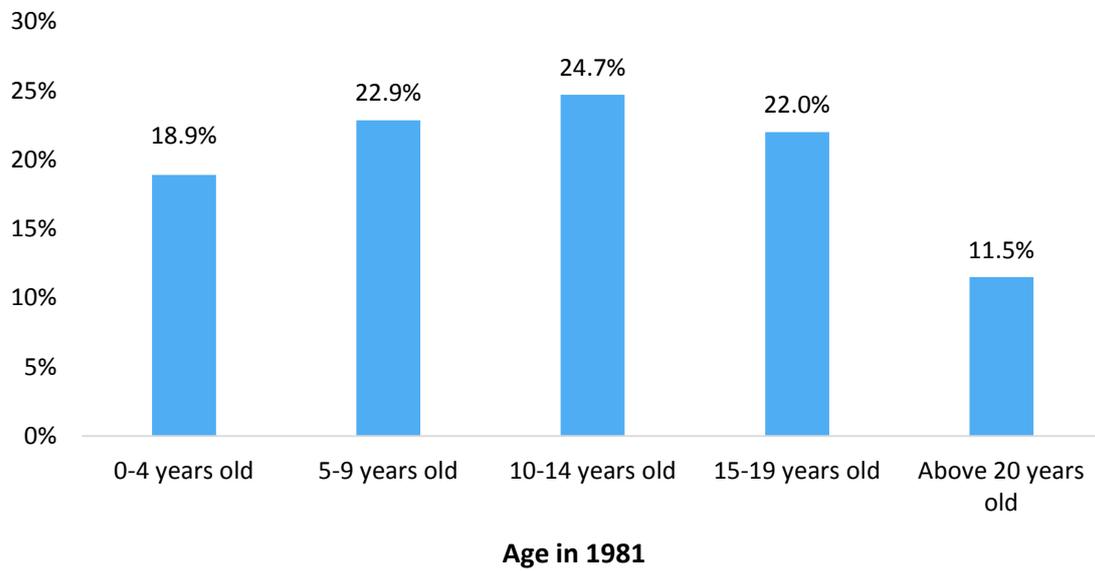


Figure 3. Age distribution of sibling cohort

2.2.2 Variables

2.2.2.1 Socio-Economic Variables

'Life-course variables' that were constructed from New Zealand Longitudinal Census (NZLC) 1981–2006 were used. The variables were harmonised (made consistent) to allow assessment of change over time; as census questions "have been developed to meet the needs of its time; therefore, each census has a different set of questions, though with a core set of questions common to all censuses." (Didham, Nissed, & Dobson, 2014). Four main socio-economic measures were identified: number of times a respondent was in the lowest *NZ Deprivation Index (NZDep)* quintile, number of times a respondent was in lowest *NZ Socio-Economic Index (NZSEI)* quintile, number of times a respondent was on *welfare* and number of times a respondent was in the lowest *household income* quintile. Household income was selected as the main measure of socio-economic position in our analysis as it was found to have the strongest effect on mortality outcomes (results available upon request). The NZDep was also subsequently investigated as a predictor in our study. These two predictor variables are in quintiles — where population is divided into five equal groups, from lowest to highest. The bottom quintile is the lowest 20 percent of the population while top quintile is the highest 20 percent. We then calculate the number of times an individual is in the lowest quintile in the six censuses between the years 1981-2006.

As our sample was selected based on individuals who were living under the same household in 1981, they consist of mainly children or young adults (about 80% of both the cohorts have age less than twenty in 1981, see Figure 2 and Figure 3). NZSEI and welfare status are only given to adults who are in the workforce and a large proportion of our sample will not have a value for these two variables. For this reason, NZSEI and welfare status are not suitable to be investigated for our sample and only household-level variables — *household income quintile* and *NZDep* as socio-economic predictors will be considered in this study. The frequencies of the number of times in lowest household income quintile between the years 1981-2006 for the two cohorts are shown in Figure 4 and Figure 5.

Table 1. SEP predictors, definition, variable names and its categories.

SEP predictors	Definition	Variable names	Categories
Household income quintile	Number of times a respondent was in lowest household income quintile between 1981 - 2006	low_inc	0 - Never in lowest household income quintile 1 - Once 2 - Twice 3 - Three times or more
NZDep index	Number of times a respondent was in lowest NZDep quintile between 1991 - 2006	low_nzdep	0 - Never in lowest NZDep quintile 1 - Once 2 - Twice 3 - Three times or more

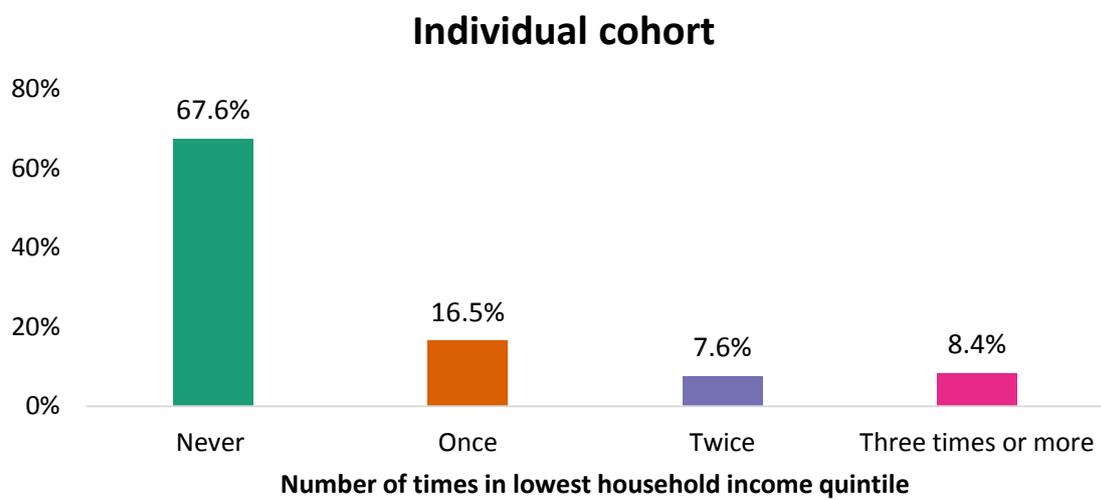


Figure 4. Number of times proportion of individual cohort in lowest household income quintile.

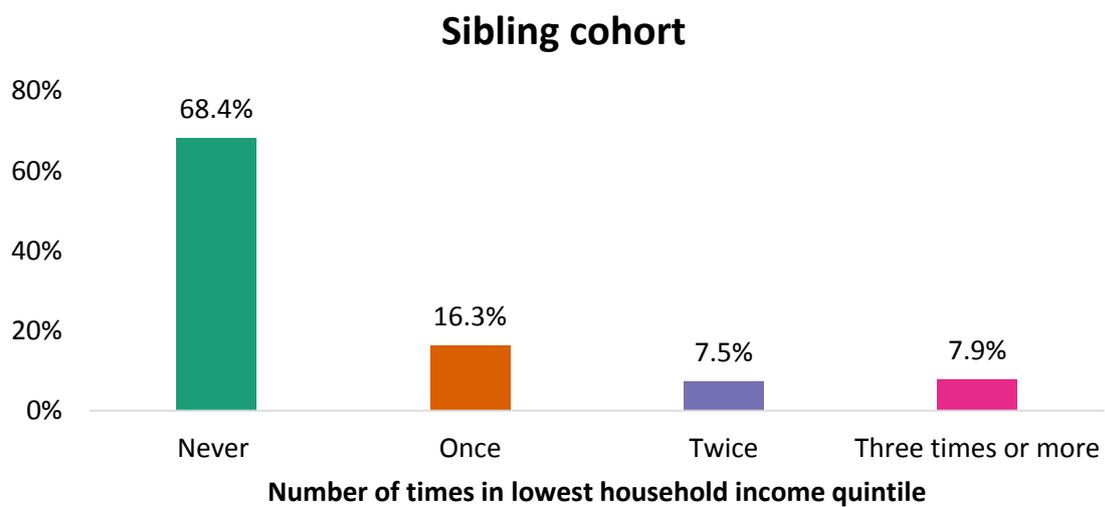


Figure 5. Number of times proportion of sibling cohort in lowest household income quintile.

2.2.2.2 Potential confounders

Additional variables were included in our analyses to mitigate any confounding effects. These variables are selected as they are expected to be both associated with one's socio-economic status and mortality outcome. All variables used are constructed life-course variables (Liu, 2015) with the exception of crowding index variables that are provided by SNZ. The factors are grouped and listed in Table 2.

Table 2. Grouped covariates, its description, variable name and categories

Covariate group	Description	Variable name	Categories
Age (years)	Age of respondent in 1981 in 5-year band	age_5yrs	1 - 0-4
			2 - 5-9
			3 - 10-14
			4 - 15-19
			5 - 20 and above
Gender	Sex of respondent	sex	1 - Male; 2 - Female
Education	Highest education attained by 2006	education_06	0 - No qualification
			1 - School qualification
			2 - Post-school qualification
			3 - Tertiary qualification
Household information	Household crowding index in years 1986, 1991, 1996, 2001 and 2006	index_86	0 - Not crowded with rooms to spare
		index_91	1 - Not crowded
		index_96	2 - Crowded and severely crowded
		index_01	999 - Missing value
		index_06	
Family and social factors	Living alone indicator in years 1986, 1991, 2001 and 2006	live_alone_86	0 - No
		live_alone_91	1 - Yes
		live_alone_96	999 - Missing value
		live_alone_01	
		live_alone_06	
	Moved residence in last year in years 1986, 1991, 2001 and 2006	moved_in_lastyear_86	0 - No
		moved_in_lastyear_91	1 - Yes
		moved_in_lastyear_96	999 - Missing value
		moved_in_lastyear_01	
		moved_in_lastyear_06	
Partnership indicator in years 1981, 1986, 1991, 2001 and 2006	partner_81	0 - Never married/widowed or bereaved/separated-divorced	
	partner_86	1 - Partnered-Married	
	partner_91	999 - Missing value	
	partner_96		
	partner_01		
Health information	Overall long term health problem indicator in years 1996, 2001 and 2006	health_96	0 - No
		health_01	1 - Yes
		health_06	999 - Missing value
	Smoking status in 2006	smoke_06	0 - No 1 - Yes 999 - Missing value
Birth order*	Birth order of sibling in the family	bthorder	Values 1-6

**only used in model for sibling cohort*

2.2.3 Outcome

2.2.3.1 Mortality records from New Zealand Census-Mortality Study (NZCMS)

Data on mortality outcomes for our cohort studies was obtained from the New Zealand Census-Mortality Study (NZCMS). The mortality records covering the five-year period from 2006–2011 were utilised on our two cohorts that were linked between years 1981 and 2006. For the 51,267 siblings identified, there were 258 observed mortality outcomes (about 0.5% of the cohort); while for the 170,697 individual cohort, there were 1,479 observed mortality outcomes (about 0.9% of the cohort). Due to the relatively low number of observed mortality outcomes, cause-specific mortality or mortality for specific ethnic groups will not be able to be investigated effectively and results are only reported on all-cause mortality.

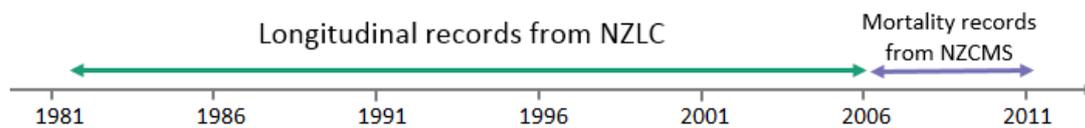


Figure 6. Timeline of 1981-2006 longitudinal records of individuals from NZLC and 2006-2011 mortality records from NZCMS

2.2.4 Bias weights

Every study comes with unavoidable biases from systematic error. Steps should be taken to account for these biases wherever possible. The linkage of census records (NZLC) and mortality with census records (NZCMS) were done anonymously and probabilistically. When we link people from one record to the other, there exist linkage biases which need to be accounted for. Linkage bias happens when the probability of a record being linked back to another varies by a factor of interest (e.g. age, ethnicity, socio-economic position) (Fawcett et al., 2002). For this reason, bias weights that were developed for NZLC and NZCMS are included in our analyses to account for the linkage biases and will be described in further details as below.

2.2.4.1 NZLC

As our sibling and individual cohorts are obtained through the NZLC records, there are biases in our sample that needs to be taken into consideration. As explained in

a report done on adjusting linkage bias in the NZLC, linkages biases in each census pairs range between 36% to 59% (Singal, 2015). Biases were significantly improved after adjusting for bias weights developed specifically for each individual to account for NZLC linkage bias. Bias weights that are based on the inverse probability of linkage given selected covariates were therefore included for each individual in our analyses.

2.2.4.2 NZCMS

The NZCMS links censuses with mortality records to create cohort of the entire NZ population. Similar to the NZLC records, there is a linkage bias in NZCMS, as people who have their mortality records linked with the census records may be different to the rest of the population in terms of their age, sex and ethnicity etc. (Fawcett et al., 2002). The 2006–2011 linked mortality records from NZCMS were utilised for our final mortality outcome in our analyses. To enable adjustments for linkage bias, the bias weights constructed specifically for NZCMS, also using the inverse probability of linkage given covariates were subsequently included for each person in our analyses.

Table 3. Definition, variable name and range values of bias weights

Variable	Definition	Variable name	Range
Bias weights	Multiplication of constructed bias weights for NZLC and NZCMS	biaswgt	Positive numbers

2.3 Data set for analyses

The identified siblings (Section 2.2.1), socio-economic measures and potential confounders (Section 2.2.2), mortality records (Section 2.2.3) and constructed bias weights (Section 2.2.4) were merged together to create longitudinal data on our two prospective cohorts that span across 1981–2006 censuses. All the relevant SAS code are found in Appendix D.

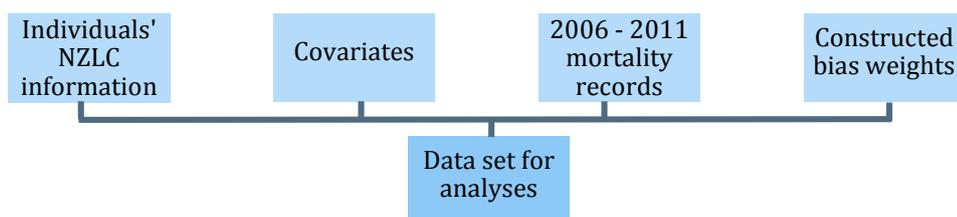


Figure 7. Merging dataset from different sources ready for analyses

Chapter 3

Methods

This chapter describes the method used for analyses in this dissertation. A series of three Cox proportional hazards regression models were fitted to our individual and sibling cohorts. This chapter begins with a description of Cox regression and the use of it in SAS software on the individual cohort to analyse the time-to-death information obtained from the NZCMS. It then follows with the description of Cox regression with modified sandwich estimator with corrected standard error used on the sibling cohort as another set of cohort analyses to take account of the familial clustering. Finally, fixed effects method is introduced into the Cox regression model to analyse the same sibling cohort that is clustered at the family level.

3.1 Cox Regression

Cox regression is a well-recognised statistical method for analysing time-to-event data, also known as survival data. These methods for modelling events are more often called survival analysis. It models the time it takes for an event to occur, where events are usually deaths, injuries, recovery from illness or onset of specific chronic disease. In our study, the time-to-event is the number of months between the five-year period between 2006 and 2011 from NZCMS and the event of interest is deaths. We are interested in estimating the mortality risk for siblings (and similarly for individuals) that have different socio-economic trajectories.

In order to investigate several variables at a time, we use the regression method presented by Cox known as proportional hazards regression analysis. David Cox introduced *proportional hazards model* and *partial likelihood* method for estimating the model (Allison, 2005). The dependent variable in the Cox proportional-hazards model is the *hazard*, or instantaneous likelihood of event

occurrence. Let T represent survival time. The hazard function for individual i at time t is given as follows:

$$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[(t \leq T < t + \Delta t) | T \geq t]}{\Delta t} \quad (1)$$

Equation (1) assesses the instantaneous risk of event at time t , conditional on survival to that time (Fox, 2002). To model the hazard as a function of the predictor variables, let $h_i(t)$ be the hazard for individual i , the proportional hazards model is given by:

$$\log h_i(t) = \mu(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

where $\mu(t)$ is the baseline hazard function, the β s are the coefficients, and the x 's are predictor variables that may vary across individuals.

The method of partial likelihood makes it possible to estimate β without specifying anything about the function μ . Cox regression analyses were used in our study due to the right-censored outcomes. If an individual does not have a linked mortality record for years 2006–2011 to their 2006 census record — they are considered to be alive or unable to be followed up and are said to be right-censored. For individuals who are censored, a value of 61 months is given as that is the length of period for mortality records from 2006–2011.

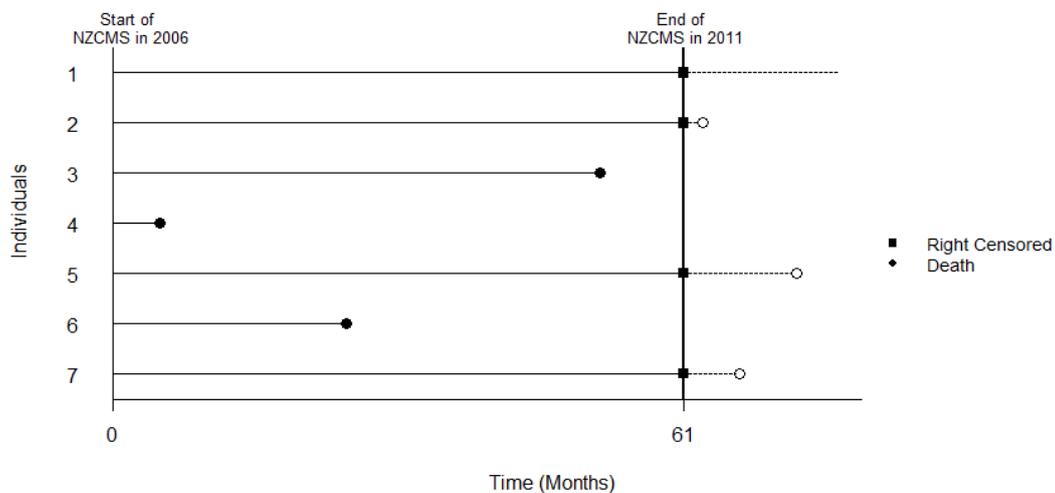


Figure 8. Example of censoring in our data – most individuals are censored at 61 months where the dotted lines are the individual’s true life span.

3.1.1 Cox regression in SAS

All of the data management and analyses are done in SAS Enterprise Guide 7.1 in SNZ Data Lab. The proportional hazards model is implemented in SAS using command `PROC PHREG`. Three series of analyses were done for each outcome on our two cohorts:

1. Cox regression
2. Cox regression with sandwich estimator corrected standard errors
3. Cox regression with fixed effects

3.1.1.1 Individual cohort

For our 170,197 individuals in the individual cohort, we fitted the Cox regression without any modification in SAS software to model their time to death.

Groups of covariates as specified in Table 2 are added to our model sequentially and hazard ratios for our main predictor `low_inc` are recorded in series. The SAS commands for fitting the model for just adjusting age and sex are done as follow:

```
PROC PHREG DATA=fe2;
  WEIGHT=biaswgt;
  CLASS low_inc sex age_5yrs /ref=FIRST;
  MODEL months*dead(0) = low_inc sex age_5yrs /TIES=EFRON;
RUN;
```

The SAS commands for fitting the model after adjusting for all the covariates are:

```
PROC PHREG DATA=fe2;
  WEIGHT=biaswgt;
  CLASS low_inc sex age_5yrs education_06 index_:
  live_alone_: moved_in_lastyear_: partner_: health_:
  smoke_06 /ref=FIRST;
  MODEL months*dead(0) = low_inc sex age_5yrs
  education_06 index_: live_alone_: moved_in_lastyear_:
  partner_: health_: smoke_06 /TIES=EFRON;
RUN;
```

All the variables used here are described in Table 2 and each specification used in the model is described in Appendix B. The commands for running all analyses and adding the covariates sequentially are included in Appendix C.

3.1.1.2 Sibling cohort

In our sibling cohort, the individuals are clustered at family level. If we assume independence among the individuals in the sibling cohort, there are repeated observations within families. Failure to address this dependence in the model could lead to serious underestimates of standard errors and p-values (Allison, 2005). We need a method commonly called *robust variance estimator* or the *modified sandwich estimator* to correct for the dependence. An option called `COVSANDWICH`(`AGGREGATE`) is available in SAS and included in `PROC PHREG` to correct for dependence when there are repeated observations.

```
PROC PHREG DATA=sibfe3 COVSANDWICH(AGGREGATE);  
  WEIGHT=biaswgt;  
  CLASS low_inc sex bthorder age_5yrs education_06 index_  
    live_alone_ moved_in_lastyear_ partner_ health_  
    smoke_06 /ref=FIRST;  
  MODEL months*dead(0) = low_inc sex bthorder age_5yrs  
    education_06 index_ live_alone_ moved_in_lastyear_  
    partner_ health_ smoke_06 /TIES=EFRON;  
  ID family_nbr_81;  
RUN;
```

The `ID` statement is used for variable that is common to all observations in the same “cluster” — in our case is the `family_nbr_81` which is the individual’s family number in 1981 in which siblings are clustered.

3.2 Cox Regression with Fixed Effects

In non-experimental studies like ours, we want to approximate a randomized experiment to statistically control for other variables that cannot be measured. By using fixed effects methods, it makes it possible to control for all possible stable characteristics of the families in the study, even without measuring them, provided that those characteristics do not change over time (Allison, 2005). The fixed effects method also addresses the dependence problem when there are repeated observations, while controlling for all stable predictor variables.

It is important to note that the fixed effects methods includes no term for the between-family variation and focuses only the within-family variation. Fitting Cox regression with fixed effects has the same model as an ordinary Cox regression but

with an extra term, α_i in it that represents the combined effects of all stable covariates.

$$\log h_i(t) = \mu(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \alpha_i$$

After modifying equation above by defining

$$\mu_i(t) = \mu(t) + \alpha_i$$

which yields

$$\log h_i(t) = \mu_i(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad (2)$$

In equation (2), the fixed effects α_i has been absorbed into the unspecified function of time, which is now allowed to vary from one family to another (Allison, 2005). Thus, each family has her own hazard function, which is considerably less restrictive than allowing each family to have their own constant.

3.2.1 Cox Regression with Fixed Effects in SAS

Equation (2) can be estimated by partial likelihood using the well-known method of stratification. In `PROC PHREG`, stratification is implemented with the `STRATA` statement.

Again, groups of covariates are added sequentially and are included in Appendix E. Our analysis adjusting for all the covariates can be done in SAS software as the following:

```
PROC PHREG DATA=sibfe3 NOSUMMARY;  
  WEIGHT=biaswgt;  
  CLASS low_inc bthorder sex age_5yrs education_06 index_  
        live_alone_ moved_in_lastyear_ partner_ health_  
        smoke_06 /ref=FIRST;  
  MODEL months*dead(0) = low_inc bthorder sex age_5yrs  
        education_06 index_ live_alone_ moved_in_lastyear_  
        partner_ health_ smoke_06 /TIAS=EFRON;  
  STRATA family_nbr_81;  
RUN;
```

The `NOSUMMARY` option is specified to exclude uninformative outputs from SAS software. The SAS statement `STRATA family_nbr_81` creates a separate baseline hazard functions fit for each stratum — `family_nbr_81`, a total of 23,934 families for the sibling cohort.

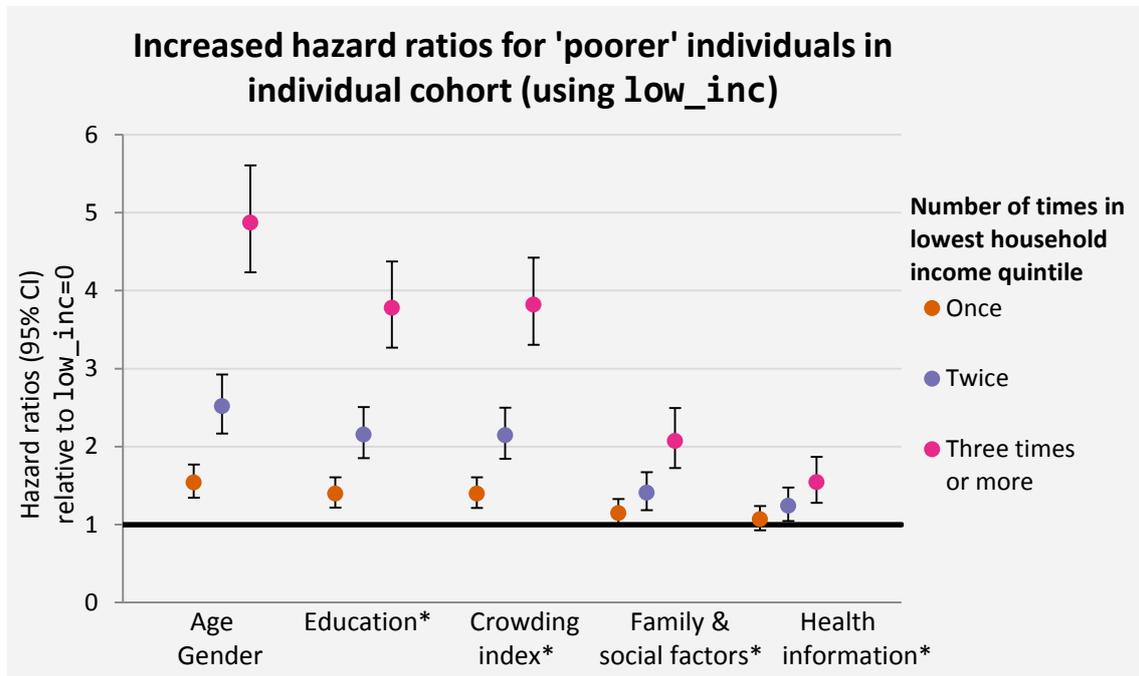
Chapter 4

Results

This chapter describes the results from the three series of Cox regression analyses of mortality using socio-economic position — household income quintile and NZDep as main predictors. The estimated hazard ratios for the three series of models with the corresponding 95% confidence intervals are summarised in Section 4.4. The chapter contains three sections with plotted hazard ratios done in Microsoft Excel while the plots of survival curves are done in SAS software. Results are plotted and tabulated using colours that correspond to the categories accordingly. All the code used to run the analyses can be found in Appendix E.

4.1 Individual cohort

Household income quintile



**groups of covariates are added sequentially*

Figure 9. Hazard ratios of an individual (in individual cohort) in lowest household income quintile relative to individual who was never in lowest household income quintile after adjusting for groups of covariates sequentially

Table 4. Hazard ratios of individuals in individual cohort relative to individuals who were never in lowest household income quintile

Number of times in lowest household income quintile	Adjusting for groups of covariates sequentially				
	Age & Gender	Education	Crowding index	Family & social factors	Health information
Once	1.54 (1.34 – 1.77)*	1.40 (1.22 – 1.61)*	1.40 (1.22 – 1.61)*	1.15 (0.99 – 1.33)	1.07 (0.93 – 1.24)
Twice	2.52 (2.17 – 2.93)*	2.16 (1.85 – 2.51)*	2.15 (1.84 – 2.50)*	1.41 (1.19 – 1.67)*	1.24 (1.04 – 1.48)*
Three times or more	4.87 (4.24 – 5.61)*	3.78 (3.27 – 4.37)*	3.82 (3.30 – 4.42)*	2.07 (1.72 – 2.50)*	2.52 (2.17 – 2.93)*

*statistically significant results at the 5% level

Figure 9 shows that the mortality risks for the ‘poorer’ individual increases as the number of times in lowest household income quintile increases. It is to note that the groups of covariates are added successively (e.g. first model adjusts for age and gender, second model adjusts for age, gender and education and so on). All estimates are tabulated in Table 4 and shows they are *all* significant at the 5% level (hazard ratio greater than one) except for individuals that were in lowest household income quintile once after controlling for family and social factors (Hazard Ratio (HR): 1.15; 95% Confidence Interval (CI): 0.99 – 1.33) and health information subsequently (HR: 1.07; 95% CI: 0.93 – 1.24).

On average, for an individual that was in lowest household income quintile *once*:

- There is an increased hazard ratio of 1.54 (95% CI: 1.34 – 1.77) compared to sibling who was *never* in lowest household income quintile after adjusting for just age and gender. The hazard ratio decreases to 1.40 (95% CI: 1.22 – 1.61) after adjusting for highest education attained in 2006; and stays the same at 1.40 (95% CI: 1.22 – 1.61) controlling for crowded households.
- The estimates are not significant after taking family and social factors into account. These estimates contains hazard ratio of one in the confidence intervals indicating that the association of mortality risks for siblings being in lowest household income quintile once compared with siblings who were never in lowest household income quintile is not significant at the 5% level.

For an individual that was in lowest household income quintile *twice* and *three times or more*:

- All estimates are significant at the 5% level.
- There are increased mortality hazard of 1.24 (95% CI: 1.04 – 1.48) and 2.52 times more (95% CI: 2.17 – 2.93) respectively when compared to an individual that was never in lowest household income quintile, after adjusting for all potential confounders.

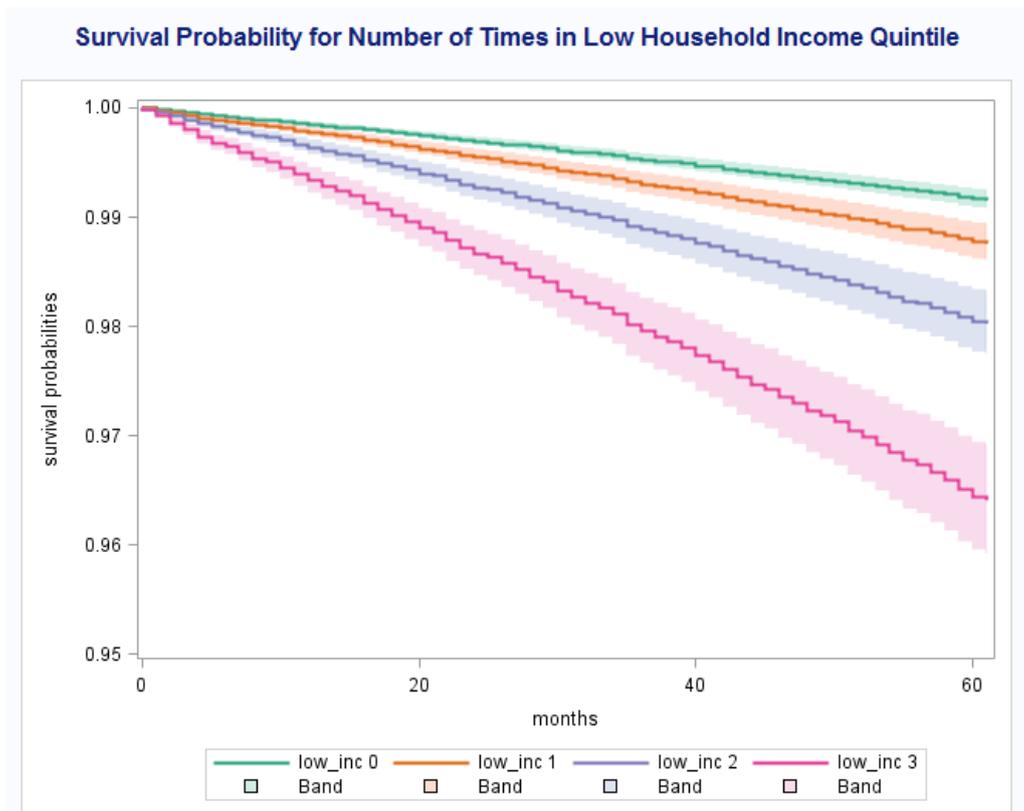
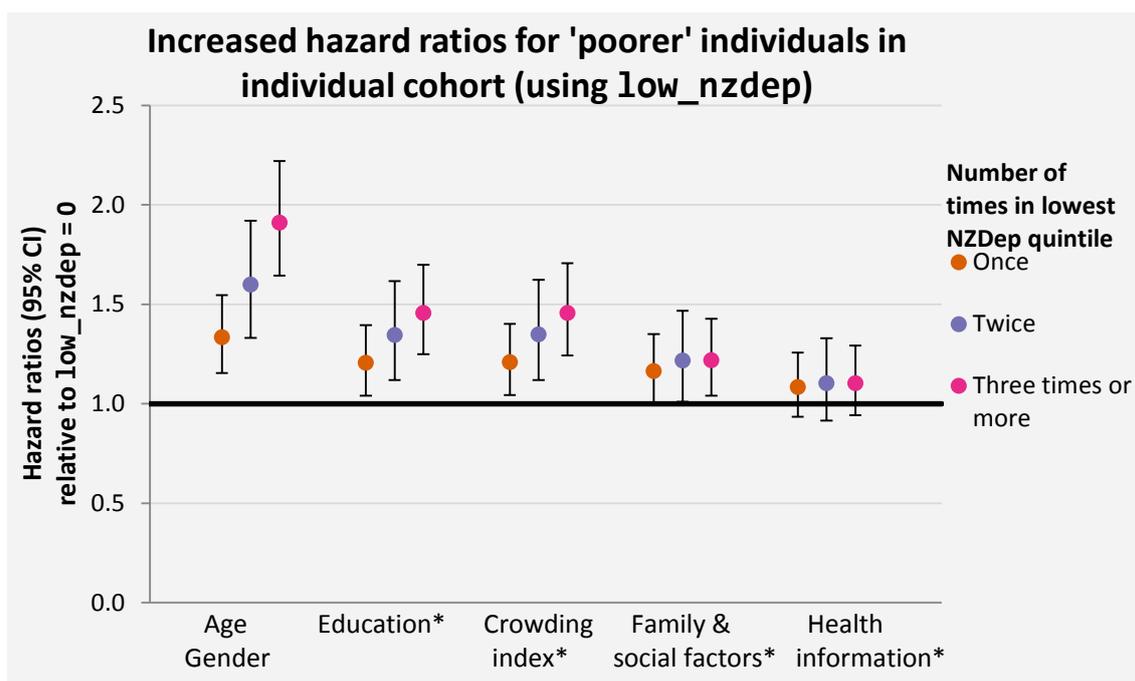


Figure 10. Survival curves of an individual from the individual cohort for all categories of low_inc with mean values for all other covariates in the model

Figure 10 shows the survival probabilities for an individual from the individual cohort that was never in lowest household income quintile is 0.995, or 99.5%. For individuals that were in lowest household income quintile once, twice and three times or more, the survival probabilities are 99%, 98.2% and 96.6% subsequently. The bands (confidence intervals) show no overlap indicating there are significant differences across all the groups.

NZDep



*groups of covariates are added sequentially

Figure 11. Hazard ratios of individuals in individual cohort in lowest NZDep adjusting for groups of covariates sequentially

Figure 11 (and Table 5) shows that the estimates for all the individuals in the individual cohort are not significant adjusting for health indicator after controlling for age and gender, highest education attained by 2006, and crowding index sequentially. The survival curve for individuals using NZDep as socio-economic measure is similar to Figure 10 and is not presented.

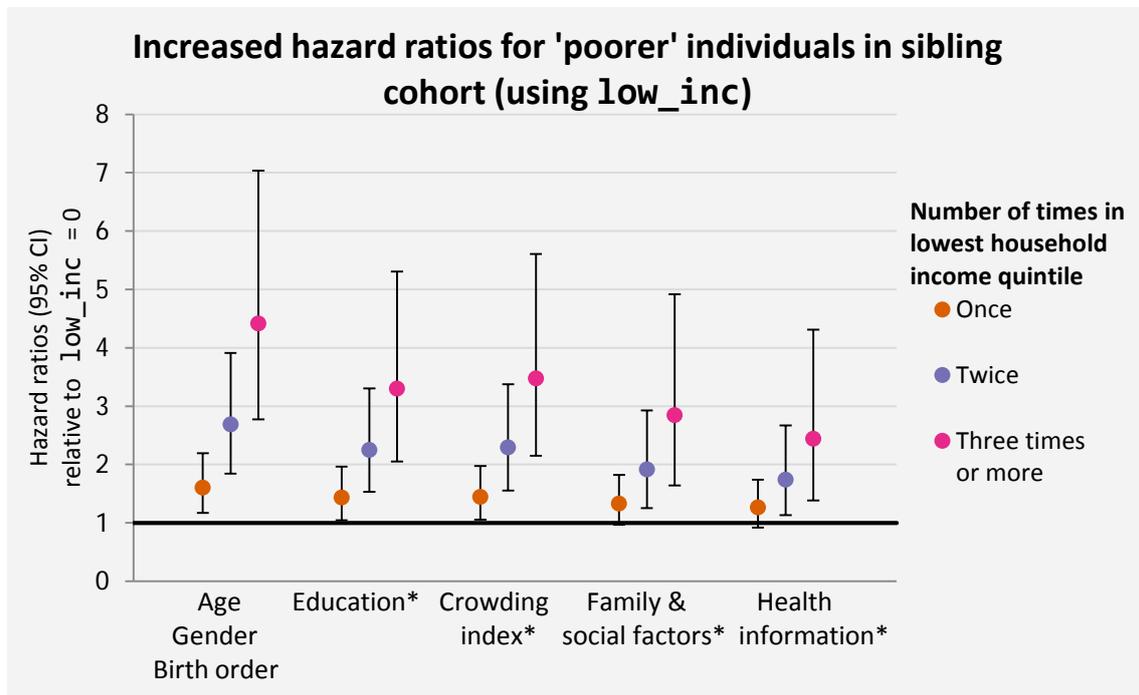
Table 5. Hazard ratios of individuals in individual cohort relative to individuals who were never in lowest NZDep

Number of times in lowest NZDep	Adjusting for groups of covariates sequentially				
	Age & Gender	Education	Crowding index	Family & social factors	Health information
Once	1.34 (1.15 - 1.55)*	1.21 (1.04 - 1.40)*	1.21 (1.04 - 1.40)*	1.17 (1.01 - 1.35)*	1.08 (0.94 - 1.26)
Twice	1.60 (1.33 - 1.92)*	1.35 (1.12 - 1.62)*	1.35 (1.12 - 1.62)*	1.22 (1.01 - 1.47)*	1.10 (0.92 - 1.33)
Three times or more	1.91 (1.65 - 2.22)*	1.46 (1.25 - 1.70)*	1.46 (1.24 - 1.71)*	1.22 (1.04 - 1.43)*	1.10 (0.94 - 1.29)

*statistically significant results at the 5% level

4.2 Sibling cohort without fixed effects

Household income quintile



*groups of covariates are added sequentially

Figure 12. Hazard ratios of siblings with robust variance estimation in lowest household income quintile relative to sibling who was never in the lowest household income quintile adjusting for groups of covariates sequentially

For the individuals in the sibling cohort, without taking the familial clustering into account, it is also apparent that the mortality risks for the 'poorer' individual increases as the number of times in lowest household income quintile increases. The results similar to the individual analyses (Figure 9) but with increased standard error and have larger confidence intervals. There are less certainties with these estimates and that is most likely due to smaller sample size of the cohort (n = 51,267 as opposed to 170,697).

Table 6. Hazard ratios of individuals in sibling cohort relative to individual who were never in lowest household income quintile

Number of times in lowest household income quintile	Adjusting for groups of covariates sequentially				
	Age, gender and birth order	Education	Crowding index	Family & social factors	Health information
Once	1.60 (1.17 - 2.19)*	1.43 (1.05 - 1.96)*	1.44 (1.06 - 1.98)*	1.33 (0.97 - 1.82)	1.26 (0.92 - 1.74)
Twice	2.69 (1.84 - 3.91)*	2.25 (1.53 - 3.31)*	2.29 (1.56 - 3.38)*	1.91 (1.25 - 2.93)*	1.74 (1.13 - 2.67)*
Three times or more	4.42 (2.77 - 7.04)*	3.30 (2.05 - 5.31)*	3.47 (2.15 - 5.61)*	2.84 (1.64 - 4.92)*	2.44 (1.38 - 4.31)*

*statistically significant results at the 5% level

Table 6 shows that all estimates are statistically significant at the 5% level (hazard ratio greater than one) except for individuals that were in lowest household income quintile once after controlling for family and social factors (HR: 1.33; 95% CI: 0.97 - 1.82) and health information subsequently (HR: 1.26; 95% CI: 0.92 - 1.74).

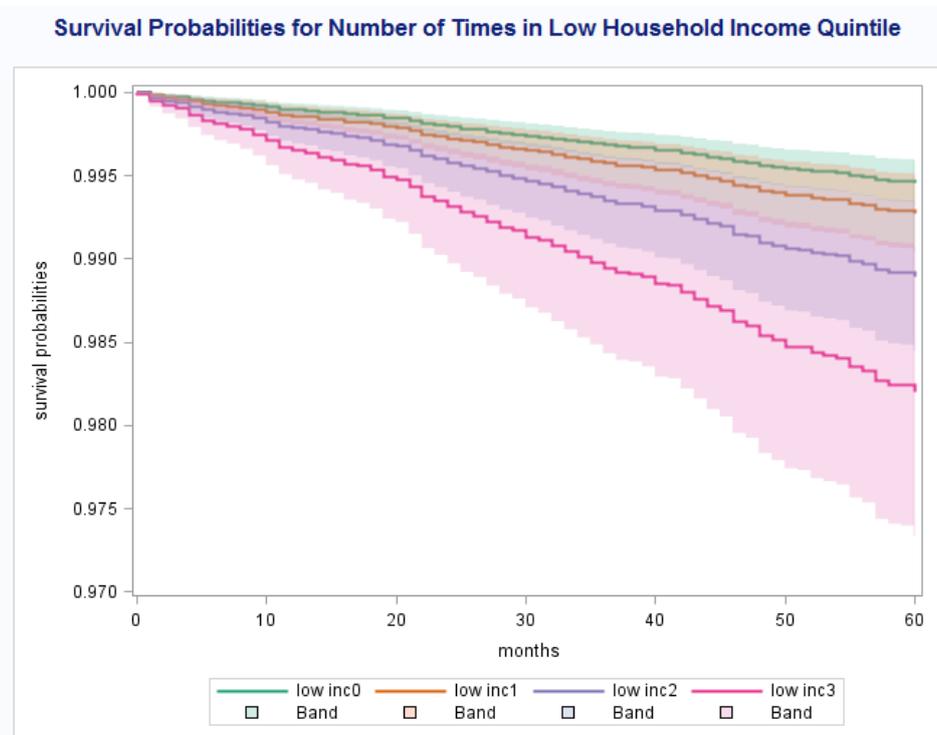
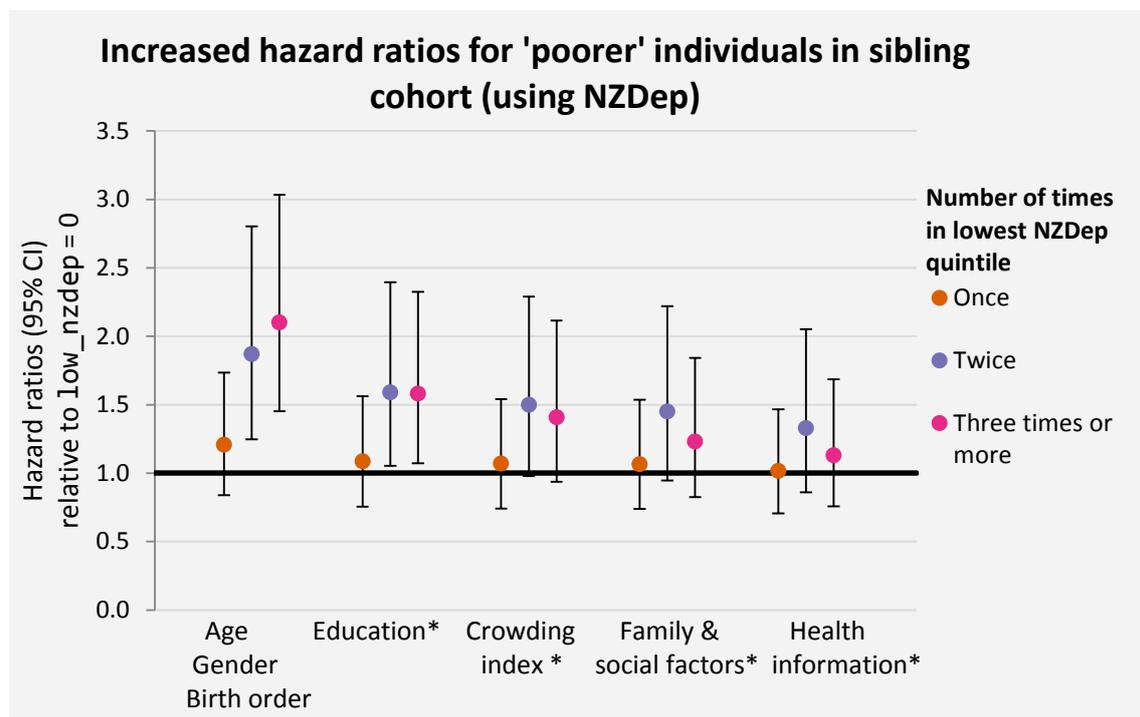


Figure 13. Survival curves of siblings for all categories of low_inc and mean values for all other covariates in the model

Figure 13 shows that for an individual in the sibling cohort that was in lowest household income quintile once, twice and three times or more between years 1981–2006, the survival probabilities are approximately 0.994, 0.990 and 0.983 subsequently. That corresponds to about survival rates of 98-99% for all groups of individuals. However there is a huge overlap in the confidence intervals (bands around survival curves) indicating the uncertainties in these estimates.

NZDep



**groups of covariates are added sequentially*

Figure 14. Hazard ratios of individuals in the sibling cohort with robust variance estimation in lowest NZDep relative to individuals who were never in lowest household NZDep quintile adjusting for groups of covariates

The hazard ratios of individuals in the sibling cohort with robust variance estimation in lowest NZDep show that only the group of individuals in lowest NZDep twice and three times or more after adjusting for age and gender and education sequentially are significant (HR greater than one).

Table 7. Hazard ratios of individuals in individual cohort relative to individual who were never in lowest NZDep

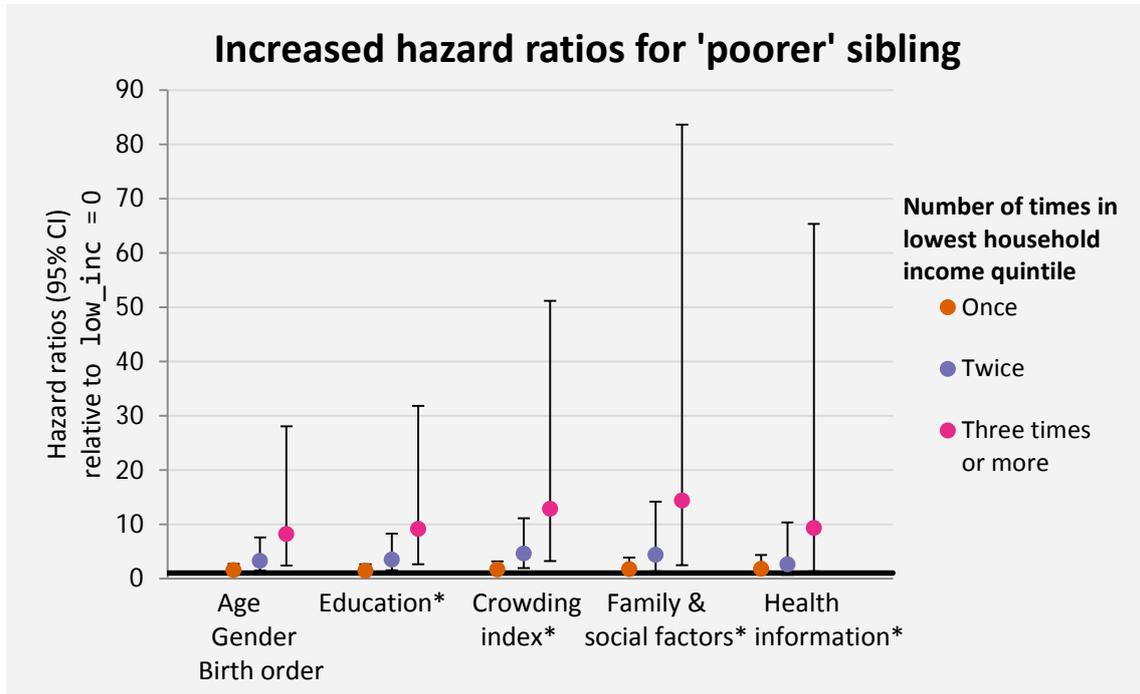
Number of times in lowest NZDep	Adjusting for groups of covariates sequentially				
	Age & Gender	Education	Crowding index	Family & social factors	Health information
Once	1.21 (0.84 - 1.74)	1.09 (0.76 - 1.56)	1.07 (0.74 - 1.54)	1.07 (0.74 - 1.54)	1.02 (0.71 - 1.47)
Twice	1.87 (1.25 - 2.80)*	1.59 (1.05 - 2.40)*	1.50 (0.98 - 2.29)	1.45 (0.95 - 2.22)	1.33 (0.86 - 2.05)
Three times or more	2.10 (1.45 - 3.03)*	1.58 (1.07 - 2.33)*	1.41 (0.94 - 2.12)	1.23 (0.82 - 1.84)	1.13 (0.76 - 1.69)

**statistically significant results at the 5% level*

The survival curves for individuals using NZDep as the socio-economic measure is similar to Figure 13 and hence are not presented.

4.3 Sibling cohort with fixed effects

Household income quintile



**groups of covariates are added sequentially*

Figure 15. Hazard ratios of siblings in lowest household income quintile relative to siblings who were never in lowest household income quintile adjusting for groups of covariates

As the hazard ratios from this model have large confidence intervals, the estimates are plotted on the log-scale for easier visualisation. However, interpretations are done in terms of hazard ratios.

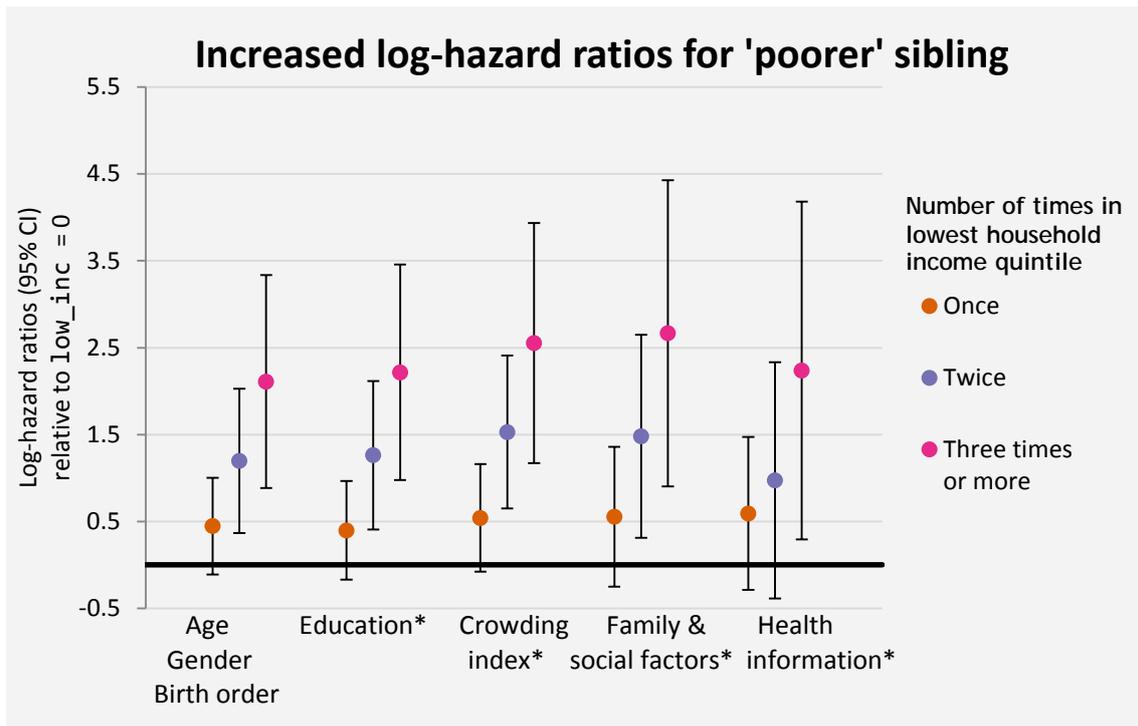


Figure 16. Log-hazard ratios of siblings in lowest household income quintile relative to siblings who were never in lowest household income quintile adjusting for groups of covariates

Figure 15 shows the hazard ratios for siblings who were in lowest household income quintile once, twice and three times or more as compared to sibling who were *never* in lowest household income quintile between the years 1981–2006. It is also apparent that the mortality risks for the ‘poorer’ sibling increases as the number of times in lowest household income quintile increases. However, despite the large sample size ($n = 51,267$), the estimates using the fixed effects method are estimated with large error when a particular covariate differs greatly across families but shows little variation within families (Allison, 2005). This is apparent for the ‘Three times or more’ group with large quoted confidence intervals (Figure 15).

Table 8. Hazard ratios of a sibling relative to sibling who was never in lowest household income quintile

Number of times in lowest household income quintile	Adjusting for groups of covariates sequentially				
	Age, gender & birth order	Education	Crowding index	Family & social factors	Health information
Once	1.56 (0.90 – 2.73)	1.49 (0.84 – 2.62)	1.72 (0.93 – 3.20)	1.74 (0.78 – 3.90)	1.81 (0.75 – 4.36)
Twice	3.31 (1.45 – 7.59)*	3.54 (1.51 – 8.31)*	4.62 (1.92 – 11.13)*	4.39 (1.36 – 14.15)*	2.65 (0.68 – 10.33)
Three times or more	8.25 (2.42 – 28.07)*	9.18 (2.65 – 31.78)*	12.85 (3.23 – 51.15)*	14.38 (2.47 – 83.65)*	9.36 (1.34 – 65.36)*

*statistically significant results at the 5% level

On average, for a sibling that was in lowest household income quintile *once* during years 1981–2006:

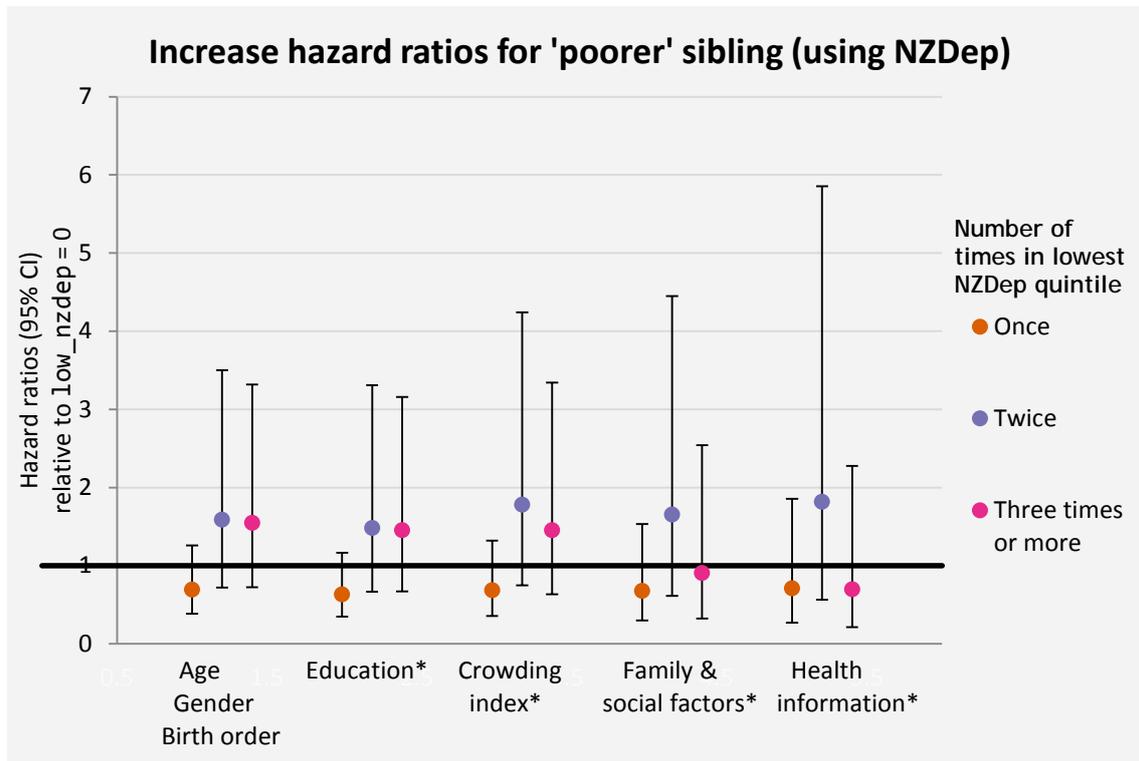
- There is an increased hazard ratio of 1.56 (95% Confidence Interval (CI): 0.90 – 2.73) compared to sibling who was *never* in lowest household income quintile after adjusting for just age, gender and birth order. The hazard ratio decreases to 1.49 (95% CI: 0.84 – 2.62) after incremental adjustment for highest education attained in 2006; increases to 1.72 (95% CI: 0.93 – 3.20) controlling for crowded households; and increases slightly to 1.74 (95% CI: 0.78 – 3.90) after taking family and social factors into account. Adjusting for health information, the hazard ratio is 1.81 (95% CI: 0.75 – 4.36) compared to sibling that was never in lowest household income quintile.
- These estimates all contain hazard ratio of one in the confidence intervals indicating that the association of mortality risks for siblings being in lowest household income quintile once compared with siblings who were never in lowest household income quintile is not statistically significant at the 5% level.

For a sibling that was in lowest household income quintile *twice* during the years 1981–2006:

- There is an increased hazard ratio of 3.31 (95% CI: 1.45 – 7.59) compared to a sibling who was *never* in lowest household income quintile after adjusting for age, gender and birth order.
- The hazard ratios are all significant and increases after education status, crowding households and family and social factors have been incrementally taken into account (Hazard Ratio (HR): 3.54, 95% CI: 1.51 – 8.31; HR: 4.62, 95% CI: 1.92 – 11.13; HR: 4.39, 95% CI: 1.36 – 14.15).
- It is interesting to note that the hazard ratio becomes *insignificant* after adjusting for health information.

Significant results are found for a sibling that was in lowest household income quintile *three times or more* during the years 1981–2006 but are quoted with large confidence intervals indicating the uncertainties in these estimates.

Survival curve is not shown for this series of model as hazard functions using fixed effects are estimated for each family (strata) and is not sensible for interpretation.



**groups of covariates are added sequentially*

Figure 17. Hazard ratios of a sibling relative to sibling who was never in lowest NZDep adjusting for groups of covariates sequentially

Figure 17 shows the hazard ratios for siblings who were in lowest NZDep once, twice or three times or more as compared to siblings who were never in lowest NZDep between the years 1981–2006. The plot shows all the 95% confidence intervals for all the estimates contain hazard ratio of one suggesting that there are no significant difference between siblings who were in lowest NZDep *once, twice or three times or more* compared to siblings who were never in lowest NZDep.

4.4 Summaries

The hazard ratios quoted and shown in Section 4.1, Section 4.2 and Section 4.3 are summarised in the following tables (Table 9 and Table 10). Each row corresponds to the hazard ratios after adjusting for groups of covariates sequentially.

Table 9. Hazard ratios of low_inc when compared to sibling/individual who was never in lowest household income quintile from three series of models on two cohorts

Number of times in lowest household income quintile	Individual cohort (95% CI)	Sibling cohort without fixed effects (95% CI)	Sibling cohort with fixed effects (95% CI)
Once	1.54 (1.34 - 1.77)*	1.60 (1.17 - 2.19)*	1.56 (0.90 - 2.73)
	1.40 (1.22 - 1.61)*	1.43 (1.05 - 1.96)*	1.49 (0.84 - 2.62)
	1.40 (1.22 - 1.61)*	1.44 (1.06 - 1.98)*	1.72 (0.93 - 3.20)
	1.15 (0.99 - 1.33)	1.33 (0.97 - 1.82)	1.74 (0.78 - 3.90)
	1.07 (0.93 - 1.24)	1.26 (0.92 - 1.74)	1.81 (0.75 - 4.36)
Twice	2.52 (2.17 - 2.93)*	2.69 (1.84 - 3.91)*	3.31 (1.45 - 7.59)*
	2.16 (1.85 - 2.51)*	2.25 (1.53 - 3.31)*	3.54 (1.51 - 8.31)*
	2.15 (1.84 - 2.50)*	2.29 (1.56 - 3.38)*	4.62 (1.92 - 11.13)*
	1.41 (1.19 - 1.67)*	1.91 (1.25 - 2.93)*	4.39 (1.36 - 14.15)*
	1.24 (1.04 - 1.48)*	1.74 (1.13 - 2.67)*	2.65 (0.68 - 10.33)
3 times or more	4.87 (4.24 - 5.61)*	4.42 (2.77 - 7.04)*	8.25 (2.42 - 28.07)*
	3.78 (3.27 - 4.37)*	3.30 (2.05 - 5.31)*	9.18 (2.65 - 31.78)*
	3.82 (3.30 - 4.42)*	3.47 (2.15 - 5.61)*	12.85 (3.23 - 51.15)*
	2.07 (1.72 - 2.50)*	2.84 (1.64 - 4.92)*	14.38 (2.47 - 83.65)*
	1.55 (1.28 - 1.87)*	2.44 (1.38 - 4.31)*	9.36 (1.34 - 65.36)*

*statistically significant results at the 5% level

Table 10. Hazard ratios of low_nzdep when compared to sibling/individual who was never in lowest NZDep from three series of models on two cohorts

Number of times in lowest NZDep quintile	Individual cohort (95% CI)	Sibling cohort without fixed effects (95% CI)	Sibling cohort with fixed effects (95% CI)
Once	1.34 (1.15 - 1.55)*	1.21 (0.84 - 1.74)	0.70 (0.39 - 1.26)
	1.21 (1.04 - 1.40)*	1.09 (0.76 - 1.56)	0.64 (0.35 - 1.17)
	1.21 (1.04 - 1.40)*	1.07 (0.74 - 1.54)	0.69 (0.36 - 1.32)
	1.17 (1.01 - 1.35)*	1.07 (0.74 - 1.54)	0.68 (0.30 - 1.53)
	1.08 (0.94 - 1.26)	1.02 (0.71 - 1.47)	0.71 (0.27 - 1.86)
Twice	1.60 (1.33 - 1.92)*	1.87 (1.25 - 2.80)*	1.59 (0.72 - 3.50)
	1.35 (1.12 - 1.62)*	1.59 (1.05 - 2.40)*	1.49 (0.67 - 3.31)
	1.35 (1.12 - 1.62)*	1.50 (0.98 - 2.29)	1.78 (0.75 - 4.24)
	1.22 (1.01 - 1.47)*	1.45 (0.95 - 2.22)	1.66 (0.62 - 4.45)
	1.10 (0.92 - 1.33)	1.33 (0.86 - 2.05)	1.82 (0.56 - 5.86)
3 times or more	1.91 (1.65 - 2.22)*	2.10 (1.45 - 3.03)*	1.55 (0.72 - 3.32)
	1.46 (1.25 - 1.70)*	1.58 (1.07 - 2.33)*	1.45 (0.67 - 3.16)
	1.46 (1.24 - 1.71)*	1.41 (0.94 - 2.12)	1.46 (0.64 - 3.34)
	1.22 (1.04 - 1.43)*	1.23 (0.82 - 1.84)	0.91 (0.33 - 2.54)
	1.10 (0.94 - 1.29)	1.13 (0.76 - 1.69)	0.70 (0.21 - 2.28)

*statistically significant results at the 5% level

4.5 Diagnostics

The goodness-of-fit and proportional hazard assumptions for the Cox regression models are examined by plotting the residuals.

Proportional hazards assumptions for our Cox regression model were examined by plotting the scaled Schoenfeld residual for our main SEP predictor (`low_inc`) against months, and the fitted loess smoothing at about zero shows we have no evidence against the assumption (Figure 18). There are four-level of observations in the plot due to the 4-level factor variable of our predictor. The Schoenfeld residuals plot is done in SAS Enterprise Guide 9.1.

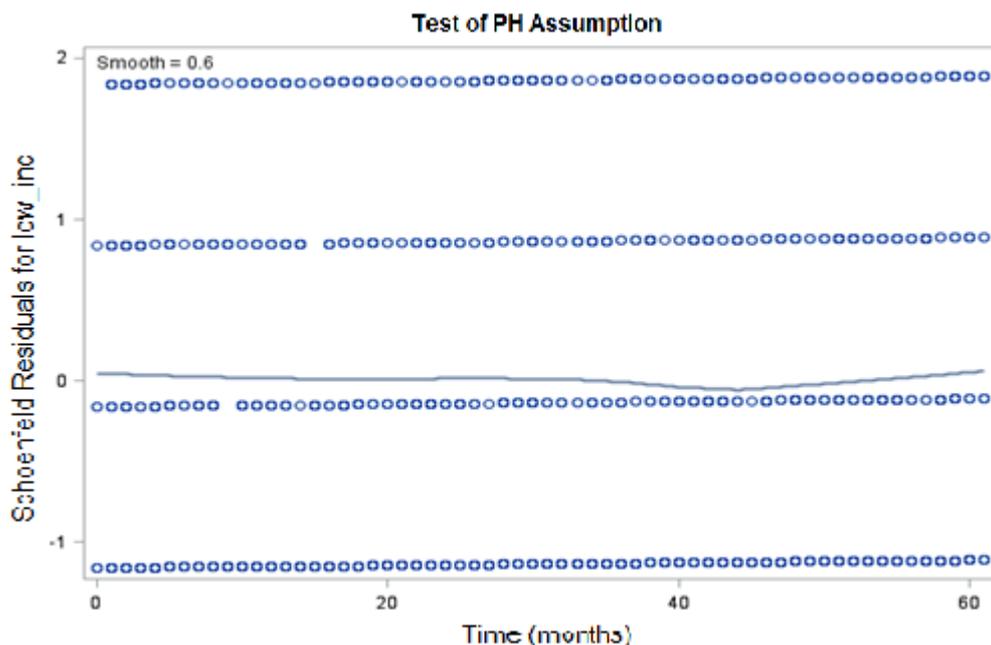


Figure 18. Schoenfeld residuals plot for `low_inc` against months with fitted loess

Tests for proportional hazards on all the other variables are included in Appendix C and shows we have no evidence against the proportional hazards assumptions.

Schoenfeld residuals plot for the other two models show similar pattern with fitted loess around zero and are not included.

The goodness-of-fit of the Cox regression are investigated by plotting the deviance residuals against the covariates. The deviance residuals plot (Figure 19) does not show a random scatter about zero as we would expect from a standard residual

plot due to the censored outcomes in our model. However, a loess smoothing was fitted at about zero indicating we have no evidence against the goodness-of-fit of our model. The residuals plots are done in R version 3.2.1 (due to SAS software being able to handle only 5000 observations for residuals plots).

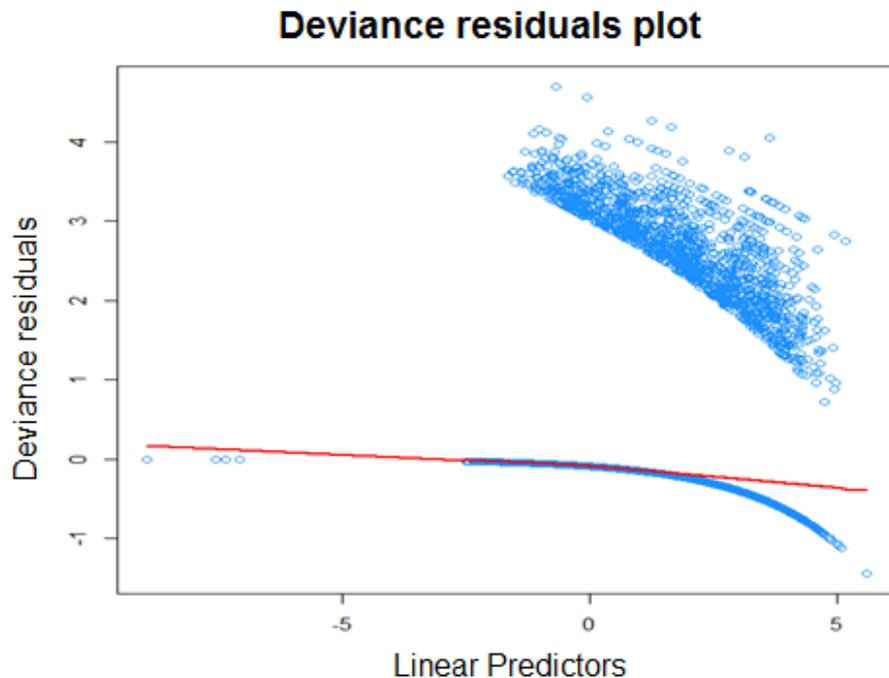


Figure 19. Deviance residuals plot against linear predictors included in the model with fitted loess.

The unusual shape in Figure 19 is caused by our censored outcomes with the top cluster corresponding to individuals who are dead and bottom cluster corresponding to individuals who are alive in the records between years 2006–2011.

The deviance residuals plots for the other two models show similar pattern as above with fitted loess around zero and hence are not included.

Chapter 5

Discussion

The results generally show that there are increasing mortality risks for individuals that cumulatively have been in lower socio-economic position (in ‘poverty’) compared to individuals who were never in poverty. This supports results found in other countries (Padyab et al., 2013), that being disadvantaged for a longer period in midlife has a significant negative effect on health. The effects vary between siblings and between two individuals. Using household income quintile as a socio-economic measure, the results are all significant for a sibling or individual who was in lowest household income quintile *twice or more* compared to a sibling or individual who was *never* in the lowest household income quintile between the years 1981–2006.

For a sibling that was in lowest household income quintile *once*, the results are insignificant compared to an individual that was also in lowest household income quintile *once*. This suggests that the significant results found comparing an individual that was in lowest household income quintile *once* to an individual that was *never* in lowest household income quintile may be confounded by factors like early childhood background.

It is interesting to note that using NZDep as a measure of socio-economic position does not give any significant results when comparing between siblings. The results are only significant for the individual cohort except for when adjusting for all groups of covariates. A possible explanation for this is that NZDep is an area-based socio-economic deprivation measure as opposed to household income which is a more direct measure of individual’s SEP.

Although New Zealand is relatively affluent when compared to other countries, these results suggest that being in lower socio-economic position does increase one’s mortality risks when compared to someone who was never in ‘low’ socio-economic position. These may be because lower socio-economic groups might

have restricted access to better medical care, or may be worrying about finances for livelihood causing cognitive stress.

5.1 Strengths and limitations

There are several strengths to this study. The availability of the data from the census allows us to conceptually make inferences of the whole population. Although computationally challenging, the longitudinal data source allows us to enhance our understanding of aspects of population change across time. With the large sample size, we are able to achieve statistically robust results.

Cox regression allows us to utilise our time-to-death information and Cox regression with fixed effects allow us to compare mortality risks between siblings. By doing a sibling cohort study, we were able to control for confounders like family background or genetic factors that could not be done otherwise.

However, there are several limitations that come with using census data. We have self-reporting issues and have to rely on information being filled out correctly. When identifying siblings, we depend on the fact that they identify as son or daughter to the occupier, but we cannot be sure if they are half sibling or non-biological child to the occupier. This may not achieve the use of siblings to control for things like genetic factors.

Moreover, factors like gestational age and birth weights are known to have effects on mortality risk but these factors cannot be included in our analyses due to limitation of the data.

5.2 Further work

Socio-economic measures used in this study were chosen based on the life-course variable that was constructed for NZLC (Liu, 2015). More sophisticated variables could be constructed as a better measure of socio-economic position. Jensen's equivalised household income is a better measure of household income as takes household composition into account. This measure of household income was

investigated but was not used due to incomplete information from SNZ resulting in large number of missing values. Hence, the life course variables were used instead.

Increased measuring power will probably be achieved if a longer follow-up period for mortality (e.g. 10-years from 2001-2011) for those linked across five censuses from 1981–2001 is used. That would increase the number of observed mortality outcomes for the cohort. Or alternatively, longer follow-up period could be achieved using more updated information from recent 2013 Census data when it is made available. But due to time constraint, analyses were done only on the 2006–2011 mortality records for the 1981–2006 linked population. Our cohorts and observed mortality outcomes are still relatively large-sized regardless.

The sibling analysis done in this study uses the fixed effects method which requires the data set to be structured in the '*long*' format where siblings are clustered within family. Alternatively, siblings could be paired with family information discarded; and analyses would be done using the difference scores between siblings as predictors.

Bibliography

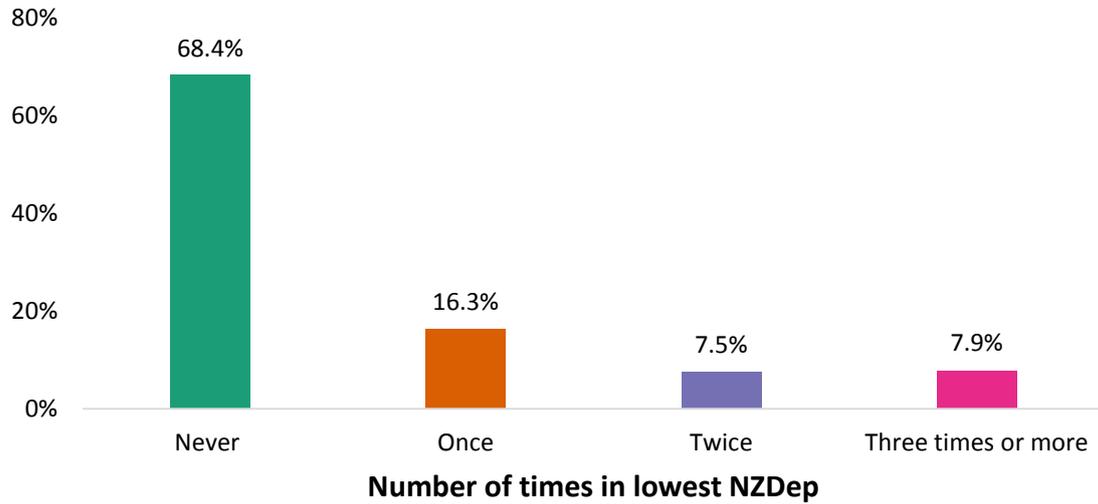
- AIHW. (2014). *Mortality inequalities in Australia 2009–2011*. Canberra: AIHW bulletin no. 124. Cat. no. AUS 184.
- Allison, P. D. (2005). *Fixed effects regression methods for longitudinal data using SAS*. SAS Institute.
- Blakely, T., M, T., J, A., Yeh, L.-C., & Huang, K. (2007). *Tracking Disparity: Trends in ethnic and socioeconomic inequalities in mortality, 1981–2004*. Wellington: Ministry of Health.
- Class, Q. A., Rickert, M. E., Lichtenstein, P., & D'Onofrio, B. M. (2014). Birth weight, physical morbidity, and mortality: a population-based sibling-comparison study. *American journal of epidemiology*, 550-558.
- Didham, R., Nissed, K., & Dobson, W. (2014). *Linking censuses: New Zealand longitudinal census 1981-2006*. Statistics New Zealand.
- Donovan, S. J., & Susser, E. (2011). Commentary: advent of sibling designs. *International journal of epidemiology*, 345-349.
- Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, 557-565.
- Fawcett, J., Blakely, T., & Atkinson, J. (2002). *Weighting the 81, 86, 91 & 96 census-mortality cohorts to adjust for linkage bias*. Department of Public Health, Wellington School of Medicine and Health Sciences.
- Fox, J. (2002). Cox proportional-hazards regression for survival data. *An R and S-PLUS companion to applied regression*, 1-18.
- Frisell, T., Öberg, S., Kuja-Halkola, R., & Sjölander, A. (2012). Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiology*, 713-720.
- Hertz-Picciotto, I., & Rockhill, B. (1997). Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics*, 1151-1156.
- Kuh, D., & Shlomo, Y. B. (2004). *A life course approach to chronic disease epidemiology (No. 2)*. Oxford University Press.

- Liu, C. (2015). *Construction of life-course variables for New Zealand Longitudinal Census (NZLC)*.
- Madsen, M., Andersen, A. M., Christensen, K., Andersen, P. K., & Osler, M. (2010). Does educational status impact adult mortality in Denmark? A twin approach. *American Journal of Epidemiology*.
- Næss, Ø., Claussen, B., Thelle, D. S., & Smith, G. D. (2004). Cumulative deprivation and cause specific mortality. A census based study of life course influences over three decades. *Journal of epidemiology and community health*, 599 - 603.
- Næss, Ø., Hoff, D. A., Lawlor, D., & Mortensen, L. H. (2012). Education and adult cause-specific mortality—examining the impact of family factors shared by 871 367 Norwegian siblings. *International journal of epidemiology*.
- Padyab, M., Malmberg, G., Norberg, M., & Blomstedt, Y. (2013). Life course socioeconomic position and mortality: A population register-based study from Sweden. *Scandinavian journal of public health*.
- Singal, R. (2015). *Adjusting for linkage bias in the New Zealand Longitudinal census*.
- Statistics New Zealand. (2014). *Linking censuses: New Zealand longitudinal census 1981–2006*. Wellington: Statistics New Zealand.
- Sudan, M., Kheifets, L. I., Arah, O. A., Divan, H. A., & Olsen, J. (2014). Complexities of sibling analysis when exposures and outcomes change with time and birth order. *Journal of Exposure Science and Environmental Epidemiology*, 482-488.
- Susser, E., Eide, M. G., & Begg, M. (2010). Invited commentary: the use of sibship studies to detect familial confounding. *American journal of epidemiology*, 537-539.
- Tan, L., Blakely, T. A., Tobias, M., & Atkinson, J. (2012). Mortality by ethnic group to 2006: is extending census-mortality linkage robust? *New Zealand medical journal*, 62-75.

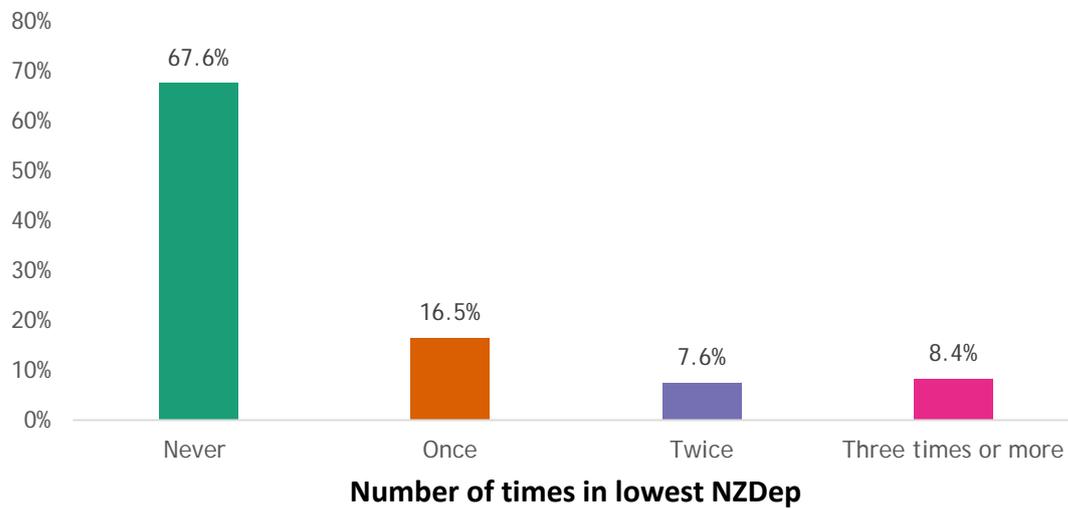
Appendix

7.1 Appendix A

Sibling cohort



Individual cohort



7.2 Appendix B

Model Specification

```
PROC PHREG DATA=fe2;
  WEIGHT=biaswgt;
  CLASS low_inc sex age_5yrs education_06
        index_: live_alone_: moved_in_lastyear_: partner_:
        health_: smoke_06 /ref=FIRST;
  MODEL months*dead(0) = low_inc sex age_5yrs education_06
        index_: live_alone_: moved_in_lastyear_: partner_:
        health_: smoke_06 /TIES=EFRON;
RUN;
```

Our individual cohort is saved in SAS in the work directory as fe2 and used in the `DATA` statement. The `WEIGHT` option specifies that each observation is weighted by the value of `biaswgt`. The variable `biaswgt` is the multiplication of the NZLC and NZCMS linkage biases as discussed in Sections 2.2.4.1 and 2.2.4.2. The `CLASS` statement specifies the classification variables used in our model - our main predictor, number of times in lowest household income quintile and all other variables. The semicolon used in the variables names (e.g. `index_:`) is a wildcard expression in SAS that takes any variable that follows the underscore. `/ref=FIRST` is specified to select the first ordered level in all class variables as reference. Lastly, the `MODEL` statement specifies `months*dead(0)` where `months` is the follow up time - 61 months between year 2006–2011 and `dead(0)` is the status indicator for whether the individual is dead or alive within the 61 months. The `/TIES=EFRON` is specified as an option in SAS software to use the Efron approximation as opposed to the SAS default that uses the Breslow approximation as it does not do well when the number of ties at a particular time point is a large proportion of the number of cases at risk (Hertz-Picciotto & Rockhill, 1997; Allison, 2005)

7.3 Appendix C

Scaled Schoenfeld residuals for all variables in the model

	rho	chisq	p
factor(low_inc)1	0.030994	1.55e+00	0.21241
factor(low_inc)2	-0.005690	3.64e-02	0.84879
factor(low_inc)3	-0.001823	3.42e-03	0.95334
factor(age_5yrs)2	0.017455	4.51e-01	0.50198
factor(age_5yrs)3	0.020886	5.65e-01	0.45207
factor(age_5yrs)4	-0.998847	8.84e+00	0.00294
factor(age_5yrs)5	-0.998852	8.84e+00	0.00294
factor(Sex)2	-0.008268	1.04e-01	0.74692
factor(education_06)1	0.002693	1.06e-02	0.91786
factor(education_06)2	-0.013204	2.78e-01	0.59824
factor(education_06)3	-0.002685	1.26e-02	0.91068
factor(education_06)999	0.014670	2.29e-01	0.63228
factor(index_86)1	-0.036801	2.07e+00	0.15017
factor(index_86)2	-0.012540	1.62e-01	0.68719
factor(index_86)999	-0.020854	6.29e-01	0.42787
factor(index_91)1	0.006422	6.20e-02	0.80330
factor(index_91)2	0.027855	7.90e-01	0.37407
factor(index_91)999	0.018280	5.52e-01	0.45738
factor(index_96)1	0.010148	1.44e-01	0.70460
factor(index_96)2	-0.004695	2.03e-02	0.88672
factor(index_96)999	-0.024761	8.09e-01	0.36836
factor(index_01)1	-0.014660	3.44e-01	0.55770
factor(index_01)2	0.008815	8.94e-02	0.76493
factor(index_01)999	-0.046015	2.36e+00	0.12461
factor(index_06)1	-0.029261	1.61e+00	0.20507
factor(index_06)2	-0.027974	8.40e-01	0.35954
factor(index_06)999	0.010970	5.86e-02	0.80871
factor(live_alone_86)1	0.001889	2.45e-03	0.96051
factor(live_alone_86)999	-0.023239	6.20e-01	0.43101
factor(live_alone_91)1	-0.000432	1.17e-04	0.99139
factor(live_alone_91)999	0.009539	3.53e-02	0.85094
factor(live_alone_96)1	0.004385	1.54e-02	0.90112
factor(live_alone_96)999	-0.022796	7.58e-01	0.38409
factor(live_alone_01)1	0.021188	4.45e-01	0.50496
factor(live_alone_01)999	-0.042387	1.40e+00	0.23683
factor(live_alone_06)1	-0.009678	8.91e-02	0.76529
factor(live_alone_06)999	0.015580	1.25e-01	0.72359
factor(partner_81)1	0.007394	4.01e-01	0.52653
factor(partner_81)999	-0.999264	8.85e+00	0.00292
factor(partner_86)1	-0.008496	8.46e-02	0.77122
factor(partner_86)999	0.019578	2.78e-01	0.59831
factor(partner_91)1	-0.040297	2.18e+00	0.13997
factor(partner_91)999	0.012894	2.50e-01	0.61682
factor(partner_96)1	0.055039	4.45e+00	0.03500
factor(partner_96)999	-0.018798	3.71e-01	0.54271
factor(partner_01)1	-0.016370	4.38e-01	0.50805
factor(partner_01)999	0.003780	4.20e-03	0.94831
factor(partner_06)1	0.004508	3.46e-02	0.85253
factor(partner_06)999	-0.024525	1.59e-01	0.69040
factor(moved_in_lastyear_86)1	-0.015606	3.40e-01	0.55988
factor(moved_in_lastyear_86)999	0.037405	1.30e+00	0.25475
factor(moved_in_lastyear_91)1	-0.006816	5.65e-02	0.81213
factor(moved_in_lastyear_91)999	0.053210	3.04e+00	0.08103
factor(moved_in_lastyear_96)1	0.029644	1.44e+00	0.22989
factor(moved_in_lastyear_96)999	-0.031659	1.14e+00	0.28619
factor(moved_in_lastyear_01)1	0.049482	3.78e+00	0.05174
factor(moved_in_lastyear_01)999	0.051745	1.74e+00	0.18732
factor(moved_in_lastyear_06)1	-0.015220	3.72e-01	0.54205
factor(moved_in_lastyear_06)999	-0.032656	5.25e-01	0.46889
factor(health_96)1	-0.000477	2.21e-04	0.98814
factor(health_96)999	0.020918	6.33e-01	0.42631
factor(health_01)1	0.012820	2.00e-01	0.65499
factor(health_01)999	0.037852	1.20e+00	0.27324
factor(health_06)1	-0.007659	7.41e-02	0.78539
factor(health_06)999	-0.003032	1.32e-02	0.90855

7.4 Appendix D

Reading in data

```
*****
      loading libnames where our data is stored
*****;

LIBNAME aa '\\wprdfs08\DataLab-MA\MAA2014-16 Life-course predictors of
mortality inequalities\Swen';
LIBNAME sib '\\wprdfs08\DataLab-MA\MAA2014-16 Life-course predictors
of mortality inequalities\Swen\Sib Project';
LIBNAME cc '\\wprdfs08\DataLab-MA\MAA2014-16 Life-course predictors of
mortality inequalities\Swen\Saved data set';
LIBNAME data1 '\\wprdfs08\RO-MAA2013-18 Linkage Bias Longitudinal
Census';
LIBNAME data2 '\\wprdfs08\RO-MAA2013-18 Linkage Bias Longitudinal
Census\Updated data dec 2013';
LIBNAME created '\\wprdfs08\RO-MAA2013-18 Linkage Bias Longitudinal
Census\Created data sets';
LIBNAME mort '\\wprdfs08\RO-MAA2013-18 Linkage Bias Longitudinal
Census\nzcms';
LIBNAME created2 '\\wprdfs08\Datalab-MA\MAA2014-16 Life-course
predictors of mortality inequalities\Swen\Created for income';
LIBNAME justin '\\wprdfs08\Datalab-MA\MAA2014-16 Life-course
predictors of mortality inequalities\Justin\Justin Modified Chris
Scripts\jdata';
LIBNAME eqinc '\\wprdfs08\Datalab-MA\MAA2014-16 Life-course predictors
of mortality inequalities\Swen\Equivalised income';

*****
      all individuals from 1981 - 2006
*****
* all individuals in 1981 census;
DATA individuals;
SET data1.individual_linkind8681_1981;
RUN;

* individuals from private dwelling;
DATA private;
SET individuals;
IF family_code_81 in ("01", "02", "03", "04", "05");
RUN;

PROC FREQ DATA=individuals; TABLES family_code_81; RUN;
PROC FREQ DATA=private; TABLES family_code_81 ; RUN;

PROC FREQ DATA=individuals nlevels; TABLES ID /noprint; RUN;
PROC FREQ DATA=private nlevels; TABLES ID /noprint; RUN;

* extracting those who can be linked from 1981-2006;
DATA spine;
SET data1.spine_dataset;
RUN;

DATA linked8106 (KEEP=ID Link8106);
SET spine;
IF LinkInd0601 = 1;
IF LinkInd0196 = 1;
```

```

IF LinkInd9691 = 1;
IF LinkInd9186 = 1;
IF LinkInd8681 = 1;
Link8106 = 1;
RENAME N=ID;
RUN;

PROC FREQ DATA=linked8106;
TABLES Link8106;
RUN;

PROC SORT data=linked8106; BY ID; RUN;
PROC SORT data=private; BY ID; RUN;

DATA indv8106;
MERGE private linked8106;
BY ID;
IF LinkInd8681 ~=. ;
IF Link8106 =1;
RUN;

PROC FREQ DATA=indv8106;
TABLES Link8106;
RUN;

*****
                individual information
*****
*** combine all necessary information together
* 1981 - 2006 info, mort info, ethnicity, income variables;

***** 1981 *****;
* extracting variables from 'spine';
DATA spine (KEEP= ID Year_of_birth Sex dwell_nbr_81 dwell_nbr_86
dwell_nbr_91 dwell_nbr_96 dwell_nbr_01 dwell_nbr_06) ;
SET data1.spine_dataset;
RENAME N=ID;
RUN;

* extracting variables from 'ethnicity_info' for 1981;
DATA eth (KEEP=ID eur06_81 Mao06_81 Pac06_81 Asian06_81 MELAA06_81
Other06_81 EthNS_81);
SET data1.ethnicity_info;
RENAME EthNS81 = EthNS_81
N=ID;
RUN;

* extracting variables from 'geogr_linkspine' for 1981;
DATA mb (KEEP=ID mb_81 CNSmb01_81);
SET data1.geogr_linkspine_8681;
mb_81=URSmb01_81;
RUN;

PROC SORT DATA=indv8106; BY ID; RUN;
PROC SORT DATA=spine; BY ID; RUN;
PROC SORT DATA=eth; BY ID; RUN;
PROC SORT DATA=mb; BY ID; RUN;

DATA info81;
MERGE indv8106 spine eth mb;
BY ID;

```

```

IF Link8106=1;
RUN;

***** 1986 *****;
* other info;
DATA ind86;
SET data1.individual_linkind8681_1986;
RUN;

* meshblock info;
DATA mb86 (KEEP=ID mb_86);
    SET data1.geogr_linkspine_9186;
    mb_86=URSmb01_86;
RUN;

* extracting variables from 'ethnicity_info' for 1986;
DATA eth86 (KEEP=ID eur06_86 Mao06_86 Pac06_86 Asian06_86 MELAA06_86
Other06_86 EthNS_86);
    SET data1.ethnicity_info;
RENAME    EthNS86 = EthNS_86
N=ID;
RUN;

PROC SORT DATA=info81; BY ID; RUN;
PROC SORT DATA=mb86; BY ID; RUN;
PROC SORT DATA=eth86; BY ID; RUN;
PROC SORT DATA=ind86; BY ID; RUN;

DATA info86;
MERGE info81 ind86 mb86 eth86;
BY ID;
IF Link8106=1;
RUN;

***** 1991 *****;
* other info;
DATA ind91;
SET data1.individual_linkind9186_1991;
RUN;

* meshblock info;
DATA mb91 (KEEP=ID mb_91);
    SET data1.geogr_linkspine_9186;
    mb_91=URSmb06_91;
RUN;

* extracting variables from 'ethnicity_info' for 1991;
DATA eth91 (KEEP=ID eur06_91 Mao06_91 Pac06_91 Asian06_91 MELAA06_91
Other06_91 EthNS_91);
    SET data1.ethnicity_info;
RENAME    EthNS91 = EthNS_91
N=ID;
RUN;

PROC SORT DATA=info86; BY ID; RUN;
PROC SORT DATA=mb91; BY ID; RUN;
PROC SORT DATA=eth91; BY ID; RUN;
PROC SORT DATA=ind91; BY ID; RUN;

DATA info91;
MERGE info86 ind91 mb91 eth91;

```

```

BY ID;
IF Link8106=1;
RUN;

***** 1996 *****;
* other info;
DATA ind96;
SET data1.individual_linkind9691_1996;
RUN;

* meshblock info;
DATA mb96 (KEEP=ID mb_96);
    SET data1.geogr_linkspine_9691;
    mb_96=URSmb06_96;
RUN;

* extracting variables from 'ethnicity_info' for 1996;
DATA eth96 (KEEP=ID eur06_96 Mao06_96 Pac06_96 Asian06_96 MELAA06_96
Other06_96 EthNS_96);
    SET data1.ethnicity_info;
RENAME    EthNS96 = EthNS_96
N=ID;
RUN;

PROC SORT DATA=info91; BY ID; RUN;
PROC SORT DATA=ind96; BY ID; RUN;
PROC SORT DATA=mb96; BY ID; RUN;
PROC SORT DATA=eth96; BY ID; RUN;

DATA info96;
MERGE info91 ind96 mb96 eth96;
BY ID;
IF Link8106=1;
RUN;

***** 2001 *****;
* other info;
DATA ind01;
SET data1.individual_linkind0196_2001;
RUN;

* meshblock info;
DATA mb01 (KEEP=ID mb_01);
    SET data1.geogr_linkspine_0196;
    mb_01=URSmb06_01;
RUN;

* extracting variables from 'ethnicity_info' for 2001;
DATA eth01 (KEEP=ID eur06_01 Mao06_01 Pac06_01 Asian06_01 MELAA06_01
Other06_01 EthNS_01);
    SET data1.ethnicity_info;
RENAME    EthNS01 = EthNS_01
N=ID;
RUN;

PROC SORT DATA=info96; BY ID; RUN;
PROC SORT DATA=mb01; BY ID; RUN;
PROC SORT DATA=eth01; BY ID; RUN;
PROC SORT DATA=ind01; BY ID; RUN;

DATA info01;

```

```

MERGE info96 ind01 mb01 eth01;
BY ID;
IF Link8106=1;
RUN;

***** 2006 *****;
* other info;
DATA ind06;
SET data1.individual_linkind0601_2006;
RENAME N=ID;
RUN;

* meshblock info;
DATA mb06 (KEEP=ID mb_06);
    SET data1.geogr_linkspine_0601;
    mb_06=URSmb06_06;
RUN;

* extracting variables from 'ethnicity_info' for 2006;
DATA eth06 (KEEP=ID eur06_06 Mao06_06 Pac06_06 Asian06_06 MELAA06_06
Other06_06 EthNS_06);
    SET data1.ethnicity_info;
RENAME EthNS06 = EthNS_06
N=ID;
RUN;

PROC SORT DATA=info01; BY ID; RUN;
PROC SORT DATA=ind06; BY ID; RUN;
PROC SORT DATA=mb06; BY ID; RUN;
PROC SORT DATA=eth06; BY ID; RUN;

* merging '06 info to all previous years' info;
* dropping EthNS_ variables as have problem merging later);
DATA info06 (DROP=EthNS_);
MERGE info01 ind06 mb06 eth06;
BY ID;
IF Link8106=1;
RUN;

*****
                life course variables
*****
* adding Chris' life course variables to our dataset;
DATA lcvar;
SET created.census_all_final;
KEEP ID low_inc_quintile_numtimes low_nzdep_numtimes
low_nzsei_numtimes fare_numtimes;
RUN;

***** 1981 *****;
DATA z81;
SET created.census1981_final;
KEEP ID NZDEP_81 EurOther_81--acc_81 welfare_81--moved_in_5years_81
nz_born_81--smoke_81 health_81--overall_dsb_81 tenure_81--phone_81
religion_none_81--religion_residual_81 Ethnicity_81--NZSEI_81;
RUN;

***** 1986 *****;
DATA z86;
SET created.census1986_final;

```

```

KEEP ID NZDEP_86 EurOther_86--acc_86 welfare_86--moved_in_5years_86
nz_born_86--smoke_86 health_86--overall_dsb_86 tenure_86--phone_86
religion_none_86--religion_residual_86 Ethnicity_86--NZSEI_86;
RUN;

```

```

***** 1991 *****;

```

```

DATA z91;
SET created.census1991_final;
KEEP ID NZDEP_91 EurOther_91--acc_91 welfare_91--moved_in_5years_91
nz_born_91--smoke_91 health_91--overall_dsb_91 tenure_91--phone_91
religion_none_91--religion_residual_91 Ethnicity_91--NZSEI_91;
RUN;

```

```

***** 1996 *****;

```

```

DATA z96;
SET created.census1996_final;
KEEP ID NZDEP_96 EurOther_96--acc_96 welfare_96--moved_in_5years_96
nz_born_96--smoke_96 health_96--overall_dsb_96 tenure_96--phone_96
religion_none_96--religion_residual_96 Ethnicity_96--NZSEI_96;
RUN;

```

```

***** 2001 *****;

```

```

DATA z01;
SET created.census2001_final;
KEEP ID NZDEP_01 EurOther_01--acc_01 welfare_01--moved_in_5years_01
nz_born_01--smoke_01 health_01--overall_dsb_01 tenure_01--phone_01
religion_none_01--religion_residual_01 Ethnicity_01--NZSEI_01;
RUN;

```

```

***** 2006 *****;

```

```

DATA z06;
SET created.census2006_final;
KEEP ID NZDEP_06 EurOther_06--acc_06 welfare_06--moved_in_5years_06
nz_born_06--smoke_06 health_06--overall_dsb_06 tenure_06--phone_06
religion_none_06--religion_residual_06 Ethnicity_06--NZSEI_06;
RUN;

```

```

PROC SORT DATA=info06; BY ID; RUN;
PROC SORT DATA=lcvar; BY ID; RUN;
PROC SORT DATA=z81; BY ID; RUN;
PROC SORT DATA=z86; BY ID; RUN;
PROC SORT DATA=z91; BY ID; RUN;
PROC SORT DATA=z96; BY ID; RUN;
PROC SORT DATA=z01; BY ID; RUN;
PROC SORT DATA=z06; BY ID; RUN;

```

```

* merging NZLC info and Chris' variables together;

```

```

DATA allvar;
MERGE info06 lcvar z81 z86 z91 z96 z01 z06;
BY ID;
IF Link8106=1;
RUN;

```

```

*****

```

```

* indiv from private dwellings;

```

```

* hard-coding the variables;

```

```

DATA sib.indvinfo;
SET allvar;
RUN;

```

```

DATA allvar;

```

```

SET sib.indvinfo;
RUN;
*****

*****
                mortality info
*****
* FULL INDV;
* merging mortality info to individuals
* only using ID from indivID to merge all mort info together first;
DATA indivID;
SET sib.indvinfo;
KEEP ID Link8106;
RUN;

***** mortality info for 1996 - 2006*****;
* 1991 doesn't have enough information and we don't need it anyway;
DATA a06; SET mort.nzlc0601_nzcms06_link; RUN;
DATA a01; SET mort.nzlc0196_nzcms01_link; RUN;
DATA a96; SET mort.nzlc9691_nzcms96_link; RUN;

DATA b06; SET mort.nzcms_lcs_2006; RUN;
DATA b01; SET mort.nzcms_lcs_2001; RUN;
DATA b96; SET mort.nzcms_lcs_1996; RUN;

* 2006 mort info;
* merging '06 mort records with 0601 IDs;
PROC SORT DATA=a06; BY ID_CTMort; RUN;
PROC SORT DATA=b06; BY ID_CTMort; RUN;

DATA c06 (DROP=N);
MERGE a06 b06;
BY ID_CTMort;
ID=N;
IF ID ~= .;
RUN;

PROC SORT DATA=indvID; BY ID; RUN;
PROC SORT DATA=c06; BY ID; RUN;

* merging mort info onto sibID;
DATA death06;
MERGE indivID c06;
BY ID;
IF Link8106=1;
RENAME ICD_Gp = ICD_Gp_06
ICD_Gp_Lab = ICD_Gp_Lab_06
AgeC_mths = AgeC_mths_06
AgeC_mths_Lab = AgeC_mths_Lab_06
AgeD_mths = AgeD_mths_06
AgeD_mths_Lab = AgeD_mths_Lab_06
W_AgEthAdj = W_AgEthAdj_06
Link=Link_06;
RUN;

* 2001 mort info;
* merging '01 mort records with 0601 IDs;
PROC SORT DATA=a01; BY ID_Cohort; RUN;
PROC SORT DATA=b01; BY ID_Cohort; RUN;

DATA c01 (DROP=N ID_Dwell);

```

```

MERGE a01 b01;
BY ID_Cohort;
ID=N;
IF ID ~= .;
RUN;

PROC SORT DATA=c01; BY ID; RUN;
PROC SORT DATA=death06; BY ID; RUN;

DATA death01;
MERGE death06 c01;
BY ID;
IF Link8106=1;
RENAME ICD_Gp = ICD_Gp_01
ICD_Gp_Lab = ICD_Gp_Lab_01
AgeC_mths = AgeC_mths_01
AgeC_mths_Lab = AgeC_mths_Lab_01
AgeD_mths = AgeD_mths_01
AgeD_mths_Lab = AgeD_mths_Lab_01
W_AgEthAdj = W_AgEthAdj_01
Link=Link_01;
RUN;

* 1996 mort info;
* merging '96 mort records with 0601 IDs;
PROC SORT data=a96; by ID_Cohort; run;
PROC SORT data=b96; by ID_Cohort; run;

DATA c96 (DROP=N);
MERGE a96 b96;
BY ID_Cohort;
ID=N;
IF ID ~= .;
RUN;

PROC SORT DATA=c96; BY ID; RUN;
PROC SORT DATA=death01; BY ID; RUN;

DATA death96;
MERGE death01 c96;
BY ID;
IF Link8106=1;
RENAME ICD_Gp = ICD_Gp_96
ICD_Gp_Lab = ICD_Gp_Lab_96
AgeC_mths = AgeC_mths_96
AgeC_mths_Lab = AgeC_mths_Lab_96
AgeD_mths = AgeD_mths_96
AgeD_mths_Lab = AgeD_mths_Lab_96
CauseDeath = CauseDeath_96
CauseDeath_Lab = CauseDeath_Lab_96
W_AgEthAdj = W_AgEthAdj_96
Link=Link_96;
RUN;

* saving mort information for all years;
DATA sib.mortinfo;
SET death96;
BY ID;
RUN;

```

```

*****
merging indiv info, mort info, bias wgts together
*****;
* bias weights;
DATA biaswgt (KEEP=ID biaswgt);
SET justin.confounders_16th_attempt_6;
RENAME w8106 = biaswgt;
RUN;

DATA biasID (KEEP=ID S1);
SET sib.allinfo;
S1 = 1;
RUN;

PROC SORT DATA=biaswgt; BY ID; RUN;
PROC SORT DATA=biasID; BY ID; RUN;

DATA indivbias;
MERGE biasID biaswgt;
BY ID;
IF S1=1;
RUN;

* hard coding bias weights;
DATA sib.indvBias (DROP=S1);
SET indivbias;
RUN;

*****
importing updated crowding index
*****;
* 2006;
DATA index06;
SET DATA2.crowding_index_2006;
RENAME N = ID;
RUN;

* 2001;
DATA index01;
SET DATA2.crowding_index_2001;
RENAME N = ID;
RUN;

* 1996;
DATA index96;
SET DATA2.crowding_index_1996;
RENAME N = ID;
RUN;

* 1991;
DATA index91;
SET DATA2.crowding_index_1991;
RENAME N = ID;
RUN;

* 1986;
DATA index86;
SET DATA2.crowding_index_1986;
RENAME N = ID;
RUN;

```

```

* merging data together;
PROC SORT DATA=index86; BY ID; RUN;
PROC SORT DATA=index91; BY ID; RUN;
PROC SORT DATA=index96; BY ID; RUN;
PROC SORT DATA=index01; BY ID; RUN;
PROC SORT DATA=index06; BY ID; RUN;

*****
* full indiv;
DATA indivinfo1 (KEEP=ID S1 DROP=index_);
SET sib.indvinfo;
S1=1;
RUN;

PROC SORT DATA=indvinfo1; BY ID; RUN;

DATA index1;
MERGE indivinfo1 (in=indata) index86 index91 index96 index01 index06;
BY ID;
IF S1=1;
RUN;

DATA index2;
MERGE index1 index91;
BY ID;
IF S1=1;
RUN;

* checks;
PROC FREQ DATA=index86; TABLES crowding_index_86 /missing; RUN;
PROC FREQ DATA=index91; TABLES crowding_index_91 /missing; RUN;
PROC FREQ DATA=index2; TABLES crowding_index_ : /missing; RUN;

* renaming the variables;
DATA index2;
SET index1;
IF crowding_index_86 = 'Unknown' THEN index_86 = 999;
IF crowding_index_86 = . THEN index_86 = 999;
IF crowding_index_86 in ('01 Severely Crowded', '02 Crowded') THEN
index_86 = 2;
IF crowding_index_86 = '03 No extra bedrooms requi' THEN index_86 = 1;
IF crowding_index_86 in ('04 One Bedroom Spare', '05 Two or more
bedrooms sp') THEN index_86 = 0;

IF crowding_index_91 = 'Unknown' THEN index_91 = 999;
IF crowding_index_91 = . THEN index_91 = 999;
IF crowding_index_91 in ('01 Crowded', '02 Severely Crowded') THEN
index_91 = 2;
IF crowding_index_91 = '03 No Extra Bedrooms Requi' THEN index_91 = 1;
IF crowding_index_91 in ('04 One Bedroom Spare', '05 Two or more
Bedrooms Sp') THEN index_91 = 0;

IF crowding_index_96 = . THEN index_96 = 999;
IF crowding_index_96 = 'Unknown' THEN index_96 = 999;
IF crowding_index_96 in ('01 Crowded', '02 severely crowded') THEN
index_96 = 2;
IF crowding_index_96 = '03 No Extra Bedrooms Requi' THEN index_96 = 1;
IF crowding_index_96 in ('04 One Bedroom Spare', '05 Two or more
Bedrooms Sp') THEN index_96 = 0;

IF crowding_index_01 = . THEN index_01 = 999;

```

```

IF crowding_index_01 = 'Unknown' THEN index_01 = 999;
IF crowding_index_01 in ('01 Crowded', '02 severely crowded') THEN
index_01 = 2;
IF crowding_index_01 = '03 No Extra Bedrooms Requi' THEN index_01 = 1;
IF crowding_index_01 in ('04 One Bedroom Spare', '05 Two or more
Bedrooms Sp') THEN index_01 = 0;

IF crowding_index_06 in ( . , 99) THEN index_06 = 999;
IF crowding_index_06 in ('01 2+ bedrooms needed', '02 1 bedroom
needed') THEN index_06 = 2;
IF crowding_index_06 = '03 No extra beds needed' THEN index_06 = 1;
IF crowding_index_06 in ('04 1 bedroom spare', '05 2+ bedrooms spare')
THEN index_06 = 0;
RUN;

* checks;
PROC FREQ DATA=index1; tables crowding_index_01: /missing; RUN;
PROC FREQ DATA=index2; tables index_06: /missing; RUN;

* full indv;
* hard coding;
DATA sib.indvCrowdIndex;
SET index2;
RUN;

* merging all info together;
PROC SORT DATA=sib.indvinfo; BY ID; RUN;
PROC SORT DATA=sib.mortinfo; BY ID; RUN;
PROC SORT DATA=sib.indvBias; BY ID; RUN;
PROC SORT DATA=sib.mortinfo; BY ID; RUN;
PROC SORT DATA=sib.indvCrowdIndex; BY ID; RUN;

DATA sib.indvAll;
MERGE sib.indvinfo sib.mortinfo sib.indvBias sib.mortinfo
sib.indvCrowdIndex;
BY ID;
RUN;

*****;
DATA test;
SET sib.indvAll;
RUN;

PROC CORR DATA=test;
VAR low_inc_quintile_numtimes low_nzdep_numtimes low_nzsei_numtimes
fare_numtimes link;
RUN;

PROC LOGISTIC data=test;
MODEL link_06 (ref='0')= low_inc_quintile_numtimes;
RUN;

PROC LOGISTIC data=test;
MODEL link_06 (ref='0') = low_nzdep_numtimes;
RUN;

*****
recoding variables for more sensible interpretation
*****;
DATA rc1 (DROP= index_81 live_alone_81 moved_in_lastyear_81
disability_81

```

```

disability_86 disability_91 smoke_86 smoke_91 smoke_01 overall_dsb_81
overall_dsb_86
overall_dsb_91 health_81 health_86 health_91);
SET sib.allindvAll;
RENAME low_inc_quintile_numtimes = low_inc
low_nzdep_numtimes = low_nzdep
low_nzsei_numtimes = low_nzsei
fare_numtimes = low_fare;
RUN;

* check if variables are coded properly;
PROC FREQ DATA=rcl; TABLES low_: index_: age_5yrs sex education_:
live_alone_: moved_in_lastyear_: partner_: health_: smoke_:; RUN;

* recoding variables;
DATA rc2;
SET rcl;
IF Communicating_socialising_96 = '2' THEN
communicating_socialising_96 = 1;
IF Communicating_socialising_96 = '9' THEN
Communicating_socialising_96 = 0;

avg_eth_density = (sum(Ethnic_density_01, Ethnic_density_06,
Ethnic_density_81, Ethnic_density_86, Ethnic_density_91,
Ethnic_density_96))/6;

IF education_06 = 9 THEN education_06 = .;

IF Everyday_Activities_96='9' THEN Everyday_Activities_96 =0;

IF internet_01 in (2,8,9) THEN internet_01= .;
IF internet_06 in (2,8,9) THEN internet_06= .;

IF job_ind_code_06 = 2 THEN job_ind_code_06 = 0;

IF live_alone_86 = 9 THEN live_alone_86 = .;
IF live_alone_91 = 9 THEN live_alone_91 = .;
IF live_alone_96 = 9 THEN live_alone_96 = .;
IF live_alone_01 = 9 THEN live_alone_01 = .;
IF live_alone_06 = 9 THEN live_alone_06 = .;

IF Long_Term_Disability_96 = 2 THEN Long_Term_Disability_96 = 0;

IF motor_81 in (7,9) then motor_81 = .;
IF motor_86 in (2,7,8,9) then motor_86 = .;
IF motor_91 in (2,7,8,9) then motor_91 = .;
IF motor_96 in (2,8,9) then motor_96 = .;
IF motor_01 in (2,8,9) then motor_01 = .;
IF motor_06 in (2,8,9) then motor_06 = .;

IF partner_81 in (0,1,2) THEN partner_81 = 0;
IF partner_81 = 3 THEN partner_81 = 1;
IF partner_86 in (0,1,2) THEN partner_86 = 0;
IF partner_86 = 3 THEN partner_86 = 1;
IF partner_86 = 9 THEN partner_86 = 999;
IF partner_91 in (0,1,2) THEN partner_91 = 0;
IF partner_91 = 3 THEN partner_91 = 1;
IF partner_91 = 9 THEN partner_91 = 999;
IF partner_96 in (0,1,2) THEN partner_96 = 0;
IF partner_96 = 3 THEN partner_96 = 1;
IF partner_96 = 9 THEN partner_96 = 999;

```

```

IF partner_01 in (0,1,2) THEN partner_01 = 0;
IF partner_01 = 3 THEN partner_01 = 1;
IF partner_01 = 9 THEN partner_01 = 999;
IF partner_06 in (0,1,2) THEN partner_06 = 0;
IF partner_06 = 3 THEN partner_06 = 1;
IF partner_06 = 9 THEN partner_06 = 999;

IF phone_81 = 9 THEN phone_81 = .;
IF phone_96 in (2,8,9) THEN phone_96 = .;
IF phone_01 in (2,8,9) THEN phone_01 = .;
IF phone_06 in (2,8,9) THEN phone_06 = .;

IF smoking_status_code_06 = 3 THEN ever_smoke_reg = 0;
IF smoking_status_code_06 in (1,2) THEN ever_smoke_reg = 1;
IF smoking_status_code_06 in (7,9) THEN ever_smoke_reg = .;

IF tenure_81 = 9 THEN tenure_81 = .;
IF tenure_86 in (8,9) THEN tenure_86 = .;
IF tenure_91 in (8,9) THEN tenure_91 = .;
IF tenure_96 in (8,9) THEN tenure_96 = .;
IF tenure_01 in (8,9) THEN tenure_01 = .;
IF tenure_06 in (8,9) THEN tenure_06 = .;

IF means_travel_work_81 in (8, 9, 10) THEN travel_work_81 = 1;
IF means_travel_work_81 in (2, 3, 4, 7) THEN travel_work_81 = 2;
IF means_travel_work_81 in (0, 1) THEN travel_work_81 = 3;
IF means_travel_work_81 in (5, 6) THEN travel_work_81 = 4;
IF means_travel_work_81 in (10) THEN travel_work_81 = 9;

if means_travel_work_86 in ('10') then travel_work_86 = 1;
if means_travel_work_86 in ('03','04','05','07','09') then
travel_work_86 = 2;
if means_travel_work_86 in ('01','02') then travel_work_86 = 3;
if means_travel_work_86 in ('06','08') then travel_work_86 = 4;
if means_travel_work_86 in ('99') then travel_work_86 = 9;

if means_of_travel_to_work_91 in (1,10) then travel_work_91 = 1;
if means_of_travel_to_work_91 in (2,3,4,7,11) then travel_work_91 = 2;
if means_of_travel_to_work_91 in (5,6) then travel_work_91 = 3;
if means_of_travel_to_work_91 in (8,9) then travel_work_91 = 4;
if means_of_travel_to_work_91 in (99) then travel_work_91 = 9;

if Means_of_Travel_to_Work_96 in ('01','02','PP') then travel_work_96
= 1;
if Means_of_Travel_to_Work_96 in
('05','06','07','08','11','12','13','14','15') then travel_work_96 = 2;
if Means_of_Travel_to_Work_96 in ('03','04') then travel_work_96 = 3;
if Means_of_Travel_to_Work_96 in ('09','10') then travel_work_96 = 4;
if Means_of_Travel_to_Work_96 in ('99') then travel_work_96 = 9;

if travel_work_code_01 in (1,2) then travel_work_01 = 0;
if travel_work_code_01 in (3,4,5,8) then travel_work_01 = 1;
if travel_work_code_01 in (6,7,15) then travel_work_01 = 2;
if travel_work_code_01 in (9,10) then travel_work_01 = 3;
if travel_work_code_01 in (99) then travel_work_01 = 9;

if travel_work_code_06 in (1,2) then travel_work_06 = 0;
if travel_work_code_06 in (3,4,5,8) then travel_work_06 = 1;
if travel_work_code_06 in (6,7,15) then travel_work_06 = 2;
if travel_work_code_06 in (9,10) then travel_work_06 = 3;
if travel_work_code_06 in (77,99) then travel_work_06 = 9;

```

```

RUN;

DATA rc3;
SET rc2;
IF low_inc in (3,4,5,6) then low_inc=3;
IF low_nzdep in (4) then low_nzdep=3;
IF low_nzsei in (4,5,6) then low_nzsei=3;
IF low_fare in (4,5,6) then low_fare=3;

IF 0<=age_81<=4 THEN age_5yrs =1;
IF 5<=age_81<=9 THEN age_5yrs = 2;
IF 10<=age_81<=14 THEN age_5yrs = 3;
IF 15<=age_81<=19 THEN age_5yrs = 4;
IF 20<=age_81<=1000 THEN age_5yrs = 5;

IF education_06=. THEN education_06=999;

IF health_96 = 9 THEN health_96 = 999;
IF health_01 = 9 THEN health_01 = 999;
IF health_06 = 9 THEN health_06 = 999;

IF live_alone_86 =. THEN live_alone_86 = 999;
IF live_alone_91 =. THEN live_alone_91 = 999;
IF live_alone_96 =. THEN live_alone_96 = 999;
IF live_alone_01 =. THEN live_alone_01 = 999;
IF live_alone_06 =. THEN live_alone_06 = 999;

IF moved_in_lastyear_86=9 THEN moved_in_lastyear_86=999;
IF moved_in_lastyear_91=9 THEN moved_in_lastyear_91=999;
IF moved_in_lastyear_96=9 THEN moved_in_lastyear_96=999;
IF moved_in_lastyear_01=9 THEN moved_in_lastyear_01=999;
IF moved_in_lastyear_06=9 THEN moved_in_lastyear_06=999;

IF partner_81 in (., 9) THEN partner_81=999;
IF partner_86=. THEN partner_86=999;
IF partner_91=. THEN partner_91=999;
IF partner_96=. THEN partner_96=999;
IF partner_01=. THEN partner_01=999;
IF partner_06=. THEN partner_06=999;

IF smoke_06 = 9 THEN smoke_06 = 999;
RUN;

PROC FREQ DATA=rc3;
TABLES age_5yrs sex low_ index_ Link_06 education_06 live_alone_
moved_in_lastyear_ partner_ health_ smoke_06 ;
RUN;

* hard code recoded indiv info;
DATA sib.indvfull;
SET rc3;
RUN;

*****
      finding siblings from individuals
*****
* restricting individuals to those identified as son/daughter to
occupier;
DATA kids;
SET sib.indvfull;

```

```

IF relationship_to_head_81='03';
RUN;

* checking if everyone have rel_.. = '03';
PROC FREQ DATA=kids; TABLES relationship_to_head_81; RUN;
PROC FREQ DATA=kids nlevels; TABLES ID /noprint; RUN;

* getting siblings out;
* restricting to only where there are more than two son/daughters in
the family;
PROC SORT DATA=kids; BY family_nbr_81; RUN;

DATA sib;
SET kids;
BY family_nbr_81;
age = input(age_81,f3.0);
IF (first.family_nbr_81 ne 1 or last.family_nbr_81 ne 1) THEN output;
RUN;

* check how many are not from code = '01';
PROC FREQ DATA=sib;
TABLES family_code_81 relationship_to_head_81;
RUN;

PROC FREQ DATA=sib nlevels; TABLES ID /noprint; RUN;

* restrict age_diff in sibling to less than 20;
PROC SORT DATA=sib; BY family_nbr_81 age_81; RUN;

* computing difference between siblings within family;
DATA sibagediff;
SET sib;
BY family_nbr_81;
age_diff = dif(age_81);
IF first.family_nbr_81 THEN DO;
age_diff=.;
END;
RUN;

* taking the sum of age_diff to get max age_diff between siblings;
PROC SQL;
CREATE TABLE aged20 as SELECT *, sum(age_diff) as sumdiff FROM
sibagediff
GROUP BY family_nbr_81;
QUIT;

DATA aged20a;
SET aged20;
absSum = abs(sumdiff);
RUN;

PROC FREQ DATA=aged20a;
TABLES sumdiff;
RUN;

* restricting age difference between siblings to less than 20;
* removing sib where age_diff >= 20;
DATA sibless20;
SET aged20;
IF ABS(sumdiff) <= 20;
RUN;

```

```

PROC FREQ DATA=sibless20 nlevels; TABLES ID family_nbr_81 /noprint;
RUN;

** need to check how many families have 6 siblings, 5 etc.. ;
PROC FREQ DATA=sibless20 ORDER=FREQ;
TABLES family_nbr_81 /NOPRINT OUT=test;
RUN;

PROC FREQ DATA=test;
TABLES COUNT;
RUN;

DATA sib.sibfull;
SET sibless20;
RUN;

PROC FREQ DATA=sib.indvfull nlevels; TABLES ID family_nbr_81 /noprint;
RUN;
PROC FREQ DATA=sib.sibfull nlevels; TABLES ID family_nbr_81 /noprint;
RUN;

PROC FREQ DATA=sib.sibfull;
TABLES age_5yrs sex low_ index_ Link_06 education_06 live_alone_
moved_in_lastyear_ partner_ health_ smoke_06 ;
RUN;

```

7.5 Appendix E

Model

```
*****
                        Individual cohort
*****;
* full INDIVIDUALS;
DATA fe1;
SET sib.indvfull;
RUN;

* creating censored variables to run cox regression;
DATA fe2 (DROP=Health_Problems_96);
SET fe1;
months = AgeD_mths_06 - AgeC_mths_06;
IF Link_06=0 then months = 61;
RENAME Link_06 = dead ;
RUN;

* replacing low_inc with other main predictors;
PROC PHREG DATA=fe2;
WEIGHT=biaswgt;
CLASS low_inc sex age_5yrs /ref=FIRST;
MODEL months*dead(0) = low_inc age_5yrs sex /TIES=EFRON;
hazardratio low_nzsei /diff=ref;
RUN;

PROC PHREG DATA=fe2;
WEIGHT=biaswgt;
CLASS sex age_5yrs /ref=FIRST;
MODEL months*dead(0) = low_inc age_5yrs sex /TIES=EFRON;
hazardratio low_nzsei /diff=ref;
RUN;

* attempt to export as CSV to try in R;
DATA fe3;
SET fe2;
KEEP ID family_nbr_81 dead months biaswgt low_ age_5yrs sex
education_06 index_ live_alone_ moved_in_lastyear_ partner_
health_ smoke_06;
RUN;

PROC EXPORT DATA=fe3
    outfile='\\wprdfs08\DataLab-MA\MAA2014-16 Life-course predictors
of mortality inequalities\Swen\Sib Project\fullindv.csv'
    dbms=csv
    replace;
RUN;

* checking if variables are coded properly;
PROC FREQ DATA= fe2;
TABLE age_5yrs sex low_ index_ dead education_06 live_alone_
moved_in_lastyear_ partner_ health_ smoke_06;
RUN;

* replacing low_inc with other main predictors;
PROC PHREG DATA=fe2;
WEIGHT=biaswgt;
```

```

CLASS low_inc sex age_5yrs /ref=FIRST;
MODEL months*dead(0) = low_inc age_5yrs sex /TIES=EFRON;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=fe2;
WEIGHT=biaswgt;
CLASS low_inc sex age_5yrs education_06 /ref=FIRST;
MODEL months*dead(0) = low_inc sex age_5yrs education_06 /TIES=EFRON;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=fe2;
WEIGHT=biaswgt;
CLASS low_inc sex age_5yrs education_06 index_:/ref=FIRST;
MODEL months*dead(0) = low_inc sex age_5yrs education_06
index_:/TIES=EFRON;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=fe2;
WEIGHT=biaswgt;
CLASS low_inc sex age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_:/ref=FIRST;
MODEL months*dead(0) = low_inc sex age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_:/TIES=EFRON;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=fe2;
WEIGHT=biaswgt;
CLASS low_inc sex age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_:
smoke_06/ref=FIRST;
MODEL months*dead(0) = low_inc sex age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_:
smoke_06/TIES=EFRON;
hazardratio low_inc /diff=ref;
RUN;

*****
                Sibling cohort
*****;
* SIBLINGS;
DATA sibfel;
SET sib.sibfull;
RUN;

* sort data set by family number and age in descending order to
construct birth order;
PROC SORT DATA=sibfel; BY family_nbr_81 descending age_81; RUN;

* construct birth order;
DATA sibfe2;
SET sibfel;
bthorder + 1;
by family_nbr_81;
IF first.family_nbr_81 THEN bthorder=1;
RUN;

```

```

* creating censored variables to run cox regression;
DATA sibfe3 (DROP=Health_Problems_96);
SET sibfe2;
months = AgeD_mths_06 - AgeC_mths_06;
IF Link_06=0 then months = 61;
RENAME Link_06 = dead ;
IF bthorder in (4,5,6) THEN bthorder=3;
RUN;

PROC FREQ DATA=sibfe3;
TABLE dead age_5yrs sex bthorder biaswgt low_: index_: education_06
live_alone_: moved_in_lastyear_: partner_: health_: smoke_06;
RUN;

PROC CORR DATA=sibfe3 (KEEP=low_: age sex bthorder index_: live_alone_:
moved_in_lastyear_: partner_: health_: smoke_:)
OUT = corrcoef (where=( _type_='CORR' ));
RUN;

* attempt to export as CSV to try in R;
DATA sibfe4;
SET sibfe3;
KEEP ID family_nbr_81 dead months bthorder biaswgt low_: age_5yrs sex
bthorder education_06 index_: live_alone_: moved_in_lastyear_:
partner_: health_: smoke_06;
RUN;

PROC EXPORT DATA=sibfe4
    outfile='\\wprdfs08\DataLab-MA\MAA2014-16 Life-course predictors
of mortality inequalities\Swen\Sib Project\sibindv.csv'
    dbms=csv
    replace;
RUN;

*****
        Cox regression without fixed effects
*****;
PROC PHREG DATA=sibfe3 COVSANDWICH(AGGREGATE);
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs /ref=FIRST;
MODEL months*dead(0) = low_inc age_5yrs bthorder sex /TIES=EFRON;
ID family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=sibfe3 COVSANDWICH(AGGREGATE);
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06 /ref=FIRST;
MODEL months*dead(0) = low_inc sex bthorder age_5yrs education_06
/TIES=EFRON;
ID family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=sibfe3 COVSANDWICH(AGGREGATE);
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06
index_:/ref=FIRST;
MODEL months*dead(0) = low_inc sex bthorder age_5yrs education_06
index_:/TIES=EFRON;
ID family_nbr_81;

```

```

hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=sibfe3 COVSANDWICH(AGGREGATE);
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_:/ref=FIRST;
MODEL months*dead(0) = low_inc sex bthorder age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_:/TIES=EFRON;
ID family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=sibfe3 COVSANDWICH(AGGREGATE);
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_:
smoke_06/ref=FIRST;
MODEL months*dead(0) = low_inc sex bthorder age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_:
smoke_06/TIES=EFRON;
ID family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;

*****
Cox regression with fixed effects
*****;
* adding covariates sequentially;
PROC PHREG DATA=sibfe3 NOSUMMARY;
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs /ref=FIRST;
MODEL months*dead(0) = low_inc age_5yrs bthorder sex /TIES=EFRON;
STRATA family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=sibfe3 NOSUMMARY;
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06 /ref=FIRST;
MODEL months*dead(0)= low_inc age_5yrs bthorder sex education_06 ;
STRATA family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=sibfe3 NOSUMMARY;
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06 index_: /ref=FIRST;
MODEL months*dead(0) = low_inc age_5yrs bthorder sex education_06
index_:/TIES=EFRON;
STRATA family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;

PROC PHREG DATA=sibfe3 NOSUMMARY;
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_:/ref=FIRST;
MODEL months*dead(0) = low_inc age_5yrs bthorder sex education_06
index_: live_alone_:
moved_in_lastyear_: partner_:/TIES=EFRON;

```

```
STRATA family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;
```

```
PROC PHREG DATA=sibfe3 NOSUMMARY;
WEIGHT=biaswgt;
CLASS low_inc sex bthorder age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_:
smoke_06/ref=FIRST;
MODEL months*dead(0) = low_inc sex bthorder age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_:
smoke_06/TIES=EFRON;
STRATA family_nbr_81;
hazardratio low_inc /diff=ref;
RUN;
```

Survival curves

```
*****
Survival curves
*****;
* FULL INDIVIDUALS;
DATA survCurve;
SET sib.indvfull;
RUN;

* creating censored variables to run cox regression;
DATA survCurve1 (DROP=Health_Problems_96);
SET survCurve;
months = AgeD_mths_06 - AgeC_mths_06;
IF Link_06=0 then months = 61;
RENAME Link_06 = dead ;
RUN;

PROC FREQ data=survCurve;
tables low_inc sex age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_: smoke_06;
RUN;

DATA survCurve2;
SET survCurve1;
sex_81 = input(sex, f3.0);
RUN;

DATA fe3;
SET survCurve2;
KEEP dead months ID biaswgt family_nbr_81 low_inc sex_81 age_5yrs
age_81 education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_: smoke_06;
RUN;

* mean for all variables;
PROC SQL;
CREATE TABLE av_data AS SELECT
mean(age_5yrs) as age_5yrs,
mean(sex_81) as sex_81,
mean(education_06) as education_06,
mean(index_86) as index_86,
mean(index_91) as index_91,
mean(index_96) as index_96,
```

```

mean(index_01) as index_01,
mean(index_06) as index_06,
mean(live_alone_86) as live_alone_86,
mean(live_alone_91) as live_alone_91,
mean(live_alone_96) as live_alone_96,
mean(live_alone_01) as live_alone_01,
mean(live_alone_06) as live_alone_06,
mean(moved_in_lastyear_86) as moved_in_lastyear_86,
mean(moved_in_lastyear_91) as moved_in_lastyear_91,
mean(moved_in_lastyear_96) as moved_in_lastyear_96,
mean(moved_in_lastyear_01) as moved_in_lastyear_01,
mean(moved_in_lastyear_06) as moved_in_lastyear_06,
mean(partner_81) as partner_81,
mean(partner_86) as partner_86,
mean(partner_91) as partner_91,
mean(partner_96) as partner_96,
mean(partner_01) as partner_01,
mean(partner_06) as partner_06,
mean(health_96) as health_96,
mean(health_01) as health_01,
mean(health_06) as health_06,
mean(smoke_06) as smoke_06
FROM fe3
WHERE education_06 ne 999 and index_86 ne 999 and index_91 ne 999 and
index_96 ne 999 and index_01 ne 999 and index_06 ne 999
and live_alone_86 ne 999 and live_alone_91 ne 999 and live_alone_96 ne
999 and live_alone_01 ne 999
and live_alone_06 ne 999 and moved_in_lastyear_86 ne 999 and
moved_in_lastyear_91 ne 999
and moved_in_lastyear_96 ne 999 and moved_in_lastyear_01 ne 999 and
moved_in_lastyear_06 ne 999
and partner_81 ne 999 and partner_86 ne 999 and partner_91 ne 999
and partner_96 ne 999 and partner_01 ne 999 and partner_06 ne 999
and health_96 ne 999 and health_01 ne 999 and health_06 ne 999 and
smoke_06 ne 999;
QUIT;

DATA test;
do low_inc=0;
output;
do low_inc=1;
output;
do low_inc=2;
output;
do low_inc=3;
output;
end; end; end; end;
RUN;

DATA test2;
SET av_data;
do i = 1 to 4;
output;
end;
run;

DATA test3;
MERGE test test2;
DROP i;
RUN;

```

```

ODS GRAPHICS ON;
PROC PHREG data=fe3 plots(overlay)=(survival);
WEIGHT=biaswgt;
CLASS low_inc ;
MODEL months*dead(0) = low_inc age_5yrs sex_81
education_06 index_: live_alone_: moved_in_lastyear_: partner_:
health_: smoke_06;
baseline covariates=test3 out=base survival=_all_ /rowid=low_inc;
RUN;
ODS GRAPHICS OFF;

PROC PRINT DATA=base (obs=10);
RUN;

DATA inc0;
SET base;
where low_inc=0;
surv0 = survival;
upper0= uppersurvival;
lower0 = lowersurvival;
RUN;

DATA inc1;
SET base;
where low_inc=1;
surv1 = survival;
upper1= uppersurvival;
lower1 = lowersurvival;
RUN;

DATA inc2;
SET base;
where low_inc=2;
surv2 = survival;
upper2= uppersurvival;
lower2 = lowersurvival;
RUN;

DATA inc3;
SET base;
where low_inc=3;
surv3 = survival;
upper3= uppersurvival;
lower3 = lowersurvival;
RUN;

PROC SORT DATA=inc0; BY months; RUN;
PROC SORT DATA=inc1; BY months; RUN;
PROC SORT DATA=inc2; BY months; RUN;
PROC SORT DATA=inc3; BY months; RUN;

data plotdata;
MERGE inc0 inc1 inc2 inc3;
BY months;
RUN;

PROC SGPLOT data=plotdata;
xaxis type=time;
title 'Survival Probability for Number of Times in Low Household
Income Quintile';

```

```

step x=months y=surv0 /legendlabel='low_inc 0'
lineattrs=(color='#1b9e77' thickness=2);
step x=months y=surv1 /legendlabel='low_inc 1'
lineattrs=(color='#d95f02' thickness=2);
step x=months y=surv2 /legendlabel='low_inc 2'
lineattrs=(color='#7570b3' thickness=2);
step x=months y=surv3 /legendlabel='low_inc 3'
lineattrs=(color='#e7298a' thickness=2);
xaxis label='months';
yaxis label='survival probabilities' min=0.92;
band x=months lower=lower0 upper=upper0 /fillattrs=(color='#66c2a5')
type=step transparency=0.7;
band x=months lower=lower1 upper=upper1 /fillattrs=(color='#fc8d62')
type=step transparency=0.7;
band x=months lower=lower2 upper=upper2 /fillattrs=(color='#8da0cb')
type=step transparency=0.7;
band x=months lower=lower3 upper=upper3 /fillattrs=(color='#e78ac3')
type=step transparency=0.7;
run;

* SIBLINGS;
DATA sibfel;
SET sib.sibfull;
RUN;

* sort data set by family number and age in descending order to
construct birth order;
PROC SORT DATA=sibfel; BY family_nbr_81 descending age_81; RUN;

* construct birth order;
DATA sibfe2;
SET sibfel;
bthorder + 1;
by family_nbr_81;
IF first.family_nbr_81 THEN bthorder=1;
RUN;

* creating censored variables to run cox regression;
DATA sibfe3 (DROP=Health_Problems_96);
SET sibfe2;
months = AgeD_mths_06 - AgeC_mths_06;
IF Link_06=0 then months = 61;
RENAME Link_06 = dead ;
IF bthorder in (4,5,6) THEN bthorder=3;
RUN;

PROC FREQ data=sibfe3;
tables low_inc sex age_5yrs bthorder education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_: smoke_06;
RUN;

DATA sibtest1;
SET sibfe3;
sex_81 = input(sex, f3.0);
RUN;

DATA sibfe4;
SET sibtest1;
KEEP dead months ID biaswgt family_nbr_81 low_inc bthorder sex_81
age_5yrs education_06

```

```

index_: live_alone_: moved_in_lastyear_: partner_: health_: smoke_06;
RUN;

PROC FREQ DATA=sibfe4;
TABLES dead months low_: bthorder sex_81 age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_: smoke_06;
RUN;

* mean for all variables;
PROC SQL;
CREATE TABLE sibav_data AS SELECT
mean(age_5yrs) as age_5yrs,
mean(sex_81) as sex_81,
mean(bthorder) as bthorder,
mean(education_06) as education_06,
mean(index_86) as index_86,
mean(index_91) as index_91,
mean(index_96) as index_96,
mean(index_01) as index_01,
mean(index_06) as index_06,
mean(live_alone_86) as live_alone_86,
mean(live_alone_91) as live_alone_91,
mean(live_alone_96) as live_alone_96,
mean(live_alone_01) as live_alone_01,
mean(live_alone_06) as live_alone_06,
mean(moved_in_lastyear_86) as moved_in_lastyear_86,
mean(moved_in_lastyear_91) as moved_in_lastyear_91,
mean(moved_in_lastyear_96) as moved_in_lastyear_96,
mean(moved_in_lastyear_01) as moved_in_lastyear_01,
mean(moved_in_lastyear_06) as moved_in_lastyear_06,
mean(partner_81) as partner_81,
mean(partner_86) as partner_86,
mean(partner_91) as partner_91,
mean(partner_96) as partner_96,
mean(partner_01) as partner_01,
mean(partner_06) as partner_06,
mean(health_96) as health_96,
mean(health_01) as health_01,
mean(health_06) as health_06,
mean(smoke_06) as smoke_06
FROM sibtest1
WHERE education_06 ne 999 and index_86 ne 999 and index_91 ne 999 and
index_96 ne 999 and index_01 ne 999 and index_06 ne 999
and live_alone_86 ne 999 and live_alone_91 ne 999 and live_alone_96 ne
999 and live_alone_01 ne 999
and live_alone_06 ne 999 and moved_in_lastyear_86 ne 999 and
moved_in_lastyear_91 ne 999
and moved_in_lastyear_96 ne 999 and moved_in_lastyear_01 ne 999 and
moved_in_lastyear_06 ne 999
and partner_81 ne 999 and partner_86 ne 999 and partner_91 ne 999
and partner_96 ne 999 and partner_01 ne 999 and partner_06 ne 999
and health_96 ne 999 and health_01 ne 999 and health_06 ne 999 and
smoke_06 ne 999;
QUIT;

DATA sibtest;
do low_inc=0;
output;
do low_inc=1;
output;
do low_inc=2;

```

```

output;
do low_inc=3;
output;
end; end; end; end;
RUN;

DATA sibtest2;
SET av_data;
do i = 1 to 4;
output;
end;
run;

DATA sibtest3;
MERGE sibtest sibtest2;
DROP i;
RUN;

* adjusting for everything;
* WITH STRATA;
ODS GRAPHICS ON;
PROC PHREG data=sibfe4 plots(overlay)=(survival) NOSUMMARY;
WEIGHT=biaswgt;
CLASS low_inc /ref=FIRST;
MODEL months*dead(0) = low_inc sex_81 age_5yrs education_06
index_: live_alone_: moved_in_lastyear_: partner_: health_: smoke_06
/TIES=EFRON;
STRATA family_nbr_81;
hazardratio low_inc /diff=ref;
baseline covariates=sibtest3 out=base1 survival=_all_ /rowid=low_inc;
RUN;
ODS GRAPHICS OFF;

* WITHOUT STRATA;
ODS GRAPHICS ON;
PROC PHREG data=sibfe4 plots(overlay)=(survival)
COVSANDWICH(AGGREGATE);
WEIGHT=biaswgt;
CLASS low_inc /ref=FIRST;
MODEL months*dead(0) = low_inc sex_81 age_5yrs
education_06 index_: live_alone_: moved_in_lastyear_: partner_:
health_: smoke_06 /TIES=EFRON;
ID family_nbr_81;
hazardratio low_inc /diff=ref;
baseline covariates=test3 out=base survival=_all_ /rowid=low_inc;
RUN;
ODS GRAPHICS OFF;

DATA inc0;
SET base;
where low_inc=0;
surv0 = survival;
upper0= uppersurvival;
lower0 = lowersurvival;
RUN;

DATA inc1;
SET base;
where low_inc=1;
surv1 = survival;

```

```

upper1= uppersurvival;
lower1 = lowersurvival;
RUN;

```

```

DATA inc2;
SET base;
where low_inc=2;
surv2 = survival;
upper2= uppersurvival;
lower2 = lowersurvival;
RUN;

```

```

DATA inc3;
SET base;
where low_inc=3;
surv3 = survival;
upper3= uppersurvival;
lower3 = lowersurvival;
RUN;

```

```

PROC SORT DATA=inc0; BY months; RUN;
PROC SORT DATA=inc1; BY months; RUN;
PROC SORT DATA=inc2; BY months; RUN;
PROC SORT DATA=inc3; BY months; RUN;

```

```

DATA plotdata1;
MERGE inc0 inc1 inc2 inc3;
BY months;
RUN;

```

```

PROC SGPLOT data=plotdata1;
xaxis type=time;
title 'Survival Probabilities for Number of Times in Low Household
Income Quintile';
step x=months y=surv0 /legendlabel='low inc0'
lineattrs=(color='#1b9e77' thickness=2);
step x=months y=surv1 /legendlabel='low inc1'
lineattrs=(color='#d95f02' thickness=2);
step x=months y=surv2 /legendlabel='low inc2'
lineattrs=(color='#7570b3' thickness=2);
step x=months y=surv3 /legendlabel='low inc3'
lineattrs=(color='#e7298a' thickness=2);
xaxis label='months';
yaxis label='survival probabilities' min=0.95;
band x=months lower=lower0 upper=upper0 /fillattrs=(color='#66c2a5')
type=step transparency=0.7;
band x=months lower=lower1 upper=upper1 /fillattrs=(color='#fc8d62')
type=step transparency=0.7;
band x=months lower=lower2 upper=upper2 /fillattrs=(color='#8da0cb')
type=step transparency=0.7;
band x=months lower=lower3 upper=upper3 /fillattrs=(color='#e78ac3')
type=step transparency=0.7;
run;

```

7.6 Appendix F

Diagnostics in R

```
# full population
full.df = read.csv(file.choose(), header=T)

# fitting cox regression and looking at residual plots
library(survival)
full.df$months = as.numeric(full.df$months)
full.df$biaswgt = as.integer(full.df$biaswgt)+0.0001
str(full.df)
attach(full.df)
s = Surv(months, event=dead)
c = coxph(s~factor(low_inc)+factor(Sex)+factor(age_5yrs)+
factor(education_06)+
factor(index_86)+factor(index_91)+factor(index_96)+factor(index_01)+fa
ctor(index_06)+
factor(live_alone_86)+factor(live_alone_91)+factor(live_alone_96)+
factor(live_alone_01)+factor(live_alone_06)+factor(partner_81)+factor(
partner_86)+
factor(partner_91)+factor(partner_96)+factor(partner_01)+factor(partne
r_06)+
factor(moved_in_lastyear_86)+factor(moved_in_lastyear_91)+factor(moved
_in_lastyear_96)+
factor(moved_in_lastyear_96)+factor(moved_in_lastyear_01)+factor(moved
_in_lastyear_06)+
factor(health_96)+factor(health_01)+factor(health_06)+factor(smoke_06)
,data=full.df, ties="efron", weights=biaswgt)
detach(full.df)
summary(c)
resid = cox.zph(c)
resid
par(mfrow=c(3,1))
plot(resid[1:3], col="red")

cox.fit = survfit(c)
plot(cox.fit, ylim=c(0.8,1))

# deviance
dev = residuals(c, type="deviance")
plot(c$linear.predictor, dev, col="dodgerblue", xlab="Linear
Predictor", ylab="Deviance Residuals")
lines(lowess(c$linear.predictor, dev), col="red", lwd=2)

# schoenfeld
sch = residuals(c, type="schoenfeld")
plot(c$linear.predictor, sch, col="blue")

# siblings
sib.df = read.csv(file.choose(), header=T)

# fitting cox regression and looking at residual plots
library(survival)
sib.df$months = as.numeric(sib.df$months)
str(sib.df)
attach(sib.df)
s = Surv(months, event=dead)
```

```

sib =
coxph(s~factor(low_inc)+factor(bthorder)+factor(age_5yrs)+factor(Sex)+
factor(education_06)+
factor(index_86)+factor(index_91)+factor(index_96)+factor(index_01)+fa
ctor(index_06)+
factor(live_alone_86)+factor(live_alone_91)+factor(live_alone_96)+
factor(live_alone_01)+factor(live_alone_06)+factor(partner_81)+factor(
partner_86)+
factor(partner_91)+factor(partner_96)+factor(partner_01)+factor(partne
r_06)+factor(moved_in_lastyear_86)+factor(moved_in_lastyear_91)+factor
(moved_in_lastyear_96)+factor(moved_in_lastyear_96)+factor(moved_in_la
styear_01)+factor(moved_in_lastyear_06)+factor(health_96)+factor(healt
h_01)+factor(health_06)+factor(smoke_06)+cluster(family_nbr_81),
data=sib.df, ties="efron", weights=biaswgt)
summary(sib)
resid = cox.zph(sib)
resid
par(mfrow=c(3,1))
plot(resid[1:3], col="red")

# deviance
dev = residuals(sib, type="deviance")
plot(sib$linear.predictor, dev, col="dodgerblue", xlab="Linear
Predictor", ylab="Deviance Residuals")
lines(lowess(sib$linear.predictor, dev), col="red", lwd=2)

# schoenfeld
sch = residuals(sib, type="schoenfeld")
plot(sib.df$months, sch, col="blue")
lines(lowess(sibc$linear.predictor, sch), col="red", lwd=2)

```