

<http://researchspace.auckland.ac.nz>

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.

<http://researchspace.auckland.ac.nz/feedback>

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

**Studies of cardiovascular disease
risk estimation:
how, and whether, to account for the
effect of drug treatment?**

Simon James Thornley

This thesis is submitted in fulfillment of the requirements of the degree of
Doctor of Philosophy in Medicine, The University of Auckland, 2014.

Abstract

Cohort studies of individuals at risk of cardiovascular disease (CVD) aim to assess subjects' risk to assist clinical decisions concerning whether patients should be treated with preventive drugs. Since treatment with these drugs is now widespread, it is uncertain how to best account for treatment when modelling CVD risk. Statins and blood pressure lowering agents are considered the principal preventive agents, since meta-analyses generally show between a 0 to 30% reduction in CVD events in the treated group, compared to placebo. The PREDICT cohort of subjects undergoing risk assessment in New Zealand primary care was studied.

The thesis consists of five analyses of this population presented in separate chapters, with another addressing statin trials. The first explores the strength of association between initiating drug use and incident CVD. In the PREDICT cohort, use of preventive drugs was common, with 21% of the total taking statins and 32% taking at least one anti-hypertensive drug at assessment. In a cohort of 56053 untreated subjects, after adjustment for commonly measured risk factors, using a Cox model, those who started both drugs were 50% more likely to have a CVD event than those who remained untreated (95% confidence interval (CI): 3% to 117% increase).

Since treated and untreated people have different risk factor profiles, the second chapter uses propensity score methods to reassess the association between statin use and CVD. The findings were, however, generally concordant with those of the first chapter.

Due to uncertainty about the causal relationship between CVD risk factors, including drug use, the third chapter describes learning Bayesian networks to explore the causal relationships between these factors. The results showed likely causal influence between age and diabetes and baseline drug use; but no relationship between drug use, cholesterol ratio, systolic blood pressure and

CVD. In the fourth chapter, the addition of drug use as a covariate in a Cox model did not improve the classification of the model, using varying cutpoints of risk to assign treatment, over a model which included standard CVD risk factors.

The fifth chapter examines the presence of publication bias in meta-analyses of statin effects, since the preliminary chapters showed drug use was not strongly associated with CVD. In all three highly cited meta-analyses, the number of reported positive trials exceeded the expected, suggesting bias.

The final chapter addresses the magnitude of the association between a novel risk factor, serum urate, and CVD. In this analysis, serum urate was convincingly associated with CVD events. A two standard deviation difference (0.45 vs 0.27 mmol/L) was associated with an adjusted hazard ratio of 1.56 (95% CI: 1.32 to 1.84), using Cox regression analysis.

Any possible beneficial effects of blood pressure or lipid lowering drugs are likely to be more than compensated for by unmeasured adverse prognostic factors. This suggests that omitting drug use information is unlikely to bias models used for disease prediction. Some of the discrepancy between observational and trial evidence of drug effects may be attributed to publication bias.

Contents

1 Literature review	1
1.1 Introduction	1
1.1.1 A brief history of cardiovascular disease	1
1.2 Study of the risk factors for CVD	2
1.3 Framingham Heart Study	2
1.4 PREDICT	5
1.4.1 Baseline information	5
1.4.2 End points	6
1.4.3 Progress with PREDICT	6
1.5 New Zealand guidelines for managing CVD risk	7
1.5.1 Critique of the New Zealand Guidelines	7
1.6 QRISK	8
1.7 The problem: the distorted natural history of cardiovascular disease	10
1.8 What factors influence the accuracy of a risk prediction model?	13
1.8.1 Epidemiological use of regression models for prediction	13
1.8.2 Measuring the validity of a model: discrimination and calibration	15
1.8.3 Critique of predictive modelling	19
1.8.4 How useful is a predictive model?	23

1.9	How effective is drug treatment for the primary prevention of CVD?	23
1.9.1	Sources of bias in meta-analyses of randomised studies	29
1.10	Summary with hypothesis, aims and contribution	32
2	Methods	36
2.1	The study sample: the PREDICT cohort	37
2.1.1	Other covariates	38
2.1.2	Outcomes	38
2.1.3	Overview of the cohort	39
3	Analysis 1: The magnitude of the association between drug initiation and CVD	41
3.1	Introduction	41
3.1.1	Drug use at baseline risk assessment	41
3.1.2	Drug use during follow-up	42
3.2	An overview of the analyses	45
3.2.1	Design	45
3.2.2	Statistical Analysis	45
3.3	Results	46
3.4	Discussion	50
4	Analysis 2: Propensity scoring to estimate the association between statin initiation and CVD risk	53
4.1	Background	53
4.2	Study design	55
4.2.1	Drug exposure	55
4.2.2	Statistical analysis	56
4.3	Results	58
4.4	Discussion	62

5	Analysis 3: Using learning Bayesian networks	66
5.1	Background	66
5.2	Study Design	68
5.2.1	Sample	68
5.2.2	Statistical analysis	69
5.3	Results	71
5.4	Discussion	73
6	Analysis 4: The clinical utility of using drug treatment variables to predict CVD	79
6.1	Background	79
6.2	Study design	80
6.2.1	Sample	80
6.2.2	Statistical analysis	81
6.3	Results	84
6.4	Discussion	90
6.4.1	Conclusion	93
7	Analysis 5: Publication bias in studies of statin use to prevent CVD	94
7.1	Background	94
7.2	Method	95
7.2.1	Identification of studies	95
7.2.2	Methods to examine publication bias	96
7.3	Results	97
7.4	Conclusion	101
8	Analysis 6: Serum urate and the risk of CVD: the use of DAGs to estimate causal effects	105
8.1	Background	105
8.2	Study design	107
8.2.1	Statistical analysis	107

8.3	Results	110
8.4	Discussion	115
9	Conclusion	121
9.1	Statement of principal findings	121
9.2	Strengths and weaknesses of the thesis	123
9.3	Strengths and weaknesses of the thesis in relation to other studies	125
9.4	Implications of the thesis	129
9.5	Unanswered questions and future research	132
A	Sensitivity analyses of the effect of serum urate on CVD incidence	133

List of Figures

1	Co-authorship form.	xv
2	Co-authorship form.	xvi
3	Co-authorship form.	xvii
4	Co-authorship form.	xviii
2.1	Recruitment periods of the cohorts used in the different chapters of this thesis.	39
3.1	Scaled rectangle diagram [75] depicting preventive drug use at risk assessment and diabetes status in the PREDICT cohort who were recruited in the years 2006 to 2009.	43
3.2	Scaled rectangle diagram [75] depicting statin use trajectory in a subset of the PREDICT cohort who were recruited in the years 2006 and 2007 and had at least two years of follow-up ($n=26408$).	44
5.1	A collider variable is influenced by more than one other variable: here the exposure and outcome. The collider is also the <i>descendant</i> of both the exposure and outcome variable.	68
5.2	DAG, derived from the <i>grow-shrink</i> algorithm. The grey box indicates the outcome variable. CVD: Cardiovascular disease. HDL: high-density lipoprotein cholesterol concentration. TC: total cholesterol concentration. BP: blood pressure.	73

6.1	<i>Net-benefit</i> associated with the use of different CVD prediction models for varying threshold probabilities (p_t) for 2-year CVD risk.	88
6.2	Scaled rectangle diagram [75] depicting model predicted high risk patients ($\geq 3\%$ 2 year risk) and comparing them with their disease status at 2 years follow-up. One model had standard risk factors, and the other had these variables as well as preventive drug use data at enrolment.	89
7.1	<i>P</i> -value evaluates the whether the difference between the observed and expected number of statistically significant trials of summary meta meta-analyses is likely to be due to chance, at different levels of significance (alpha). (Solid black line: observed number of significant studies; black dotted line: expected number of significant studies; grey dotted and dashed line: <i>P</i> -value derived from chi-square test for difference between observed and expected numbers of significant trials).	104
8.1	Directed acyclic graph, showing causal relationships between the exposure serum urate, and cardiovascular disease (CVD) outcomes, with observed (black font) and unobserved (grey font) variables. From the PREDICT study. BP: blood pressure.	108
8.2	Density plot comparing the distributions of observed (dashed line) and imputed urate values (solid line).	113
8.3	Cox-modelled, survival estimates (with 95% confidence bands) by time for serum urate at the 16th centile (0.27 mmol/L), compared to the 84th centile (0.45 mmol/L) for a non-smoking male in the 'Other' ethnic group, with HDL cholesterol = 1.34 mmol/L, HbA1c = 6%.	114
A.1	134

A.2	135
A.3	136
A.4	137

List of Tables

3.1	Proportion of PREDICT cohort, using statins at baseline and during follow-up.	43
3.2	Total PREDICT cohort, by statin and anti-hypertensive treatment status in the six months after enrolment. Data are complete unless indicated otherwise.	47
3.3	Crude and adjusted associations from a Cox model of CVD incidence including categories of anti-hypertensive and statin use (n=43366; 412 CVD events).	49
4.1	PREDICT cohort, by statin treatment in the six months after enrolment, with propensity matched sample and comparison of standardised mean differences in variables. Data are complete unless indicated otherwise.	59
4.2	Crude and multivariate associations between clinical and demographic factors and statin use in the six months after assessment, derived from a logistic model (n=56051).	60
4.3	Crude and adjusted hazard ratios for model of association between statin use and CVD, with and without propensity score matching.	61
5.1	Sample characteristics by cardiovascular disease status (learning Bayesian network analysis): numbers (% of column sample unless otherwise stated).	72

5.2	Lists of causal links and estimated beta-coefficients from linear or logistic regression, adjusted for variables, as indicated in the DAG. Also odds ratios for logistic regression models.	74
6.1	Sample characteristics by CVD status: count (% of column sample unless otherwise stated).	85
6.2	Crude and adjusted hazard ratios derived from Cox models of CVD events, which included commonly measured risk factors as covariates with and without drug treatment variables.	86
7.1	Summary of recent meta-analyses of randomised-trials which assess the effect of statin use on risk of CVD event, conducted in people at raised risk of a CVD event.	98
7.2	Evidence for publication bias of trials reported as significant at the 5% significance level.	100
8.1	Urate and CVD analysis: baseline characteristics, by sex.	111
8.2	Observed mean serum urate, by sex, ethnic group, diabetes and smoking status.	112
8.3	Measures of association with CVD from a Cox proportional hazard model. $n = 78707$	113
9.1	Summary of the analyses.	123

Acknowledgements

I thank my supervisors Roger Marshall and Rod Jackson for their patient help and guidance. I gratefully acknowledge the R community, who have provided the software which has been used to conduct the analyses presented here. The PREDICT community also deserves mention for allowing me to use its data, and for the support with cleaning and merging data from a number of different sources. I am, of course, indebted to the love and support from my family: Emily, Patrick and Lauren.

List of acronyms

ACE	angiotensin converting enzyme
AHT	anti-hypertensive drugs
AIC	Akaike's information criterion
BIC	Bayesian information criterion
BP	blood pressure
CHD	coronary heart disease
CI	confidence interval
CRAN	The comprehensive R archive network
CVD	cardiovascular disease
DAG	directed acyclic graph
GP	general practitioner
HbA1c	haemoglobin A1c
HDL	high density lipoprotein
HR	hazard ratio
IQR	interquartile range
LDL	low density lipoprotein

N/A	not applicable
NB	net-benefit
NHI	national health index
NRI	net reclassification improvement
NZ	New Zealand
OR	odds ratio
RCT	randomised controlled trial
RR	risk ratio
SD	standard deviation
TC	total cholesterol
U. S.	United States of America

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5 is drawn from the publication:

Thornley S, Marshall RJ, Wells S, Jackson R. Using Directed Acyclic Graphs for Investigating Causal Paths for Cardiovascular Disease. *J Biomet Biostat.* 2013; 4.

Nature of contribution by PhD candidate

Extent of contribution by PhD candidate (%)

CO-AUTHORS

Name	Nature of Contribution
Roger Marshall	Study design, editing draft, and responding to reviewers' comments.
Susan Wells	Study design and editing draft
Rod Jackson	Study design and editing draft

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.


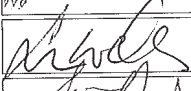
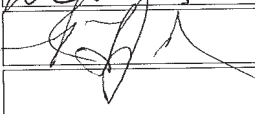
Name	Signature	Date
Roger Marshall		24/03/2014
Susan Wells		Click here <i>25/3/2014</i>
Rod Jackson		25/03/2014
		Click here
		Click here
		Click here

Figure 1: Co-authorship form.

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 7 is drawn from the publication:

Thornley S, Marshall RJ, Barzi F. Is publication bias present in the reported beneficial effects of statins for primary prevention of cardiovascular disease? A simulation study. In press. *BMJ Open*; 2014.

Nature of contribution by PhD candidate

Designing study, carrying out analyses, drafting manuscript, responding to reviewers' comments.

Extent of contribution by PhD candidate (%)

80

CO-AUTHORS

Name	Nature of Contribution
Roger Marshall	Study design, editing draft, and responding to reviewers' comments
Federica Barzi	Study design and editing draft

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Roger Marshall		24/03/2014
Federica Barzi		24/03/2014
		Click here
		Click here
		Click here
		Click here

Figure 2: Co-authorship form.

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 8 is drawn from the publication:

Thornley S, Marshall RJ, Jackson R, et al. Is serum urate causally associated with incident cardiovascular disease? *Rheumatology*. 2013; 52:135-42.

Nature of contribution by PhD candidate	Designing study, carrying out analyses, and drafting manuscript.
Extent of contribution by PhD candidate (%)	80


CO-AUTHORS

Name	Nature of Contribution
Roger Marshall	Study design, editing draft and responding to reviewers' comments
Rod Jackson	Study design and editing draft
Dudley Gentles	Study design and editing draft
Nicola Dalbeth	Study design and editing draft
Sue Crengle, Andrew Kerr	Study design and editing draft

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Roger Marshall		24/03/2014
Rod Jackson		26/03/2014
Dudley Gentles		25/03/2014
Nicola Dalbeth		
Sue Crengle		31/03/2014
Andrew Kerr		31/03/2014

Last updated: 25 March 2013

Figure 3: Co-authorship form.

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 8 is drawn from the publication:

Thornley S, Marshall RJ, Jackson R, et al. Is serum urate causally associated with incident cardiovascular disease? *Rheumatology*. 2013; 52:135-42.

Nature of contribution by PhD candidate	Designing study, carrying out analyses, and drafting manuscript.
Extent of contribution by PhD candidate (%)	80

CO-AUTHORS

Name	Nature of Contribution
Nicola Dalbeth	Study design and editing draft
Susan Wells	Study design and editing draft

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.


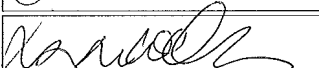
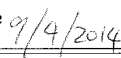
Name	Signature	Date
Nicola Dalbeth		4/04/2014
Sue Wells		Click here 
		Click here

Figure 4: Co-authorship form.

Chapter 1

Literature review

1.1 Introduction

1.1.1 A brief history of cardiovascular disease

The first hint of coronary artery disease was recorded in 1768, when the famous physician William Heberden described *angina pectoris*, the sensation of being strangled in the chest [78]. At the time, he did not make the connection between these symptoms and the heart.

Historical evidence suggests that although the condition was first described at this time, it was still extremely rare. Heberden himself records only encountering 20 cases of angina during his twenty or so years of clinical practice. Records indicate that few cases were described from the late 18th century until 1912, when Herrick reported a case-series of six with coronary thrombosis, a blood clot in the arteries supplying the heart muscle. Even in the early 20th century, death from coronary disease was a rare event. Michaels reports evidence that deaths from hardening of the arteries (arteriosclerosis) were 200 times more common in 1962 as compared to 1901-10 [78].

A common objection to evidence of a rapid rise in the number of cases of coronary artery disease, was that life expectancy was also increasing in the

early 20th century, and that later populations were much older, and thus more likely to develop disease. Although the number of people aged 50 years or older did increase in the UK over this period, the magnitude was much smaller (threefold; 4,790,000 in 1901 to 14,158,000 in the early 60s) than that observed for increasing rates of disease [78].

By the early 1960s, however, the epidemic of cardiovascular disease (CVD: coronary artery disease, stroke and peripheral vascular disease) was established in Western industrialized nations, and doctors and scientists were seeking an explanation for this rise in incidence.

1.2 Study of the risk factors for CVD

To help investigate causes and predictors of disease, a cohort was followed in the U. S. town of Framingham, Massachusetts, from the early 1950s [59]. From the findings [58], the major risk factors for heart disease were derived and prevention strategies started. Major risk factors included: cigarette smoking, raised systolic blood pressure, lipids and diabetes; along with those which are non-modifiable: sex and age [59]. The Framingham study investigators also developed a series of CVD risk prediction scores. Since this time, a number of other cohort studies of CVD risk from other parts of the world have developed risk prediction equations, including QRISK in the United Kingdom [51], SCORE [21] in continental Europe, and the Reynold's risk score [97] from the U. S.

1.3 Framingham Heart Study

The Framingham study is important since predictive models derived from the study are recommended for use in New Zealand primary care for the assessment and management of patients considered to be at risk of CVD [83].

The Framingham study sample consists of individuals who were exam-

ined between the years of 1968 and 1971, when measurement of high-density lipoprotein was initiated. Participants were aged between 30 and 74 years, and prediction scores were based on the information supplied by 8491 individuals [25], who were followed up for twelve years. CVD risk factors were measured by a variety of means. Blood pressure was measured twice on the left arm, with the patient seated, so that the mean of the two data points formed the recorded value. Total and serum HDL cholesterol were estimated from peripheral blood samples. Smoking status was self-reported, and diabetes status consisted of a fasting glucose ≥ 7 mmol/L, or, if they were prescribed insulin or oral-hypoglycaemic drugs.

Disease outcomes were closely monitored. The Framingham study population was under continuous surveillance for CVD events, by means of medical histories, physical examinations, hospital records, and communication with personal physicians [25]. If new events were suspected, a panel of three physicians examined the evidence for the claim. Events included coronary heart disease (coronary death, infarction, and angina), cerebrovascular events (ischaemic or haemorrhagic stroke, and transient ischaemic attack) and peripheral vascular disease (intermittent claudication), as well as the incidence of heart failure.

In one analysis from Framingham, sex-specific multivariate Cox models were used to estimate the incidence of a first CVD event [25]. Covariates in the models included age, total cholesterol, high density lipoprotein (HDL) cholesterol, systolic blood pressure, anti-hypertensive drug use, current smoking and diabetes status. All continuous predictor variables were transformed, by taking the natural logarithm, to minimise the effect of extreme observations. Diastolic blood pressure, body mass index and serum triglyceride levels were rejected because they were not 'statistically significant'. The study did not state clearly how this criteria was applied, for example was it part of a stepwise procedure, as it is clear that the collinearity between some variables are likely to be high (diastolic and systolic blood pressure, for example).

The ability of the model to discriminate between individuals who did and did not develop disease during follow-up was measured using Harrell's *c*. The calibration of the model, that is the degree of agreement between observed and predicted disease frequencies, was also assessed using a modified Hosmer-Lemeshow χ^2 statistic (9 degrees of freedom).

Final models were stratified by sex, and included age, total cholesterol, high density lipoprotein, systolic blood pressure (treated or untreated), smoking and diabetes status. Harrell's *c* was 0.763 in men and 0.793 in women. The calibration statistics were 13.48 for men and 7.79 for women in the study, which indicated excellent goodness-of-fit [25]. A model which excluded laboratory tests, replacing total and HDL cholesterol for body mass index, resulted in discrimination statistics (Harrell's *c*) slightly lower than the laboratory counterparts, of 0.749 for men, and 0.785 for women. These results indicate that the model had a fair ability to distinguish cases from non-cases (discrimination is discussed in further detail in section 1.8.2 on page 16). Goodness-of-fit statistics were similarly excellent [25].

The Framingham models have some strengths countered by limitations. First, the study was carried out during a period of time when drug treatment to prevent CVD was relatively rare, so that assessments were less likely to be contaminated by the subjects being exposed to treatment during follow-up. In addition, the clinical measurements and outcome measures were standardised, which were subjected to scrutiny by a team of clinical investigators. This contrasts to the PREDICT study, the subject of this thesis, which will be discussed in the next section, in which national lists of hospital treatment and mortality are relied upon for recording of disease outcomes.

The Framingham study was, however, limited by relatively small numbers of individuals, who experienced few events. No information about subjects' ethnic group was presented, so these differences were not investigated. Similarly, socioeconomic status was not reported. Statistical procedures did not account for potential threats to validity of the inferences, such as the use age as

the time scale for the study, as is now generally recommended for such analyses. Despite these limitations, the models derived from the Framingham study have generally been considered the standard by which all others have been assessed.

1.4 PREDICT

This thesis uses data from the New Zealand PREDICT study [96]. This is a prospective cohort, which enrolls participants through general practice visits in which CVD risk assessment is undertaken, as part of routine clinical care. Software provided by the PREDICT group allows family doctors to give their patient an estimate of their 5-year predicted cumulative incidence of CVD, based on a modified Framingham equation [1]. The raw Framingham scores are modified if, for example, an individual identifies with an ethnic group (Māori, Pacific and South Asian) which, from local research, is associated with raised risk of disease [83].

1.4.1 Baseline information

The recorded clinical information used by the general practitioner is stored and linked with national health information, which enables further baseline covariates and outcome data to be determined. The variables available from the risk assessment include: age, gender, smoking status, diagnosed diabetes status, systolic blood pressure, premature family history of CVD, and total: low density lipoprotein cholesterol ratio. Drug use at risk assessment and during follow-up is available from a national database in which information from redeemed prescriptions from community pharmacies are recorded. These data form the baseline information, at study enrolment. The PREDICT study was established in 2002, with ongoing recruitment, but drug use data has only been reliably available since the start of 2006, so that cohorts in this thesis are limited to subgroups that were recruited from this year on.

1.4.2 End points

Outcome information is drawn from national lists of cause-specific mortality and hospital discharge records. Codes from these sources which indicate a first diagnosis of CVD (angina, acute coronary syndrome, ischaemic or haemorrhagic stroke, or peripheral vascular disease) are considered disease outcomes, along with the date of event. Individuals who die during follow-up from causes other than CVD are considered censored.

Compared to studies which scrutinise outcomes individually with a team of investigators, such as Framingham; outcomes in PREDICT are more likely to be subject to measurement error. Since coding is conducted independently of the study, there may be changes in how conditions are coded as standards for coding change with time and between individual assessors.

1.4.3 Progress with PREDICT

Although there have been many publications resulting from the PREDICT study to date, no risk prediction equations of CVD status have been published because of insufficient follow-up time, although this is one of the aims of the research programme. At the time of writing, about 400,000 individuals had been enrolled in the study. Although a national programme recommends CVD screening of patients from the age of 35 years onward, depending on risk group, the PREDICT tool is not universally used in general practice in New Zealand. The PREDICT population is, therefore, considered to be a convenience sample, albeit a large one. Covariates used as predictors are largely complete, however, a range of other variables are available, such as laboratory test results, which are linked from regional databases held by laboratories. These have some missing information due to the nature of clinical care and the regional coverage of these sources.

1.5 New Zealand guidelines for managing CVD risk

National care pathways for general practitioners probably influence those enrolled into the PREDICT study. CVD risk screening guidelines, in NZ, were updated in 2009 [84]. Summarised, this consensus document recommends the screening of people without known risk factors from the age of 45 years in men, and at 55 in women. Māori, Pacific and Indian people are recommended to be screened 10 years earlier, and similarly, those with known risk factors, such as cigarette smoking, gestational diabetes or polycystic ovary syndrome, blood pressure $\geq 160/95$ mmHg, or prior total cholesterol: HDL ratio ≥ 7 , impaired glucose tolerance or impaired fasting glucose, a family history of diabetes in a first degree relative, body mass index $\geq 30\text{kg/m}^2$ or truncal obesity (waist circumference ≥ 100 cm in men or ≥ 90 cm in women), an estimated glomerular filtration rate of $\geq 60\text{mL}/\text{min}/1.73\text{m}^2$, or a family history of premature coronary disease (father or brother < 55 years or mother or sister < 65). Anyone with a diagnosis of diabetes is recommended to be screened annually from the time of diagnosis. At the clinical interaction, 5-year cumulative incidence of CVD is calculated from the clinical data and the historic Framingham cohort study [1], with modifications [84].

Analyses of the observed and predicted risk of disease (calibration), were undertaken with a mean follow-up of 2.11 years [96]. This study revealed that the New Zealand adjusted risk score generally overestimates risk scores, whereas the original Framingham score overestimates risk for Europeans but underestimates risk for Māori, Pacific and Indian people.

1.5.1 Critique of the New Zealand Guidelines

In an editorial, cardiologist Harvey White pointed out some of the deficiencies and contradictions of current guidelines [137]. The predictions used in New Zealand, based on Framingham derived equations, come from a cohort with a restricted age range from those who participated in the U. S. study [1]. People

outside the age range between 35 to 74 years have risk extrapolated from other studies, or have figures extrapolated from risk assessed at the limits of age. People who have had a previous CVD event, or therapeutic procedure, familial dyslipidemias, or diabetes with nephropathy or renal impairment judged to be at high ($\geq 20\%$ 5 year) risk, so that screening is deemed unnecessary. Framingham risk calculations are also adjusted for individuals with adverse factors that were not included in the original study design. Those in high risk ethnic groups, with a family history of CVD, a history of diabetes, with microalbuminuria or poor glycemic control (HbA1c consistently $\geq 8\%$), have their 5-year cumulative incidence prediction raised by an additional 5% over baseline levels. These increases are relatively arbitrarily applied outside of the usual statistical framework for developing risk prediction models.

This critique highlighted the *ad hoc* nature of CVD risk assessment in New Zealand. White's editorial suggests the need for suitable equations derived from local studies.

1.6 QRISK

In the United Kingdom, a cohort study to estimate CVD risk has been conducted. Until recently, the data was captured very differently from PREDICT. It was retrospectively extracted from medical records and linked to events, also extracted from medical records. At present, however, data is captured in a similar manner to the PREDICT study, using a QRISK calculator available to the primary care clinician. The QResearch data source has provided larger numbers of individuals for analysis than PREDICT, for example; but the study faces different problems of widespread missing data, and few covariates available for analysis. One publication [49], for example, describes the cohort and the derived models. The study consists of 563 enrolled 'QResearch' practices, spread throughout England and Wales. In one risk estimation analysis, the study population, over 3 million individuals, was split randomly, by prac-

tice, with two-thirds contributing to a derivation and one-third to a validation sample on which the prediction model was derived and its accuracy examined [49].

A range of variables were available for analysis, including:

- age
- area-based socioeconomic deprivation (Townsend score)
- ethnic group
- clinical conditions (including diabetes, treated hypertension, atrial fibrillation, chronic renal disease, and rheumatological conditions)
- family history of coronary heart disease (in first degree relative aged < 60 years)
- clinical data (systolic blood pressure, body size, total and HDL cholesterol).

Final models were sex-specific and included the following candidate risk factors: smoking status, ethnic group, systolic blood pressure, ratio of serum total serum cholesterol to HDL, body size, area-based socioeconomic status, treated hypertension, rheumatoid arthritis, atrial fibrillation, type-2 diabetes, and chronic renal disease. Much data was missing, for example, less than 30% of participants had their serum cholesterol recorded, so that analyses were carried out using multiple imputation to fill missing values. This means that a distribution of predicted values for the missing data points is calculated, given other baseline covariates and whether a disease event had occurred during follow-up.

Results were summarised in terms of both ten year, and life-time risk. Socioeconomic status, body mass index, systolic blood pressure, smoking, South Asian ethnic group, and the clinical conditions all contributed to the final model. The comparison of lifetime and 10 year risk was informative. Ten year risk was

heavily dependent on age, with the median increasing sharply between 50 and 60 years. Median lifetime risk, calculated from date of entry in the study, in contrast, was less age-dependent, with lower values at both ends of the age spectrum. The receiver-operating-characteristic statistic, was high compared to the Framingham *c* statistic, with estimates for the 10 year calculations at 0.842 among men, and 0.828 among women. These figures indicate good model discriminative ability, so that, on average, the model assigns higher risk status to those who become eventual cases compared to those who do not.

The QRISK study has incorporated a larger number of risk factors into prediction equations, and seems to perform better than the simpler Framingham counterpart when assessed according to indices of calibration and discrimination. However, the quality of the data is probably better in the American study, with data collection standardised, which is not the case in routine clinical practice. The statistical power, however, is much greater in the UK study, with a broader range of risk factors considered. Another detraction from the QRISK estimates is the reliance of the calculations on statistical imputation to account for missing data. Since little information is given about why subjects had missing information, it is difficult to assess the likely nature of the bias due to this feature of the study.

1.7 The problem: the distorted natural history of cardiovascular disease

In 2011, Paul Glasziou, a professor of clinical epidemiology, and other authors, drew attention to the potential problem posed by the use of drugs by subjects in studies which consider time to first CVD event [69]. If the goal of a study is to predict the natural history of disease, the authors argued, then any treatment received by patients will reduce the risk of disease and lead to a distorted natural history, when statistical methods are used to link baseline characteristics with disease events. Moreover, measures of association between risk factors

and disease are likely to be distorted. Glasziou argued that people with high levels of the risk factor, such as systolic blood pressure, are more likely to be treated. When treated subjects are enrolled into a study, they will have a lower blood pressure recorded at enrolment to the study than if they had otherwise been untreated. This treatment is then likely to distort the estimated relationship between the risk factor and disease incidence.

No study is exempt from this problem. Even the first Framingham study, which began in the 1950s and had an analysis published in 1976 [59], had anti-hypertensive drugs in common use during follow-up, with beta-blockers released in 1967 and thiazide diuretics in the late 1950s. Since this time, further blood pressure lowering drugs are commonly used for clinical care, and statins, which lower serum cholesterol, appeared in clinical use in the late 1980s [69]. Statin use, however, was not widespread in New Zealand until the late 1990s, starting first in patients with CVD, and later used in people considered to be at high risk of CVD.

Since the PREDICT study has recruited subjects since the early 2000s, the analysis of these data is also affected by drug treatment. How best to account for these drug effects forms the basis of this thesis. Other aspects of risk prediction, such as the possible link between novel risk factors and CVD, such as serum urate, will also be examined.

Glasziou and colleagues identify that authors have, in the past, worked around the issue of drug treatment in a number of different ways. In the majority of surveyed studies, use of anti-hypertensive agents was identified as a baseline exposure with about half of these including the variables in the final model. In one of the Framingham studies, two separate terms were included for systolic blood pressure, one accounting for the treated and the other for the untreated [25]. The main concerns from the Liew review were the lack of accounting for treatment during follow-up. Some studies included treatment as a baseline variable, such as blood pressure lowering, however, items that were not included were people initiating treatment during follow-up, and

other classes of drugs thought to be beneficial to the prevention of CVD, such as anti-platelet agents (aspirin and clopidogrel), statins, and smoking cessation therapy. These were not considered either at baseline nor follow-up in the studies which were reviewed [69].

Liew and Glasziou's review concluded that the only options were to:

1. favour old studies and discard new ones
2. monitor treatment uptake and use a penalised Cox model to account for study treatment, or
3. study cohorts over short time periods with larger numbers.

Elsewhere, in communication to the *British Medical Journal*, these authors asserted that:

'... risk prediction that ignores treatment effects is dangerous nonsense.' [70]

Other authors, such as Professor John Simes, from the University of Sydney, have taken up Liew and Glasziou's second option in the analysis of the results of randomised controlled trials of the effect of fibrates on the prevention of cardiovascular disease events [106]. In this analysis, the authors assume a beneficial effect of statins used during follow-up, and adjust for this information to modify the estimation of the effect of fibrates, which is the *a priori* subject of the trial.

This view assumes that drugs exert powerful influence on disease risk, which are likely to distort model predictions. On the surface, these arguments seem sensible, but are they?

Liew, Doust and Glasziou appropriately argue that main aim of risk prediction of CVD is to assess the natural history of disease, in the absence of drug treatment. However, an assumption of their overall argument is that drugs convey large beneficial effects on risk of CVD. Their listed options are all based on this assumption. In a later section in this chapter, this assumption will be contested from the published results of meta-analyses of randomised

controlled trials (section 1.9 on page 23). This argument is supported by a later chapter which addresses whether publication bias is present in meta-analyses of trials of statins for the primary prevention of cardiovascular disease (chapter 7 on page 94).

A broader discussion of which factors influence the effect of prediction models will be considered in the next section.

1.8 What factors influence the accuracy of a risk prediction model?

This issue of treatment during follow-up, with the proposed solutions, are based on a number of assumptions that will now be examined. The first is that drugs intended to prevent CVD have a beneficial effect, and that this improvement is large enough to distort the accuracy of prediction models. The second is that it is useful to include follow-up information in a prediction model and that this information will improve model accuracy. In order to examine these notions, the accuracy of prediction models is explored together with the magnitude of effects, associated with baseline risk factors, which are likely to influence disease risk.

1.8.1 Epidemiological use of regression models for prediction

Regression models are used for two main purposes. Most often, they control for confounding when assessing the statistical evidence for causal association between an exposure and a disease outcome.

Causal modelling involves different considerations when choosing covariates to include in the model, than their other main use in studies which aim to predict disease risk. Variables which are causally influenced by the exposure of interest should not be included in the causal model. These variables are called *mediators*.

In addition to the problem of mediators, *barren proxies* [90] are variables which exert no influence on either exposure or disease, but are themselves influenced by a variable or variables which influence disease, or exposure, or both. These variables are again best avoided when estimating causal effects. In this situation, or assessing the strength of association between exposure and disease, directed acyclic graphs either derived from informed scientific knowledge or from computer algorithms may help decide on the structure of the model. This approach is explored further in chapters 5 and 8.

Another use for regression models is to *predict* disease status from a number of risk factors, or signs or symptoms. From a statistical perspective, it is not so important whether a variable is classified as a mediator or confounder. The accuracy with which the model fits the data, penalised for the complexity of the model, is the most accepted criterion by which models are assessed, and are used to select variables.

When developing prediction models, with a range of candidate predictors, various different variable selection procedures are used, such as stepwise backward, forward or best subsets, which minimise an information criterion, such as Mallows's C_p . Other criteria include Akaike's information criterion (AIC) or the Bayesian information criterion (BIC). These measures take the difference between the model fit (represented by the model χ^2 , with the complexity of the model represented by the number of degrees of freedom of the model. This means that for candidate covariates, which include only one degree of freedom, such as gender, the incremental change in model $\chi^2 > 2$ degrees of freedom, is equal to a P -value of less than 0.157 with the outcome, required to justify inclusion in the model [110]. For BIC, the P -values for selection are much lower, with the increase in model $\chi^2 > \log(n)$, where n is the number of events in the Cox regression, for example. So, for a scenario in which there is 100 outcomes, the equivalent P -value selection criterion is less than 0.032. Thus, the strength of the conditional association between the candidate covariate (or predictor) and the outcome, forms the statistical evidence on which the variable selection

decision is made.

This information has some relevance to the study of causal structures using directed acyclic graphs (DAGs), in that those variables which are directly related to the outcome will continue to demonstrate an association with the disease, even when applied to different populations, in different circumstances. DAGs represent an alternative method to traditional regression to examine possible causality [90] and will be explored further in chapters 5 and 8. In the next section, more traditional methods of assessing model performance will be discussed.

1.8.2 Measuring the validity of a model: discrimination and calibration

Competing models may be ranked by their performance, particularly when estimating their ability to predict. In general, clinical prediction models are poor predictors of disease status when applied to individuals, and they show relatively poor discriminatory performance. That is, that if a cut-point is used to define a threshold for treatment, generally, many who are deemed to be high risk, from the use of a particular cut-point, will not develop disease [10].

Two measures of model prediction are used, and give complementary information about model performance. *Discrimination* refers to the ability of a model to distinguish those who go on to experience an event during follow-up from those who do not, whilst *calibration* refers to the long run predicted frequency of disease events from the model, compared to the observed frequency in the study population.

Calibration

Calibration of survival models, is calculated at arbitrarily defined time points (for example, five years) during follow-up in which predicted and observed survival is compared. To account for censoring, the cohort is divided into

time intervals, which generally consist of about 50 subjects per interval. Mean model predicted survival are compared with the observed, which is derived from the Kaplan-Meier estimate [44]. The observed and predicted risk categories are plotted and agreement is then measured by a regression of the observed on the predicted survival (the fit may be summarised in the *calibration slope* statistic).

Discrimination

This measure of validity is often measured in survival data by using Harrell's *c* statistic, with the 'c' short for concordance. It is defined, in survival analysis, by considering the proportion of usable subject pairs, in which at least one of the individuals has experienced the event [44]. If the predicted survival is higher for the patient that lived longer, then the predictions concur with outcomes. These pairs score 1 and, conversely, 0 is assigned to those pairs that have predicted and observed measures of survival that disagree. Concordance may not be assessed in pairs in which one subject has developed the outcome, but the other has been censored before the outcome occurred in the comparator.

The interpretation of the *c*-statistic corresponds to the probability that a randomly selected case will have a higher ranking risk score (linear predictor) than a non-case. The *c* statistic has similar properties to the area under an receiver-operating-characteristic curve, often used to assess the performance of logistic models. A value of 0.5 indicates no discrimination, and 1.0 indicates perfect separation of patients with differing survival outcomes. In a review of a large series of cardiovascular risk prediction models, the area-under-the-curve statistics ranged between 0.68 (95% CI: 0.63 to 0.74), to 0.86 (95% CI: 0.86 to 0.96), with most models in the 0.7 to 0.8 range [107].

Critique of discrimination and other indices

Some authors have criticised the use of discrimination as an artificial construct, not of clinical relevance [22]. Clinicians are primarily interested in the proba-

bility of disease given that an individual is high risk, rather than the inverse (probability of being high risk given disease status) [22], which is estimated by the c statistic.

It is also argued that since concordance is a rank statistic, so it is insensitive to changes in predicted probability, which may be large and clinically significant, yet the c statistic associated with the model will not change if the rank order of predictions are maintained. As an example, if an established model with a two standard deviation difference in linear predictor yields an odds ratio of 16, such as one derived from the Framingham cohort [1], a further independent predictor, added to the original model, with an odds ratio of 2 (such as cigarette smoking), will result in minimal change in the c -statistic (both models will have a c -statistic of 0.84) [22]. In this scenario, an independent factor with an odds ratio of 9 (for a two standard deviation comparison) will need to be added to a model to alter the c -statistic from 0.84 to 0.90. Nevertheless, the c -statistic is independent of the incidence of disease, thus allowing for comparison between different models, derived from differing populations. The statistic is also frequently reported in CVD risk prediction studies, and may be used to assess the clinical utility of a prediction model, derived from decision analysis (for further discussion see section 6 on page 79).

The validation of the model is generally considered over optimistic if carried out on the same dataset, used to derive the model, due to potential overfitting. Overfitting refers to models which fit random, rather than systematic variation in the data that may be difficult to detect. For this reason, fitting an established model to a new population is generally considered the best means of testing the validity of a model and detecting overfitting [110].

Due to the impracticalities of conducting two separate studies, internal validation is more commonly carried out. Frequently used methods include *data-splitting*, *cross-validation*, and *bootstrapping*. Splitting involves dividing the data into a training portion to derive the model, then testing the discrimination of the model on the remainder. Cross-validation involves partitioning the total

available data into subsets, using one subset to estimate the predictive model, and another to test its performance. The testing is carried out repeatedly, and the results averaged. While cross-validation has been shown superior to data-splitting, both methods are generally considered inferior to the third method: bootstrapping [44].

The bootstrapping technique may be applied to both measures of both *discrimination* and *calibration*. Harrell [46] describes the bootstrapping process for the c statistic as follows:

1. Derive the model based on all n subjects. Let c_{app} denote the *apparent* c statistic of the model. The ‘stopping rule’ for adding further predictors, devised by Harrell [44], is to only test a pre-specified number of predictors, and delete them if the total change in chi-square between the model with and without the variable $\chi^2 < 2 \times \text{d.f.}$. Harrell reasons that if the increment in χ^2 is that small, the improvement to the model accuracy is likely to not be improved, because Akaike’s Information Criterion is a commonly used to adjudicate the extra information derived from competing models, derived from the χ^2 by the formula $\text{AIC} = \chi^2 + 2 \times \text{d.f.}$
2. Sample the original population with replacement to derive a second sample of size n .
3. Fit a full model on the sampled population and derive a new c statistic, called c_{boot} . Use a stopping rule as described in the first step.
4. This second model is then evaluated on the original dataset, and the new c is called c_{orig} .
5. The optimism in the fit, derived from the bootstrap sample is $c_{\text{boot}} - c_{\text{orig}}$.
6. Repeat steps 2 through 4 about 200 times.
7. Average the optimism estimate O .

8. $c_{\text{app}} - O$ is the nearly unbiased estimate of the *expected value* of the discrimination of the original model from which c_{app} was derived, accounting for overfitting.

How accurate are risk models for CVD?

Models of CVD risk derived from Framingham have been most closely studied and tested in other populations. Studies which compared observed 10-year risk with predicted, based upon Framingham equations, show that the predicted to observed ratios of disease varied between a 53% under-estimation of risk (95% CI: 73% to 33%), to a 187% overestimate (95% CI: 91% to 331%), depending on the incidence of disease in the population [10]. This finding suggests that the Framingham equation does not accurately predict risk in populations with different burdens of CVD, and that much of what causes cardiovascular disease remains unexplained.

1.8.3 Critique of predictive modelling

From a clinical perspective, a CVD risk prediction model is designed to inform a one-off clinical decision. It assists the clinician to answer the question ‘is my patient at sufficient risk of a CVD event to justify the risks (and potential benefits) of initiating, among other things, drug (or lifestyle) treatment to reduce this risk?’. It considers a single clinical interaction, so that future disease events, such as the diagnosis of diabetes, that occur after risk assessment, or during follow-up, may change the need for treatment.

It seems logical that every study which aims to predict prognosis is conducted within a specific environment and healthcare context which may influence risk of disease during follow-up. If drug treatment is considered necessary to include in prediction models, then other changes to the natural history, of similar magnitude, must also be included into such a model. For example, population tobacco control measures may lead to individuals within a cohort stopping smoking at a greater rate than they would have otherwise and ‘dis-

torting' the natural history of disease. Should such an intervention also be introduced into a Cox model to recreate the 'natural history' of disease? It raises the point that every prediction model is developed within a particular environment that may change over time, so complicating the question of what exactly constitutes a 'natural history', which suggests lack of treatment, and is difficult to define.

As discussed in section 1.8.2 on page 15, overfitting is a common consequence of predictive modelling [44]. This leads to differences in prediction performance when a model is used to predict disease events in a population other than that which it has been developed on. This optimism has been dealt with by statistical methods such as cross-validation and boot-strap resampling. Although these methods deal with the problem to some extent, they also have limits. The methods assume that when carrying out prediction modelling, relationships between risk factors and the outcome, which are specified by beta-coefficients, remain consistent. That is, if, in the future, a person's risk factor profile changes, their risk will accordingly change. This is not necessarily the case, and carries a causal assumption. Therefore, it is speculated that if causal variables are used in the model, the model will be more reliable and accurate in its predictions applied to different populations. Therefore, although prediction and causation have been considered separate, it is argued that they are linked.

To illustrate this issue further, an example is presented. Suppose that red wine drinking in a population is a protective factor, not because of its causal influence on CVD, but due to its association with other healthy behaviours, which are causally related with CVD. Suppose further that this variable is selected in a regression model used for prediction, and that this model continues to be used, without further development, like the Framingham model. If environmental conditions change, such that red wine intake were to become instead associated with unhealthy behaviours, the model may then become unreliable, as the association between red wine drinking with the truly causal variable (healthy behaviour) has changed. Drug treatment is one such exam-

ple, in which trial data suggests beneficial effects of treatment, yet observational studies suggest increased risk associated with drug use (for more details, see section 1.9 on page 23). These examples suggest that causal considerations should be taken into account when designing prediction models [120], or models should be regularly updated, so that the causal assumption implicit in regression models is not so necessary to depend upon.

From this argument, the information contained in drug treatment variables, either at baseline or during follow-up, may be subject to causal influence from unobserved variables, such as a patient's or doctor's propensity to take treatment. It is, therefore, plausible that including these variables will cause more harm than good when such a model is applied to risk prediction in future populations, due to the possible influence of collider bias. This point is discussed further in chapter 8 on page 105.

Time-varying covariates

An issue that complicates estimating the association between drug use and CVD incidence is that, during the time they are studied, they stop and start taking drugs. This may be accounted for in an analysis of risk, using an extension of the Cox model, which allows for time-varying exposures (or covariates) [35]. This does, however, complicate the interpretation of the statistical model and the measures of association derived from it. Hazard ratios extracted from a Cox model with time-varying terms do not compare disease hazards between groups of people who are either treated or not at a certain time point (usually baseline). Time-dependent hazard ratio estimates will instead compare, at any given time during follow-up, the hazard of outcome for an individual with treatment, compared to no treatment, at the same time point.

Fisher and Lin discussed a number of problems which arise with the use of time-varying exposures, such as drug treatment, in observational studies [35]. The first was that of time-varying confounding. In one example, researchers assessed the likely benefit of giving up smoking after being diag-

nosed with coronary artery disease [14]. The investigators were surprised to find that the time-varying survival analysis yielded an association for smoking cessation that indicated increased, rather than reduced risk of mortality, as would be expected. When individual subject's histories were analysed, they found that a hospital admission or severe health condition often immediately preceded smoking cessation. Unless the adverse health event, here, a 'time-varying confounder', was also accounted for in the time-varying model, then counter-intuitive results were obtained.

This raises an issue of relevance to the problem considered in this thesis. Drug treatment that is started during follow-up is likely to occur due to deterioration in health status, such as an admission to hospital or receiving a new diagnosis as a result of developing new symptoms. If this is so, then adverse prognostic information may also act as a time-varying confounder and offset the beneficial effects of drugs, just as it was shown to do in the smoking cessation study [14]. This poses a problem with including information about drug treatment during follow-up, since the factors that lead to initiation of drug treatment may not be available in the database, as with PREDICT.

Fisher and Lin [35] also point to the need to consider lag time (between exposure and likely risk of disease), and the functional form of the relationship between the time-dependent factor and the outcome under study. From a causal perspective, changes in treatment status that occur during follow-up are likely to be related to treat (or not) at baseline. Adjusting for subsequent treatment changes, like undertaking a per protocol analysis for a randomised trial (compared to the usually favoured 'intention-to-treat' method), may bias estimates of the effect of baseline exposure status, since this method introduces adjustment for changes which are likely to be on the causal pathway between baseline (drug) exposure and disease status.

These issues render the use of time-varying information in the present study difficult to implement and interpret. Whilst objections here are mainly posed at the use of regression models with time-varying exposures for causal inference,

the same issues are likely to apply for modeling for prediction.

1.8.4 How useful is a predictive model?

Indices of model performance, such as the *c*-statistic, or calibration slopes, may not, of themselves, lead to decisions of how clinically useful a model is, or allow for comparisons between models. In order to take a model to the next step, that is decide its clinical usefulness, one may consider the clinical consequences of clinical decisions made as a result of using the model to decide a cut-point which is used to treat or not to treat patients [133]. An analysis of the *net-benefit* of a model with and without drug use information is presented in chapter 6 on page 79.

1.9 How effective is drug treatment for the primary prevention of CVD?

In order to decide which drugs are important to use in predictive modelling, those drugs which have strong evidence to reduce risk of CVD events in patients free of the disease (primary prevention) need to be identified. The major classes of drug thought to reduce the risk of CVD are: statins, anti-hypertensive agents, smoking cessation aids and anti-platelet agents.

To help decide the efficacy of drugs, meta-analyses of individual randomised controlled trials are often used [43]. Meta-analyses of randomised trials are very good at reducing type-2 error, by increasing effective sample size; however, they may increase bias (such as from publication bias, measurement error, selection or loss to follow-up), by combining the results of studies with differing designs. The evidence of the effect for each class of drug used to reduce CVD risk is now considered in turn.

This section does not seek to exhaustively review the evidence for the drug treatment for the primary prevention of cardiovascular disease. Instead, it re-

views the most frequently cited meta-analyses in major journals.

Statins

In recent years, statins have been considered the strongest candidate drugs to reduce the incidence of CVD, in people judged to be at moderate to high risk of disease [71]. Many randomised studies have been conducted, and these results have been analysed together in several meta-analyses.

One such meta-analysis was conducted by Ray and colleagues [95], to assess the effect of statins on all-cause mortality in people who were at intermediate or high risk of disease. This study excluded patients with disease, and included a range of subjects between the ages of 50 to 75 years. The study only included published studies. A total of 244,000 person years, from 65,229 individuals were included, drawn from 11 randomised trials. The random effects pooled risk ratio was 0.91 (95% CI: 0.83 to 1.01), comparing those on statins with placebo treated controls. No evidence of between study heterogeneity was found ($I^2 = 23\%$; 95 % CI: 0 - 61%). Similarly, no evidence of publication bias was found, either from an Egger test or funnel plot. Meta-regression showed no relationship between percentage or absolute change in lipid level (low density lipoprotein cholesterol), or baseline lipid level and total mortality. At face value, the study does not convincingly support the use of statins to reduce CVD risk, since the summary measure of effect did not show a significant difference between the two treatment groups. Alternatively, it is also possible that the study is underpowered to detect a small effect on total mortality.

A second meta-analysis was published in the Cochrane library [114]. In contrast to that carried out by Ray, the Taylor review included studies which had recruited up to 10% of individuals with CVD (stroke, angina, myocardial infarction or stroke). In the total mortality analysis, the authors included thirteen randomized trials with over 48060 patients, with 1077/24408 (4.4%) dying in the statin group and 1223/23652 (5.1%) in the placebo group. All cause mortality was reduced, with a pooled odds ratio of 0.86 (95% CI: 0.79 to 0.94), indi-

cating a significant difference in outcomes. This study does, however, support the use of statins to prevent CVD.

In one other meta-analysis of randomised trials, which reported beneficial effects of statins for primary prevention, the selection criteria included trials with up to 50% of subjects with coronary heart disease [79].

The influential Cholesterol Treatment Trialists' study, reported in 2012 [18], that in individuals without vascular disease, the summary measure of effect of pooled studies indicated that statin use reduced CVD incidence with a hazard ratio of 0.75 (95% CI: 0.70 to 0.80), comparing people with an LDL-cholesterol level 1.0 mmol/L lower with those 1.0 mmol/L higher, in statin trials for which individual patient records were available. The authors used an unusual method to combine the effect of trials which compared trials one statin with another to those that compared statin to placebo.

The meta-analysis with the design most appropriate to answer the question of the efficacy of statins, in a population without CVD, would be that written by Ray [95]. From a theoretical perspective, no good reason stands out to include patients with existing disease in a meta-analysis, designed to answer the question of the efficacy of statins in a primary prevention population. So, taking the conclusion of the Ray study, it is likely that statins have no large effect on survival when used in people without CVD, since the effect is small (9% survival difference between the two groups which was not statistically significant). In contrast, however, both the Cholesterol Treatment Trialists' and the Taylor Cochrane review support the use of statins to prevent CVD in low risk populations. The discordance between the Ray and Taylor study conclusions, however, show that there is some uncertainty in the benefits of these drugs to improve overall survival. Further, these studies may be affected by publication bias, which was not detected in the original studies. This possibility is explored in detail in chapter 7 on page 94.

Antihypertensive drugs

Like the evidence that statins reduce the incidence of CVD, that for blood pressure lowering agents reducing the incidence of CVD or all-cause mortality is inconsistent. In people without CVD, one meta-analysis, in which individuals, without coronary disease, with mild hypertension (defined as a systolic blood pressure between 140 and 159 mmHg, or a diastolic blood pressure between 90 and 99 mmHg, or both) were assigned to active treatment or placebo, reported a null effect of treatment status on disease, whether coronary heart disease, stroke, or total mortality [29].

Other meta-analyses have reached opposing conclusions. For example, one conducted by Wright, which included a larger subset of patients with hypertension (blood pressure > 140/90 mmHg), 70% of whom had no underlying vascular disease at baseline, reported beneficial effects for all CVD-related outcomes (mortality, coronary heart disease, stroke, and all CVD events) for low-dose thiazide diuretics, ACE-inhibitors, but not for beta-blockers and calcium channel blockers [142]. Evidence from other meta-analyses report a similar null association, such as the study by Wiysonge [140]. This analysis reported no effect of beta blockers on all-cause mortality (pooled risk ratio of 0.99, 95%CI 0.88 to 1.11), compared to placebo treated controls. In contrast to total mortality, a small beneficial effect on CVD was reported (pooled relative risk: 0.88, 95%CI 0.79 to 0.97). No meta-analysis of the effect of blood pressure drugs, outlined here, assessed publication bias in the sample of selected trials.

Perhaps the largest meta-analysis of the effects of blood pressure lowering drugs in people with CVD and without was conducted by Law, Morris and Wald [65]. The authors pooled data from studies which compared active drugs with placebo and those that compared two or more active treatment groups. The authors concluded that reducing systolic blood pressure by 10mmHg, using drugs, would translate to a 25% reduction in CHD events, with a one third reduction in stroke events. Although this paper is frequently cited as evidence of the benefits of blood pressure reduction, the methods used are likely to in-

introduce bias, since the authors adjust for variables collected after randomisation (change from baseline systolic blood pressure). As is discussed further in section 1.9.1 on page 29, adjustment for variables which are collected after randomisation are unlikely to give accurate estimates of the treatment effect.

From this brief review, several factors suggest that one can not be confident about the effect of anti-hypertensive drug treatment to prevent CVD. The first is the inconsistency in the results of meta-analyses. If meta-analyses are all, theoretically, reporting the same overall effect of treatment and disease, it is puzzling that different studies arrive at different conclusions. In addition, the effects of drug treatment, in relative terms are small.

The differences in conclusions based on summary measures of effect of the drugs, between CVD and total mortality outcomes are also interesting. It can be assumed that total mortality effect (RR) is a weighted average of the effect of the drug on CVD death (RR_{CVD}) and its effect on death from other causes ($RR_{Other\ causes}$):

$$RR = wRR_{CVD} + (1 - w)RR_{Other\ causes}$$

where the weight w is:

$$P(\text{CVD death}|\text{untreated})/P(\text{All deaths}|\text{untreated}).$$

The weight w can be considered the proportion of all deaths that are attributable to CVD. So, if it is assumed that the effect of the drug on death from other causes is 1 (no effect), then the effect of the drug on total mortality will be located between the measure of association for CVD death and all-cause mortality. If the statistical power is great enough, based on the total sample size and number of outcomes, it may be expected that a significant reduction in CVD deaths would also be accompanied by a significant reduction in total

mortality. From an epidemiological standpoint, total mortality is likely to be less prone to measurement error, because a person's vital status is likely to be more consistently recorded, compared to diagnosing a myocardial infarction, for example. The discordant significance of summary measures of effect between total mortality and CVD suggests either that the drugs are increasing the risk of other causes of death, or that bias is likely to be present when trials report CVD outcomes.

Also, in meta-analyses of RCTs in low CVD risk patients, they will generally have low statistical power to assess the effects on total mortality. If CVD is a relatively small component of all cause mortality, then even a significant beneficial effect on CVD risk will not necessarily demonstrate a benefit on overall mortality.

Anti-platelet agents

Aspirin is commonly advocated for use to prevent the onset of CVD. Like the other drugs discussed so far, the evidence that this drug is beneficial in the primary prevention setting is again inconsistent.

In the most comprehensive meta-analysis of study findings, which included only individuals without disease, the authors found that although evidence suggested a reduction in CVD events (15% relative reduction in myocardial events [95% CI: 6 to 22%]), there was no effect of the drug on total mortality, due to an increased risk of the complications of bleeding [101]. However, the authors continued to advocate aspirin treatment to prevent coronary disease in those who were at high risk (greater than 1.5% annual risk). This was proposed by stating that the relative benefit of aspirin (at varying absolute risk levels), would translate into greater absolute risk reduction, in high risk patients, whereas the bleeding risk was assumed constant. The assumptions of this analysis are at odds with the manuscript's report of the summary statistical data for total mortality [127]. In this reference, it is argued that total mortality provides a better estimate of risk and benefit of treatment than simulations

based on disease specific outcomes.

Thus, for aspirin, there is no compelling evidence for an overall survival benefit from taking the drug, although there may be a reduction in CVD events, balanced by increased adverse events. A later meta-analysis, which addressed the effect of aspirin for primary and secondary prevention of CVD, re-inforced this view, with no significant difference detected in vascular mortality for patients with no clinical evidence of CVD at baseline [7].

Smoking cessation aids

The incidence of first CVD event is likely to be influenced from drugs designed to help smokers quit. Cigarette smoking is consistently found to increase risk of CVD in observational studies [50, 1, 25, 27]. The absolute benefits of drugs to help people stop smoking, however, are likely to be small, as the therapy only benefits those in the population who smoke, and success in giving up smoking is rare [104]. However, meta-analyses of randomised studies consistently highlight an almost doubling of six monthly quit success when compared to those assigned to placebo [104]. At a population level, however, the effect of smoking cessation on the incidence of CVD is more modest when tested in a population of smokers [3].

1.9.1 Sources of bias in meta-analyses of randomised studies

A major threat to the validity of the conclusions of meta-analyses of randomised trials is from publication bias. In a systematic review and meta-analysis of this subject, in which authors were given access to a source of published and unpublished trial data, revealed that:

‘Trials with positive findings were more likely to be published than trials with negative or null findings (odds ratio 3.90; 95% confidence interval 2.68 to 5.68)’ [56].

This study contrasts with the relatively infrequent reporting of publication bias in meta-analyses. Part of the lack of detection of bias may be related to the limits of existing methods, and alternative analytical techniques are used in chapter 7 on page 94 to explore whether bias is likely to influence the results of meta-analyses of the effects of statin use.

Other studies support the idea that industry funded studies are more likely to be biased than independently funded trials. A review found that of 95 industry funded studies including statins (either compared with placebo, or comparing to other statins), the results and conclusions were 20 times more likely to support the sponsor's drug, compared to the comparator group (odds ratio = 20.2 [95% confidence interval 4.4 to 93.0])[9]. This review included only three larger phase 3 trials, and was conducted almost exclusively on smaller phase 2 studies, which may be more likely to be biased. This finding, does, however, raises suspicion of the findings of meta-analyses which are based on industry sponsored trials. Other studies support the idea that industry funded studies are more likely to be biased than independently funded trials. A review found that of 95 industry funded studies including statins (either compared with placebo, or comparing to other statins), the results and conclusions were 20 times more likely to support the sponsor's drug, compared to the comparator group (odds ratio = 20.2 [95% confidence interval 4.4 to 93.0])[9]. This review included only three larger phase 3 trials, and was conducted almost exclusively on smaller phase 2 studies, which may be more likely to be biased. This finding, does, however, raises suspicion of the findings of meta-analyses which are based on industry sponsored trials.

Some evidence of bias is evident in the way some meta-analyses are conducted. For example, the Cholesterol Trialists' Collaboration [18, 17, 8] has published studies which have inflated measures of effect, using post-randomisation variables (for example: low-density lipoprotein cholesterol measured one year after the start of treatment). These studies combine studies which compare one statin with another, and those that compare statin to placebo. The au-

thors summarise the effects of each trial by raising the relative risk of the effect of the drug by the power of the mean difference in low density lipoprotein cholesterol between the two groups after one year. In seeking to pool as much data as possible, the authors have introduced further complicating factors into their analyses. Firstly, a relative risk raised to the power of a mean difference in risk factor is no longer a relative risk and should not be labelled as such. This analysis also entails the assumption that the benefits of statins are mediated entirely through the change in the cholesterol risk factor. This assertion about the effects of statins is not universally accepted [76]. Some evidence of bias is evident in the way some meta-analyses are conducted. For example, the Cholesterol Trialists' Collaboration [18, 17, 8] has published studies which have inflated measures of effect, using post-randomisation variables (for example: low-density lipoprotein cholesterol measured one year after the start of treatment). These studies combine studies which compare one statin with another, and those that compare statin to placebo. The authors summarise the effects of each trial by raising the relative risk of the effect of the drug by the power of the mean difference in low density lipoprotein cholesterol between the two groups after one year. In seeking to pool as much data as possible, the authors have introduced further complicating factors into their analyses. Firstly, a relative risk raised to the power of a mean difference in risk factor is no longer a relative risk and should not be labelled as such. This analysis also entails the assumption that the benefits of statins are mediated entirely through the change in the cholesterol risk factor. This assertion about the effects of statins is not universally accepted [76].

A method that adjusts the summary measure of effect of treatment by 'post-treatment' variables is likely to bias results and be inferior to traditional meta-analyses which compare disease outcomes by treatment allocation. As discussed by Gelman and Hill [38], controlling for post-treatment variables leads to comparisons that are frequently not what was intended to answer the research question. In the Trialists' studies, adjusting the overall treatment effect

by the mean change from before to after treatment LDL compares those who have a reduction either due to dietary or drug means, with those who have not reduced their level, who may or may not have received the drug. The 'post-treatment' adjusted risk ratio is therefore not addressing the issue of whether those who were treated with drugs were better off than those who were not.

While other authors have questioned the conclusions of the Trialists' studies [108], the objections of these authors relate mainly to the lack of statistical evidence that the benefits of statins are mediated through reduction in cholesterol levels. Questioning the validity of this aspect of the statistical analysis has not been raised in relation to the Trialists' assessment of statin efficacy.

Guidelines which support the use of drugs to prevent CVD, generally, do not discuss the possibility of publication bias (see [83] for example), and in some cases, overlook the shortcomings of analyses that justify their recommendations. In a recently published North American guideline, for example, the Trialists' study was extensively referenced in support of lowering the threshold at which to initiate statin treatment in individuals judged to be at high risk of CVD [111]. The guideline indicated that the summary study was the highest possible level of evidence, without discussion of the validity of the study design, raised here.

1.10 Summary with hypothesis, aims and contribution

The conclusion of this brief literature review is that uncertainty exists about whether drug treatment prevents CVD. The differences in frequencies of CVD or mortality among the treated and untreated in drug trials is small, and it is widely accepted that small effect sizes are more likely to result from bias, rather than represent a true effect [55]. Moreover, evidence of inconsistency in results of meta-analyses of drugs designed to reduce blood pressure are reported, such that some drugs show evidence of benefit, while others do not.

It seems plausible, then, that drug treatment is unlikely to seriously distort the accuracy of CVD risk prediction equations, since the effects of the drugs, from randomised trials, are small. Or, the findings are, at least, not large enough to produce consistently positive findings in meta-analyses.

This conclusion, that any beneficial effects of drugs to prevent the onset of CVD were small and inconsistent, may be at odds with the understanding of the medical community, and to suggestions for a polypill to reduce the incidence of CVD [135]. At the time of writing, no studies appear to have been published of the effects of the polypill to reduce the incidence of disease.

The aim of this thesis, therefore, is divided into research themes. First, the proportion of people who are treated with drugs will be estimated, along with the magnitude of the association between drug use and disease incidence from the PREDICT cohort. This analysis seeks to quantify the magnitude and the direction of the effect of drug treatment, since only a large association between drug use and CVD survival is likely to bias risk estimation. This is estimated for both statins and anti-hypertensive drugs in chapter 3 on page 41. This study was initiated, analysed and written by the author, but editing was contributed by the author's supervisor and co-supervisor. This analysis has not been submitted for publication.

To address the potential weaknesses of the first analysis, a second similar analysis was conducted to address potential threats to the study validity. A potential weakness of the analysis in chapter 3 is that unmeasured confounding by indication for the drug may bias the measures of association between drug and disease. That is people at higher risk than measured from available information were selected for treatment. Several different methods will be used to enhance causal inference from observational data, such as by *propensity score matching*, in which people who are initiated on statins are matched with subjects who share similar characteristics, but are not taking these drugs. This analysis is presented in chapter 4 on 53. This analysis was carried out to compare the risk of treatment in this cohort with drug effects reported in meta-

analyses. The study was initiated, analysed and drafted by the author, with editorial and design assistance from the author's supervisors. It has not, at the time of writing, been submitted for publication.

A limitation of both previous chapters was that they rely on statistical models which are not designed to assess causal questions, but instead provide evidence of statistical association. Learning Bayesian networks are an alternative to traditional regression methods, and are used to explore the likely causal relationships between variables commonly used to predict CVD risk, including drug treatment assessed at baseline (chapter 5 on page 66). This analysis was initiated by the author, with editorial and methodological assistance from supervisors and Dr Susan Wells. The study has been published during the course of this work [123].

Other themes explored in this thesis include the clinical utility of CVD risk prediction equations as a decision making tool. A metric of clinical utility, *net-benefit*, is used to decide whether a patient's drug use is important in their assessment of risk, used to guide treatment decisions (chapter 6 on page 79). This study was initiated, designed and drafted by the author, with editorial assistance from the author's supervisor. It has not been submitted for publication, but has been presented in an epidemiology conference [121].

Since they are considered the highest level of evidence, meta-analyses of randomised controlled trials will be examined for publication bias, using methods which obviate some of the problems of existing techniques [57] (chapter 7 on page 94). This study was initiated, analysed and preliminary writing completed by the author, with editorial and design input from the author's supervisor and Dr Federica Barzi. The study has been submitted for publication (*BMJ Open*), but at the time of writing was still in peer review.

To compare the influence of drug treatment with other possibly important predictors of CVD risk, chapter 8 explores the statistical evidence for a causal link between serum urate and CVD. This study was designed, analysed and preliminary drafts were written by the author. Editorial and analytical assis-

tance was given by the author's supervisor and co-supervisor, with other researchers (Sue Crengle, Dudley Gentles, Susan Wells, Nicola Dalbeth and Andrew Kerr) providing editorial assistance. This work was published during the course of work toward this thesis [125].

A number of other published works have emerged from the work undertaken toward this thesis. They include methodological points, summarised in editorials ([124, 120]), a criticism of a meta-analysis of the effect of aspirin identified during the literature review [127], and a review and critique of the evidence for the dietary cause of CVD [126].

The conclusions drawn in this thesis are the author's own, although editorial assistance has been given by the author's supervisor and co-supervisor. The author had no input into the design of the original PREDICT cohort study, and did not conduct the data linkage necessary to undertake these analyses. The author is indebted to members of the PREDICT team, Romana Pylypchuk and Tadd Clayton, who assisted with these tasks. All studies were, however, designed, analysed and drafted by the author.

Chapter 2

Methods

Liew, Doust and Glasziou [69] argue that treatment received during follow-up is a major problem of contemporary CVD risk prediction studies. From the introductory literature review, however, this problem is too difficult to untangle using available statistical methods, due to the issues raised by Fisher and Lin [35] about the difficulty of estimating time-varying effects from observational studies. Particularly, when exposure is likely to be confounded in a time-varying manner by other prognostic variables. From the review of the effects of drug treatment, it is not possible to say with confidence what the effects of individual drugs are, so accounting for these effects during follow-up, in the manner proposed by Liew seems to be of limited value.

The rest of this thesis, apart from one chapter about serum urate and CVD incidence, addresses how best to deal with baseline treatment status or that initiated after risk assessment. So, the magnitude of the association between drug treatment and CVD is investigated in the PREDICT cohort, to ascertain whether people who take the drugs or are initiated on to them are at higher or lower risk of disease, conditional on other commonly measured risk factors for CVD. This chapter provides an overview of the PREDICT sample available for analysis.

2.1 The study sample: the PREDICT cohort

The analyses presented in the following chapters are based on subgroups of a cohort assembled by general practices in the Auckland and Northland regions of New Zealand using 'PREDICT', a web-based, clinical decision support system for CVD risk assessment and management [96]. Cohort participants were patients attending a GP who were assessed for CVD risk (currently using a modified Framingham Heart Study CVD risk prediction score [1]). The information, from participating GPs, was simultaneously recorded in patient records and on a secure PREDICT web server. Data on each patient was then linked to national health databases, using an encrypted unique identifier, the New Zealand national health index (NHI). Databases that were linked included: hospital discharge diagnoses, mortality, drug dispensing, and laboratory test results. Enrolees between 1 January 2006 and 15 October 2009 were selected for this analysis since comprehensive dispensing data were available for this period. Subsets of this population were selected, depending on the nature of the study design. Follow-up was limited to the end of 2009 (31 December). In the following chapters, individuals aged under 30 or 80 years and over at the time of screening were excluded from analyses, since coding of endpoints was thought to be less consistent in older age groups, and CVD is rare under the age of 30 years. People with a history of prior CVD or heart failure, identified by a family doctor as having a diagnosis of CVD or hospitalisation with CVD in the last five years, or those dispensed a loop diuretic in the six months before assessment, who were assumed to have a diagnosis of congestive heart failure, were excluded. These individuals were left out since they were likely to have already developed one of the CVD endpoints and were therefore no longer at risk of disease. In the analysis which considered initiation of drug treatment after risk assessment, people already treated with either statins or blood pressure lowering agents, or both drugs, were excluded from the study sample.

2.1.1 Other covariates

As part of the PREDICT protocol, patients had their age, sex, ethnic group, smoking and diabetes status, blood pressure (both systolic and diastolic), and most recent serum lipid profile (serum total: high density lipoprotein cholesterol ratio) recorded. In the chosen cohort, these variables were completely recorded with no missing data. Other covariates were incompletely recorded and were made available by linking information to a community laboratory test database. Serum urate concentration, HbA1c (%), individual components of the lipid profile (such as low density lipoprotein cholesterol, triglyceride and HDL concentration) were available from this data source, although not all individuals had these tests available. In the following analyses, tests that were conducted up to five years before baseline CVD risk-assessment were included, or up to two weeks after this event.

2.1.2 Outcomes

CVD events (acute coronary syndromes, coronary procedures, stroke, transient ischaemic attack, haemorrhage stroke, peripheral vascular disease, peripheral arterial procedures and congestive heart failure) were identified by using unique alphanumeric codes to link to hospital discharge and mortality records. International Classification of Disease Codes (ICD) which included these conditions or a procedure to exclusively treat one of these conditions were used, as previously described [96]. In some analyses, follow-up time was recorded from the date of baseline assessment, whereas in other analyses, in which treatment was defined in the first six months after follow-up, it was started after this time. Subjects who had an event between assessment and six months afterward were excluded. The analysis was of the outcome 'any CVD event', including immediate death that was coded as caused by CVD. Loss to follow-up, such as from external migration, could not be tracked. If individuals stayed in New Zealand, however, it was very likely that their event would be

recorded, because New Zealand has very few acute-care private medical care facilities to treat CVD, so most diagnoses of CVD are likely to be recorded in lists from public hospital discharge diagnoses.

2.1.3 Overview of the cohort

The recruitment period for the PREDICT cohort started in 2002, however, as drug dispensing information (loop diuretic use) is used to exclude people with disease, the analyses use subjects who are recruited after these data became available (Jan 2006; with drug based exclusions applying from July 2005). The methods in chapters 3, 4 and 8 are based on survival estimation, so they include a larger number than those analyses of chapters 5 and 6 which require observed outcomes (figure 2.1). Chapters 5 and 6, therefore, are based on a much more restricted cohort recruited during the 2006 calendar year.

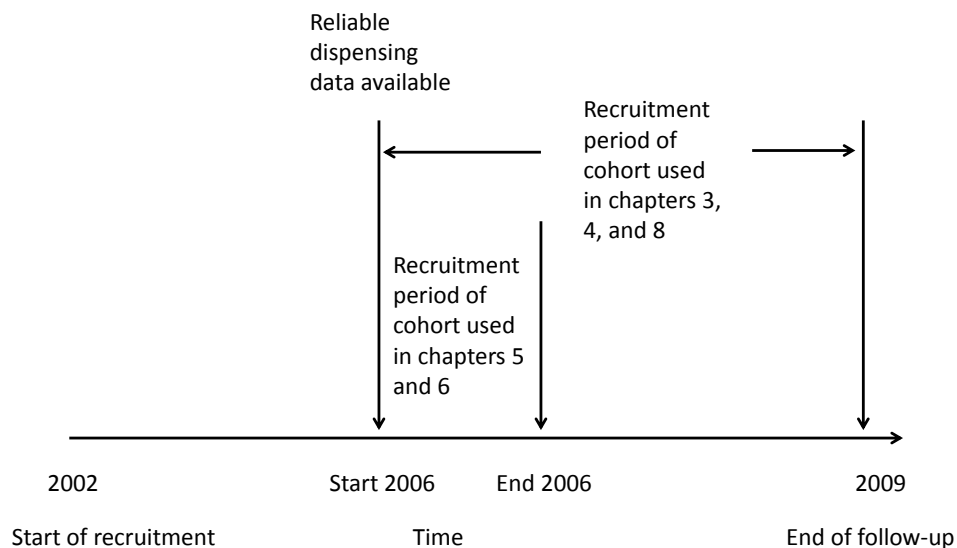


Figure 2.1: Recruitment periods of the cohorts used in the different chapters of this thesis.

This description relates to data available for all of the analyses conducted in this work, however, subtle variations in the sample and methods underlie each analysis, so they will now be discussed in turn. First, some basic exploratory analyses were carried out, documenting the proportion of individuals in the cohort who at baseline took drugs which potentially prevent CVD.

Chapter 3

Analysis 1: The magnitude of the association between drug initiation and CVD

3.1 Introduction

3.1.1 Drug use at baseline risk assessment

The aim of this chapter is to assess the magnitude of the association between drug initiation and CVD risk. Before this is carried out, a brief descriptive analysis of those treated on preventive drugs, both at baseline and during follow-up, will be presented.

First, the magnitude of the prevalence of drug treatment at baseline was estimated. In the 79027 individuals who were enrolled over the selected time period, who were free of CVD or heart failure and who had complete data available on which to make a CVD risk assessment (3 individuals had a missing total: HDL-cholesterol ratio), 21% of the total were taking statin treatment in the six months before baseline assessment (16257 / 79027). A higher proportion

of the total (32%) were taking at least one anti-hypertensive agent at baseline (25503/ 79027). About 12% of the total (9516/ 79027) were treated with two drugs at baseline.

Of the drugs that will not be considered further, 0.6% (467/ 79027) were taking nicotine replacement therapy, whereas 16% (12920/79027) were recorded as taking aspirin at baseline. In the case of aspirin, however, many more individuals may have been sourcing the drug 'over-the-counter', and were not recorded as using it.

The relationships between drug use at baseline are depicted in a scaled rectangle diagram [75] (figure 3.1). The outer square represents the total PREDICT cohort. Inside this are the different drug use categories which are proportional in size to the total PREDICT sample, along with the overlap between use of different agents, and diabetes status. Numbers in the figure represent the total in each combination of diabetes and drug use category. There is considerable overlap between use of the different drugs, with many subjects taking all three classes at baseline. Further, the majority of people with diabetes are taking at least one preventive drug.

3.1.2 Drug use during follow-up

Since statins are the most homogeneous of the drugs used to prevent CVD, and are generally considered to have the best evidence of efficacy, this section will consider the issue of treatment at baseline, and during follow-up with these drugs. Table 3.1 shows that about 10% of patients not treated with statins go on to receive treatment during follow-up, so that during follow-up, about 20% of individuals receive treatment. This table, does not, however, show individual treatment trajectory.

To examine individual trajectories of statin treatment, a subset of the total sample (79030 individuals) was examined, who had at least two years follow-up ($n=26408$), and were recruited between the 1st of January 2006 to the 31st of December 2007 (this differs from the more restricted, cohorts recruited over

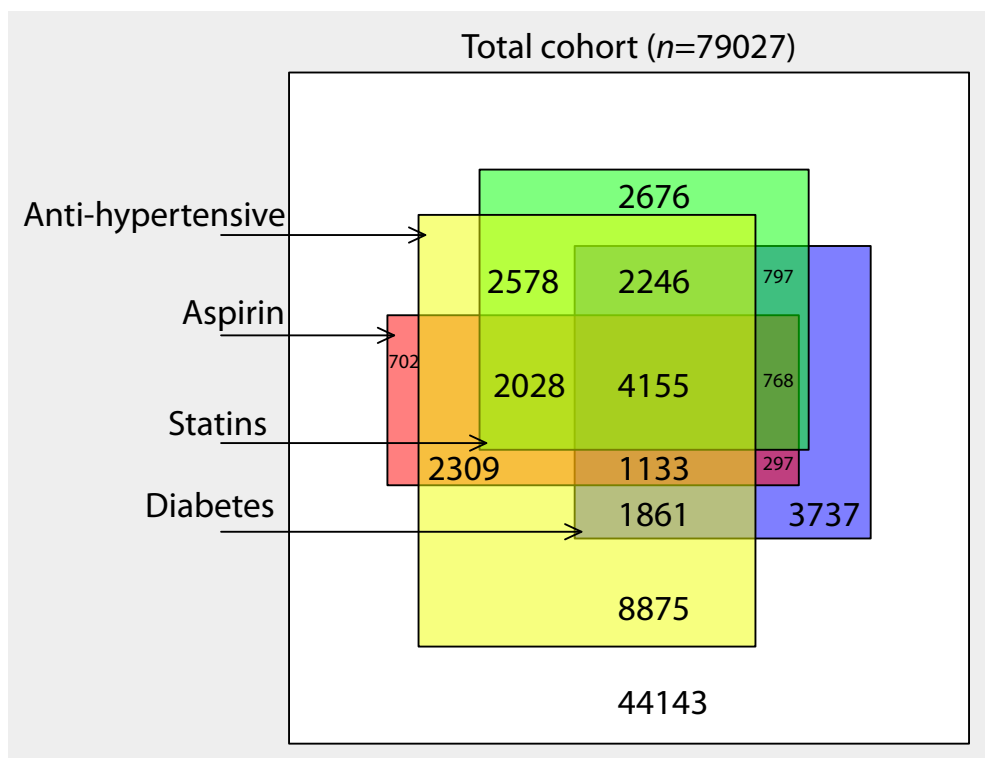


Figure 3.1: Scaled rectangle diagram [75] depicting preventive drug use at risk assessment and diabetes status in the PREDICT cohort who were recruited in the years 2006 to 2009.

Table 3.1: Proportion of PREDICT cohort, using statins at baseline and during follow-up.

Time after enrolment (months)	Total at risk (n)	Proportion taking statin	Total not taking statins at enrolment	Proportion taking statins, of those not treated at enrolment
0	79030	0.20	62889	0.00
6	70369	0.23	56053	0.09
12	53763	0.24	42724	0.10
18	38919	0.24	30779	0.11
24	26408	0.26	20610	0.13
30	15525	0.14	11769	0.09
36	6033	0.21	5195	0.14
42	2900	0.13	2489	0.09

one year ($n=6256$) analysed in chapters 5 and 6). Figure 3.2 is a scaled rectangle diagram [75] depicting the proportion of the subset who received treatment at baseline, between 7 and 12 months after baseline and later, at 19 to 24 months follow-up. The outer rectangle represents the total cohort, with the overlapping inner rectangles representing treatment status (any dispensing during the relevant period). The diagram shows considerable overlap between the three different groups, suggesting that the majority of people who are treated at baseline, remain on the drugs during follow-up. For this reason, drug use, either at baseline or in the first six months of follow-up, is used in the following chapters.

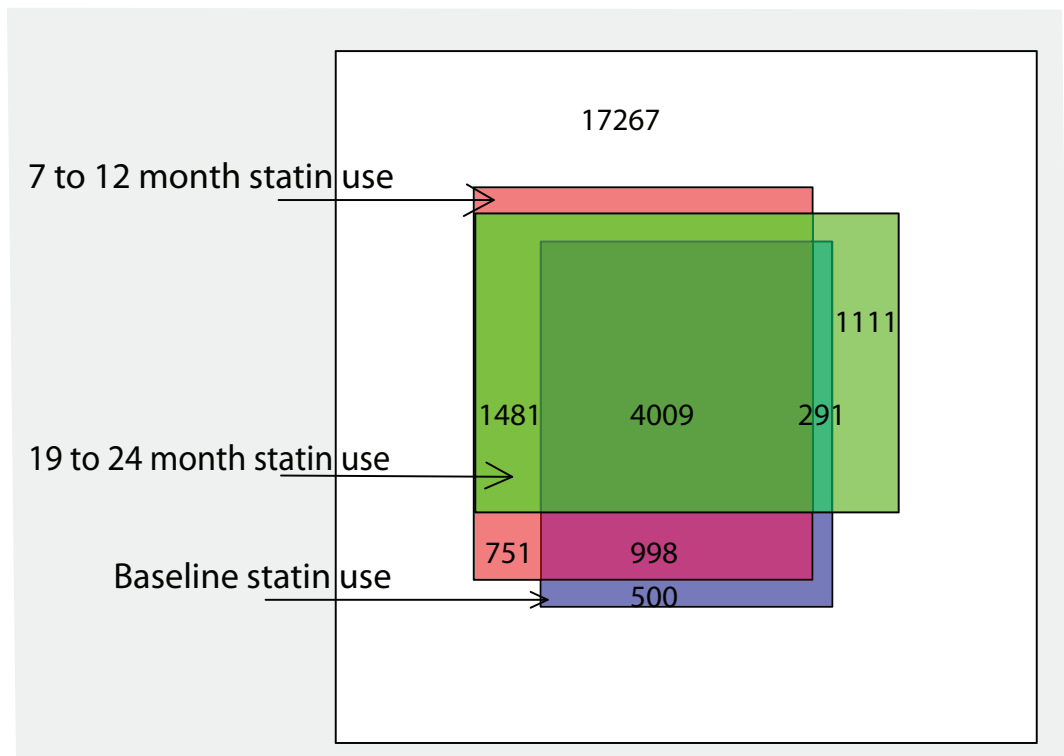


Figure 3.2: Scaled rectangle diagram [75] depicting statin use trajectory in a subset of the PREDICT cohort who were recruited in the years 2006 and 2007 and had at least two years of follow-up ($n=26408$).

3.2 An overview of the analyses

3.2.1 Design

The aim of this chapter is to estimate the magnitude of the association between drug use and CVD incidence. This analysis was carried out on the cohort described earlier (section 2.1 on page 37), but excluded all participants who had had been dispensed either an anti-hypertensive agent or statin in the six months before enrolment in the study. Usually measured baseline variables of CVD risk were included, with preventive drug initiation defined as redeeming a prescription of the drug within six months of the risk assessment. Individuals who had an event during this six month period were excluded, and follow-up time was restricted to the period after the drug treatment variable had been defined, that is, from six months after CVD risk assessment. The rationale for shifting the time of observation forward by six months was to ensure a temporal separation between drug treatment status (the exposure of interest) and the outcome (CVD survival). This obviates a potential bias of reverse causation for those who may have experienced an event during this six month period.

3.2.2 Statistical Analysis

Statistical analysis centred on developing a survival model for incidence of any CVD event from 6 months after assessment. A Cox regression model was used, with time-to-disease-event from 6 months after baseline enrolment considered the disease outcome. Individuals were censored if they either died from other causes, or reached the end of follow-up without experiencing a CVD-related hospital admission or death. Restricted cubic splines were used to investigate the relationship between continuous variables and time-to-event, as a check on the modelling assumption of linearity. Hazard ratios for continuous variables were reported by comparing the relative hazard between the risk at the 16th and 84th centiles of the variable (one standard deviation either side of the mean for a normal distribution). This allows direct comparison to relative hazards of

binary variables [124]. Drug treatment was defined as one of the following categories: 'none', 'statin', 'anti-hypertensive', or 'both agents'.

3.3 Results

A total of 43366 individuals were enrolled during the period of the study, and met the selection criteria. Table 3.2 shows the characteristics of those enrolled in the analysis, by treatment status at 6 months follow-up. The majority were untreated by the start of follow-up (87%; 37744/43366), with the remainder initiating one or both treatments. Women were more likely to initiate statins or both treatments, whereas men were more likely to receive anti-hypertensive drugs only. Older subjects were more likely to be given CVD drugs, particularly both together or anti-hypertensive drugs only. Pacific people were about twice as likely as other ethnic groups to be prescribed both drugs. People with diabetes were much more likely than non-diabetics to be initiated on one or both drugs (odds ratio: 9.73; 95% CI: 8.60 to 9.72, comparing the odds of starting both drugs compared to none).

Among the 13677 participants who had at least one HbA1c test available, those given both drugs were more likely to have higher mean HbA1c recordings than others. Patients treated with one drug had a median Framingham risk score almost twice those who remained off treatment. However, those on both treatments had, on average, almost twice the predicted 5-year cumulative incidence of disease, compared to those on one drug alone.

Strong associations were observed between those taking a specific drug, and the CVD risk factor that is influenced by that drug. For example, those given statins only were more likely to have a high total-to-HDL cholesterol ratio (and positive family history of premature CVD) compared to other categories of drug use. Similarly, those initiated to blood pressure lowering drugs had higher levels of systolic blood pressure than all other categories.

During follow-up (median 408.4 days; interquartile range 203 to 688 days),

Table 3.2: Total PREDICT cohort, by statin and anti-hypertensive treatment status in the six months after enrolment. Data are complete unless indicated otherwise.

Factor	None (col. %*)	AHT only (col. %*)	Statins only (col. %*)	Both (col. %*)	Total (col. %*)
Total	37744	2124	2300	1198	43366
Sex					
Men	15500 (41.1)	949 (44.7)	898 (39.0)	453 (37.8)	17800 (41.0)
Age at risk assessment (years)					
Median (IQR)	51.1 (45.2, 58.8)	54.6 (47.3, 61.7)	53.5 (46.5, 60.5)	55.5 (48.3, 62.7)	51.5 (45.3, 59.2)
Ethnic group					
Other	20539 (54.4)	1054 (49.6)	1 266 (55.0)	465 (38.8)	23 324 (53.8)
Māori	5 950 (15.8)	383 (18.0)	314 (13.7)	178 (14.9)	6825 (15.7)
Pacific	8 686 (23.0)	535 (25.2)	477 (20.7)	449 (37.5)	10 147 (23.4)
Indian	2 569 (6.8)	152 (7.2)	243 (10.6)	106 (8.8)	3 070 (7.1)
Current smoker?					
Yes	7287 (19.3)	387 (18.2)	540 (23.5)	295 (24.6)	8 509 (19.6)
Diagnosis of diabetes?					
Yes	2279 (6.0)	321 (15.1)	346 (15.0)	461 (38.5)	3407 (7.9)
Family history of premature CVD?					
Yes	5 370 (14.2)	398 (18.7)	605 (26.3)	221 (18.4)	6594 (15.2)
Systolic blood pressure (mmHg)					
Median (IQR)	125 (116, 136)	145 (130, 160)	128 (120, 138)	140 (130, 160)	128 (118,140)
Framingham 5-year risk of CVD (% from Anderson et al. risk equation [1])					
Median (IQR)	3.2 (1.6, 6.0)	5.5 (3.0, 9.5)	5.3 (3.0, 9.5)	9.3 (5.1, 14.5)	3.5 (1.7, 6.6)
Serum HDL-cholesterol (mmol/L) (n=37231)					
Median (IQR)	1.38 (1.14, 1.66)	1.33 (1.10, 1.61)	1.26 (1.06, 1.51)	1.25 (1.05, 1.50)	1.36 (1.13, 1.65)
Total-to-HDL-cholesterol ratio					
Median (IQR)	3.9 (3.1, 4.8)	3.8 (3.1, 4.7)	5.0 (4.1, 5.9)	4.6 (3.8, 5.5)	3.9 (3.2, 4.9)
HbA1c (%) (n=13 677)					
Median (IQR)	5.8 (5.5, 6.1)	6.0 (5.6, 7.0)	6.0 (5.6, 7.0)	7.0 (6.0, 9.0)	5.8 (5.5, 6.3)
Serum urate (mmol/L) (n=16 401)					
Median (IQR)	0.35 (0.29, 0.41)	0.35 (0.29, 0.42)	0.35 (0.30, 0.42)	0.37 (0.31, 0.44)	0.35 (0.29, 0.41)
Follow-up time (days)					
Median (IQR)	401 (190, 681)	495 (303, 756)	426 (261, 713)	483 (287, 737)	408 (203, 688)
CVD events?					
Yes	305 (0.8)	46 (2.2)	24 (1.0)	37 (3.1)	412 (1.0)
Death (all-cause)?					
Yes	140 (0.4)	20 (0.9)	10 (0.4)	13 (1.1)	183 (0.4)

*Unless otherwise indicated.

SD: standard deviation.

412 CVD events occurred. The results of the Cox model analysis are summarised in table 3.3. Hazard ratios for the drug treatment variable were progressively attenuated with increased number of variables included in the model. Of the drug treatment categories, the highest risk group was those taking both drugs after baseline assessment (adjusted hazard ratio 1.50; 95% CI: 1.03 to 2.17). Similarly, a progressive reduction in the magnitude of the blood pressure and diabetes measures of association occurred with increasing numbers of covariates added to the model. The point estimate for the 'statin only' category was reversed in the fully adjusted models. No evidence was found of time-varying associations using the Grambsch and Therneau test for the proportional hazard assumption. Similarly, no evidence of interaction was found between age and drug category, diabetes status and drug category, or age and diabetes status from likelihood ratio tests.

In contrast, adjustment resulted in increased magnitude of variables associated with ethnic group, such that in the full model, the effects of the Indian and Māori ethnic groups were associated with a 40% increased risk, compared to 'Others'; whereas Pacific peoples were at almost twice the risk of disease (compared to 'Others').

Although diabetes is a factor in the above model, the imbalance observed in diabetes diagnosis by drug treatment status suggested that it was also worthwhile to conduct an analysis excluding those diagnosed with diabetes at baseline ($n=39959$, number of events=364). The results were not substantially different: adjusted hazard ratios for use of 'both drugs' was 1.40 (95% CI: 0.89 to 2.20), for 'anti-hypertensive drugs only' was 1.39 (95% CI: 0.97 to 1.97), and for 'statins only' was 0.89 (95% CI: 0.57 to 1.38), compared to those who remained untreated.

Table 3.3: Crude and adjusted associations from a Cox model of CVD incidence including categories of anti-hypertensive and statin use (n=43366; 412 CVD events).

Exposure	Low risk	High risk	Crude hazard ratio (95% CI)	Adj.† hazard ratio (95% CI)	Adj.* hazard ratio (95% CI)
Sex	Women	Men	1.52 (1.23 to 1.87)	1.94 (1.57 to 2.39)	1.78 (1.43 to 2.20)
Age (years)	42.1 ⁺	62.8 ⁺⁺	3.86 (3.17 to 4.69)	5.13 (3.92 to 6.71)	4.83 (3.67 to 6.34)
Ethnic group	Other	Māori	1.15 (0.87 to 1.52)	1.51 (1.13 to 2.02)	1.43 (1.06 to 1.91)
	Other	Pacific	1.41 (1.13 to 1.77)	1.96 (1.55 to 2.47)	1.97 (1.56 to 2.49)
	Other	Indian	0.77 (0.50 to 1.20)	1.37 (0.87 to 2.15)	1.42 (0.91 to 2.24)
Total to HDL-cholesterol ratio	2.9 ⁺	5.3 ⁺⁺	1.40 (1.21 to 1.62)	1.45 (1.27 to 1.66)	1.24 (0.97 to 1.58)
Diabetes status	No	Yes	1.56 (1.16 to 2.11)	1.19 (0.88 to 1.62)	1.06 (0.77 to 1.45)
Smoking status	Non-smoker	Smoker	1.73 (1.41 to 2.14)	1.94 (1.56 to 2.41)	1.89 (1.52 to 2.36)
Systolic blood pressure (mmHg)	110 ⁺	142 ⁺⁺	2.03 (1.74 to 2.37)	1.61 (1.37 to 1.90)	1.46 (1.22 to 1.74)
Drug use at 6 months					
Statins alone	None	Statin use	1.16 (0.76 to 1.75)	0.97 (0.64 to 1.47)	0.83 (0.55 to 1.27)
Anti-hypertensives alone	None	AHT use	2.19 (1.61 to 2.99)	1.70 (1.24 to 2.33)	1.34 (0.96 to 1.87)
Both statins and anti-hypertensive drugs	None	Use of both drugs	3.25 (2.31 to 4.57)	2.02 (1.41 to 2.90)	1.50 (1.03 to 2.17)

+ 16th centile. ++ 84th centile.

† Adjusted for age, sex, ethnic group, smoking status, and diabetes status.

*Adjusted for all covariates shown in table

(drug treatment, lipids and systolic blood pressure further added to the model).

AHT: anti-hypertensive drug; CI: confidence interval; Adj.: adjusted;

CVD: cardiovascular disease; HDL: high density lipoprotein.

3.4 Discussion

In this analysis, after adjusting for confounders, those treated by both a statin and anti-hypertensive drug were at raised risk (HR: 1.50; 95% CI: 1.03 to 2.17) of CVD compared to those who took neither drug class. Subjects initiated on anti-hypertensives were also at raised risk (HR: 1.34; 0.96 to 1.87) while those on statins alone were at a lower risk (HR: 0.83; 0.55 to 1.27), although single treatment hazards were not statistically significant. Since treatment with anti-hypertensives and statins are likely to be beneficial, the results suggest that, at least, for statins and anti-hypertensives combined and anti-hypertensives alone, those initiated on treatment are at an intrinsic higher risk than is adequately adjusted for in the statistical model. As shown in table 3.2, people who were given one or both drugs had a poorer prognostic profile, particularly with diabetes status and overall CVD risk calculated by a Framingham risk equation [1].

The strengths of this analysis included the large database of patients who were risk assessed and managed in routine primary care, yet with largely complete data available. The use of a standard risk tool linked to routine health data is likely to reduce both selection bias, and loss to follow-up, other than from external migration. This analysis differs from most other CVD risk prediction studies, as it separates temporally baseline clinical and demographic characteristics from drug treatment, and CVD events. This means that the drugs could not have influenced baseline risk factors, as would have happened if drug use and risk factors were recorded at the same point in time. Most other studies measure all predictor variables, including drug treatment, at the same point: enrolment into the study [49, 97]. This design, which incorporates the intervention of providing absolute risk and access to drug treatment guidelines, is more likely to provide insight into what factors lead clinicians to initiate treatment based on assessment of CVD risk, and avoid the distortion that use of drugs have on baseline risk factor profiles.

A limitation of the analysis was that it did not address changes in treatment during follow-up. Time-varying effects of changing treatment during follow-up were not adjusted for, as time-varying confounding is likely to bias measures of association [35], and patients were known to, in general, have a high probability of staying on treatment if they received the drug in the six month period after assessment [77]. If patients with a generally poor prognosis, classified in this analysis as untreated, are likely to go onto treatment during follow-up, this may lead to underestimation of the beneficial effects of drug treatment. The lack of information available about change in subjects' prognosis during follow-up, which were likely to influence decisions to start drug treatment during the study, precluded an analysis of the time-varying nature of exposure. Further, follow-up time was relatively short. If the increased risk in the combination treatment variable represents the difference between adverse prognostic factors and beneficial effects of the treatments, then doctors are actually appropriately targeting those at higher than average risk with this treatment.

There are few other studies with which to directly compare these results. Compared with 17 analyses which report CVD risk assessment and were carried out in the era between 1991 and 2010 [2, 138, 25, 4, 15, 72, 51, 49, 52, 97, 98, 141, 66, 143, 37, 89, 21], only two, both conducted by Reynolds [98, 97], report information about prevalence of statin use at baseline assessment. Unfortunately neither Reynolds study reported the associated hazard ratio. Other studies, like QRISK, have deliberately omitted individuals treated with statins [51], and some have reported measures of association with blood pressure, and most indicate increased, rather than reduced, risk of disease. For example, in the QRISK cohort, blood pressure treatment was associated with an adjusted hazard ratio of 1.85 (95% CI:1.79 to 1.91) [49]. Similarly, a Framingham analysis reported increased risk associated with high blood pressure at enrolment, comparing treated with untreated subjects at baseline [25].

In conclusion, this analysis shows an apparent increased CVD risk among

people taking antihypertensive drugs alone or combined with statins, and a modest decrease in statin treated patients. This analysis does not provide a satisfactory resolution of whether treatment should be included in risk prediction models. However, for it to be worth including a new independent covariate in a predictive regression model, to improve discrimination in a clinically significant way, the associated hazard ratio for a binary variable must generally be large, more extreme than 2 or 0.5 [48]. The adjusted association between initiation on single agents was not significantly associated with disease incidence, and initiation on to both drugs was marginally significant. So, in this cohort, this information provides little overall influence on disease risk. Taken with arguments that including drug treatment in risk prediction models may reduce their reliability [120], there is not a strong case for treatment to be included.

Despite these findings, people who are started on drugs have different characteristics than people who are not. This makes the findings of the analysis somewhat model dependent. This may be addressed by the use of propensity scores, which are considered in the next chapter.

Chapter 4

Analysis 2: Propensity scoring to estimate the association between statin initiation and CVD risk

4.1 Background

In the previous chapter, those who were initiated on drug treatment after having their risk assessed went on to have a higher incidence of disease compared to those who were not. A possible explanation for these results is that those subjects initiated on treatment had a poorer adverse prognostic profile, and that traditional regression methods did not adequately adjust for these differences. This chapter aims to address this problem using propensity score methods.

Propensity score methods are designed to address the problem of bias in causal inference when confounders do not have similar distributions in the treated and untreated groups which are being compared. The differences in

distribution is referred to as a ‘lack of balance’ [38]. The imbalance in confounders between the two groups means that the form of the regression model is relied upon more, than if the samples were balanced.

The rationale for using this method is illustrated in the following example adapted from Gelman and Hill [38]. Consider an observational study in which the true model for the relationship between treated and controls for a continuous outcome y , binary effect of treatment θ , and continuous confounding exposure x is given by:

$$\begin{aligned} \text{treated: } y_i &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \theta + \text{error}_i \\ \text{controls: } y_i &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \text{error}_i \end{aligned}$$

Assume, that the true relationship between x and the outcome y is as indicated, by the quadratic for both treated ($i = 1$) and untreated ($i = 0$) groups. After averaging values over both treatment groups, and rearranging, leads to the following expression for the treatment effect in the sample:

$$\hat{\theta} = \hat{y}_1 - \hat{y}_0 - \beta_1(\hat{x}_1 - \hat{x}_0) - \beta_2(\widehat{x_1^2} - \widehat{x_0^2})$$

If the differences between treated and untreated groups have similar levels of the confounder x , then $(\hat{x}_1 - \hat{x}_0)$ will be close to 0, so that the effect $\hat{\theta}$ will be unbiased. The magnitude of the bias is related to the size of β_1 and β_2 and the mean difference in distributions of x in the two groups. If x is balanced between the two groups, a linear model, without the quadratic term, will also give an unbiased estimate of the treatment effect θ , since $(\widehat{x_1^2} - \widehat{x_0^2})$ will be close to 0. In this way, propensity score analysis uses either weighting of observations, or matching, to balance the confounding factors among the treated and untreated groups, so making the causal effect estimate less dependent on

the structure of the model [5].

Austin [5] also argues that it is easier to assess whether a propensity score model has been correctly specified, since the distribution of covariates can be compared between the treated and untreated, compared to a usual regression approach. Similarly, he argues that the two step approach, of first checking for baseline balance in covariates, then treatment effect, more closely mimics the structure of a randomised trial.

4.2 Study design

The aim of this analysis was to determine the association between the initiation of statin treatment in a cohort of people who were untreated with this drug in the six months before undergoing their risk assessment, and the incidence of CVD. This differs from that of the previous chapter which considers both blood pressure lowering and statin treatment. This is due to the necessity to focus on a binary treatment variable, which propensity methods are centred on. In addition to the age and diagnosis (heart failure and previous CVD diagnosis) exclusions already described, patients were eliminated if they were treated with statins in the six months before undergoing risk assessment. Therefore, the selected sample is similar to that described in chapter 3, but fewer subjects were discarded, as those taking anti-hypertensive drugs at baseline assessment were retained in this sample. Matching and weighting by propensity to be treated (propensity score) were carried out to address the issue of lack of balance from potential confounders.

4.2.1 Drug exposure

Baseline statin and anti-hypertensive treatment was classified as redeeming a prescription for either therapy in the 6 month period after being enrolled in the cohort, so that other baseline variables were separated temporally from statin treatment, and so were unable to be influenced by the drug. Anti-hypertensive

drugs included: beta-blockers, calcium channel blockers, ACE inhibitors, angiotensin-2 blockers, alpha blockers and thiazide diuretics. Statins available at this time, in New Zealand, were either simvastatin or atorvastatin.

4.2.2 Statistical analysis

Statistical analysis consisted of developing a survival model for the incidence of any CVD event from 6 months after assessment. A Cox regression model with time-to-disease event from baseline enrolment was included as the outcome. Individuals were censored if they either died from other causes, or reached the end of follow-up without experiencing a CVD-related hospital admission or death. Restricted cubic splines were used to investigate the relationship between continuous variables and time-to-event as a check on the modelling assumption of linearity. Hazard ratios for continuous variables are reported by comparing the relative hazard between the risk at the 16th and 84th centiles of the variable (one standard deviation either side of the mean for a normal distribution). This allows direct comparison to relative hazards of binary variables [124].

To allow for the effects of imbalance in covariates among those treated and untreated with statins, propensity score matching was used to improve estimation of the causal effect of statin use. This involved building a logistic regression model to predict the probability of being initiated on treatment with a statin, matching each statin treated individual with an individual who was not treated. Those who were not chosen for matching were discarded. Matching was carried out by finding someone who was the 'nearest neighbour' based on the model predicted propensity (log-odds) for treatment. The covariates included in the propensity model included: age, sex, diabetes status, systolic blood pressure, total: high density lipoprotein cholesterol ratio, smoking status, and anti-hypertensive drug use.

Imbalance between statin treated and untreated samples was assessed by using standardised mean differences. Values which indicate differences which

are 20% or greater than the average standard deviation of the two groups, after matching, are generally thought to lead to indicate ‘large’ residual imbalance in covariates [99].

A Cox regression model was then used to estimate the effect of statin treatment on the ‘propensity matched’ sample. The matching was accounted for in the Cox analysis, by including separate baseline hazards for each matched pair. Variance inflation factors were checked in adjusted models to determine whether collinearity was likely to be present.

Because the matching process severely limited the sample size, a weighting by propensity score procedure to estimate the magnitude of the association between statin use and CVD [144] was used. A stabilized-inverse-probability of treatment weighting analysis was carried out, which assigns a weight to each individual based on the propensity score, which is then used in the Cox regression analysis [42].

An assumption of the Cox model is that hazard ratios apply throughout the duration of follow-up and do not change, sometimes described as the *proportional hazards assumption*. Violations of this assumption were checked by correlating the scaled Schoenfeld residuals with a suitable transformation of time [118]. If evidence of non-proportional association was found, then stratification (specifying a separate baseline hazard for the levels of the variable with non-proportional association) was used to improve model fit.

The terms included in the risk prediction model were those which were commonly incorporated into the standard Framingham model (covariates included: age, sex, total cholesterol to high density lipoprotein cholesterol ratio, systolic blood pressure, diabetes status, and current smoking status) as well as statin use.

All analyses were done using R software (version 2.14.1) [94]. The library `rms` was used for the Cox regression analysis. The ‘matching’ function from the `Matchit` library [54] of R software was used for matching on propensity scores by pairing nearest neighbours based on the linear predictor of a logistic

model derived to estimate risk of statin treatment, with or without a caliper restriction. Testing for proportional hazards was done using the `cox.zph` R function.

4.3 Results

A total of 56053 individuals were enrolled during the period of the analysis, and met the selection criteria. Table 4.1 shows the characteristics of the statin users and non-users, with the untreated shown in both raw form and propensity matched with statin users. Mean follow-up time was about one year in both groups.

Compared with the total cohort (see table 4.1), those who were classified as untreated (with statins) had a lower proportion that had been diagnosed with diabetes (9% compared to 25%) at enrolment. The mean age-at-risk assessment was 2 years older (about one fifth of a standard deviation), for those who went on to receive statins, and many more of this group also took antihypertensive drugs (31% compared to 21%). Total-to-HDL cholesterol was one unit higher in the group that went on to be treated with statins with the greatest mean differences seen in this variable.

For propensity score matching, an equal number of individuals were selected to match each person who went on to receive statin treatment. Imbalances, initially present between statin users and non-users were improved after propensity score matching, particularly for HbA1c, diagnosis of diabetes, and age at risk assessment. Some imbalance remained, however, despite matching. After matching, people who went on to statin treatment had a lower HDL-cholesterol compared with those were not treated (1.27 vs 1.36), and statin treated patients were more likely to have reported a positive family history than matched untreated patients (23% vs 15%).

After adjustment for the factors in table 4.1 in a logistic model, with the out-

Table 4.1: PREDICT cohort, by statin treatment in the six months after enrolment, with propensity matched sample and comparison of standardised mean differences in variables. Data are complete unless indicated otherwise.

Factor	No statin; (col.)	SMD (%) (before PSM)	Statin (col. %*)	No statin; propensity matched (col. %*)	SMD (%) (after PSM)	Total (col. %*)
Total	50973		5080	5080		56053
Sex						
Men	22361 (43.9)	5.2	2098 (41.3)	2071 (40.8)	1.1	24459 (43.6)
Age at risk assessment (years)						
Mean (SD)	53.7 (10.4)	20.1	55.7 (10.1)	56.0 (10.1)	2.5	53.8 (10.4)
Ethnic group						
Other	28361 (55.6)	9.9	2575 (50.6)	2564 (50.5)	0.4	30936 (55.2)
Māori	3263 (6.4)	2.3	766 (15.1)	808 (15.9)	2.3	3712 (6.6)
Pacific	8115 (15.9)	7.9	1290 (25.4)	1271 (25.0)	0.9	8881 (15.8)
Indian	11234 (22.0)	9.2	449 (8.8)	437 (8.6)	0.8	12524 (22.3)
Serum HDL-cholesterol (mmol/L) (<i>n</i> =48350)						
Median	1.38 (1.14,	26.7	1.27 (1.07,	1.36 (1.13,	22.7	1.36 (1.13,
(IQR)	1.67)		1.50)	1.65)		1.65)
Total-to-HDL-cholesterol ratio (<i>n</i> =56051)						
Median	3.8 (3.1, 4.7)	63.1	4.7 (3.9, 5.6)	4.7 (3.8, 5.7)	0.2	3.9 (3.2, 4.8)
(IQR)						
Current smoker?						
Yes	9175 (18.0)	9.6	1109 (21.8)	1119 (22.0)	0.5	10284 (18.4)
Systolic blood pressure (mmHg)						
Median	130 (120,	32.5	130 (120,	133 (120,	1.3	130 (120,140)
(IQR)	140)		148)	148)		
HbA1c (%) (<i>n</i> =19727)						
Median	5.8 (5.5, 6.3)	50.4	6.2 (5.8, 8.0)	6.1 (5.7, 7.0)	14.6	5.9 (5.5,6.5)
(IQR)						
Serum urate (mmol/L) (<i>n</i> =22 551)						
Median	0.35 (0.29,	16.3	0.37 (0.3,	0.36 (0.30,	3.3	0.35 (0.29,
(IQR)	0.41)		0.43)	0.43)		0.42)
Diagnosis of diabetes?						
Yes	4721 (9.2)	43.6	1290 (25.4)	1243 (24.4)	2.1	6011 (10.7)
Taking antihypertensive drug at baseline?						
Yes	11103 (21.8)	21.3	1582 (31.1)	1581 (31.1)	0.0	12685 (22.6)
Premature family history of ischaemic heart disease?						
Yes	7530 (14.8)	21.8	1183 (23.3)	749 (14.7)	21.9	8713 (15.5)
Follow-up time (days)						
Median	397.4 (189.4,	19.9	443.4 (271.4,	362.4 (142.4,	28.5	401.4 (198.4,
(IQR)	682.4)		724.1)	658.6)		685.4)

*Unless otherwise indicated.

SD: standard deviation

SMD: Standardised mean difference.

PSM: Propensity score matching.

Table 4.2: Crude and multivariate associations between clinical and demographic factors and statin use in the six months after assessment, derived from a logistic model ($n=56051$).

Factor	Low risk	High risk	Crude odds ratio (95% CI)	Adjusted odds ratio* (95% CI)
Gender	Men	Women	0.90 (0.85 to 0.95)	1.02 (0.95 to 1.08)
Age at assessment (years)**	43.3 ⁺	64.9 ⁺⁺	1.71 (1.58 to 1.84)	2.23 (2.04 to 2.43)
Ethnic group				
Māori	Other	Māori	1.04 (0.96 to 1.13)	0.96 (0.87 to 1.05)
Pacific	Other	Pacific	1.26 (1.18 to 1.36)	1.34 (1.24 to 1.44)
Indian	Other	Indian	1.52 (1.36 to 1.69)	1.79 (1.59 to 2.01)
Smoke cigarettes?	No	Yes	1.27 (1.19 to 1.36)	1.26 (1.16 to 1.36)
Diagnosis of diabetes?	No	Yes	3.33 (3.11 to 3.58)	3.07 (2.85 to 3.32)
Antihypertensive treatment?	No	Yes	1.62 (1.52 to 1.73)	1.19 (1.11 to 1.28)
Systolic blood pressure (mmHg)	110 ⁺	148 ⁺⁺	1.96 (1.85 to 2.07)	1.70 (1.59 to 1.81)
Total: HDL cholesterol ratio	2.9 ⁺	5.3 ⁺⁺	2.81 (2.68 to 2.95)	3.29 (3.12 to 3.48)

⁺16th centile

⁺⁺84th centile

*Adjusted for all other factors in the table.

**Restricted cubic spline used (4 df).

HDL: High density lipoprotein.

come being treatment with a statin, the strongest indicators of statin use were serum lipids (ratio of total to HDL cholesterol), being diagnosed with diabetes, systolic blood pressure and being older. Other demographic and laboratory indices had a relatively weaker association (table 4.2).

Two-way interactions between variables which were strongly associated with statin treatment were checked (total: HDL cholesterol ratio, diabetes diag-

Table 4.3: Crude and adjusted hazard ratios for model of association between statin use and CVD, with and without propensity score matching.

No propensity score matching; total sample (n=56051; 655 CVD events)				
Factor	Low risk	High risk	Crude odds ratio (95% CI)	Adjusted odds ratio* (95% CI)
Age at assessment (years)**	43.3 ⁺	64.9 ⁺⁺	3.66 (3.26 to 4.09)	4.74 (3.98 to 5.64)
Diagnosis of diabetes?	No	Yes	1.78 (1.45 to 2.18)	1.28 (1.04 to 1.58)
Systolic blood pressure (mmHg)	110 ⁺	148 ⁺⁺	2.20 (1.92 to 2.53)	1.74 (1.39 to 2.18)
Total: HDL cholesterol ratio	2.9 ⁺	5.3 ⁺⁺	1.32 (1.17 to 1.48)	1.39 (1.24 to 1.55)
Statin use at 6 months after assessment	No	Yes	1.65 (1.33 to 2.03)	1.04 (0.83 to 1.29)
Propensity matched results (n=10160; 187 CVD events)**				
Statin use at 6 months after assessment	No	Yes	0.88 (0.56 to 1.80)	0.80 (0.52 to 1.24) [†]
Stabilized-inverse-probability of treatment weights (n=56051; 655 CVD events)				
Statin use at 6 months after assessment	No	Yes	1.34 (1.05 to 1.71)	1.11 (0.86 to 1.41)

⁺16th centile

⁺⁺84th centile

*Separate baseline hazards were estimated for sex, ethnic group, and smoking status. Adjusted for all covariates shown in table

**Measures of association are adjusted for matching, by using separate baseline hazards for each matched pair.

[†]Adjusted for sex, age, smoking status, systolic blood pressure, total-to-HDL cholesterol ratio

nosis, and age at assessment), but none were significant at the 5% level when using a likelihood ratio test. Nagelkerke's R^2 for the model was 0.14, with an overall receiver-operating-characteristic curve of 0.76. The Brier score was 0.077, indicating accurate predicted frequencies of observed events.

During follow-up (median 401.4 days; interquartile range 487 days), 655 CVD events occurred, 16% (103) of which occurred among the group who used statins in the first six months after assessment.

In the raw un-matched analysis, statin use at baseline was associated with a 65% increase in risk of outcome, which reduced to no difference in hazards (HR 1.04; table 4.3) when other confounding factors were adjusted for. Due

to non-proportional hazards, sex, ethnic group and smoking status were given separate baseline hazards for all differing combinations of such variables. In the adjusted analysis, age and systolic blood pressure had the strongest association with the outcome (table 4.3).

The results of propensity score analysis varied with the technique employed. During propensity score matching, pairs were within 0.2 of a standard deviation on the logit scale for the propensity score, indicating acceptable matching [6]. The unadjusted hazard ratio for statin treatment indicated similar risk in treated and untreated groups, with the point estimate indicating lower risk among those who were treated (0.88; 95% CI: 0.56 to 1.80), although confidence intervals were wide (table 4.3). The adjusted estimate for the propensity matched sample was similar to the corresponding unadjusted estimate. Only a limited subset of variables were adjusted for, because addition of terms such as ethnic group and diabetes status lead to variance inflation due to collinearity. This problem was presumably caused by the propensity score matching procedure, as variance inflation factors only increased in the adjusted matched analysis, but not during the non-matched or propensity score weighted analysis. The numbers of subjects and events were limited to about one fifth of the original sample, and this cohort was generally higher risk (187 events out of 10160 individuals, compared to 655 events from the total sample of 56051).

Weighting the Cox regression by a transformation of the propensity score (stabilized-inverse-probability of treatment), resulted in a crude association which indicates increased risk in those who started statin treatment, compared to those who remained untreated. The adjusted measure resulted in a null association (table 4.3).

4.4 Discussion

After adjusting for confounders, those who started on statins were at a similar risk (HR=1.04) of a CVD event, compared to those who remained untreated.

The findings of the propensity score studies differed, but the unadjusted stabilized inverse probability of treatment weight analysis, which retained the total sample, showed increased risk in the statin treated, compared to the untreated group (HR: 1.34). In comparison, the adjusted propensity weighted estimate showed no difference between the two groups, whereas both adjusted and unadjusted matched estimates showed a modest reduction in risk, which was not significant.

In principle, the marginal measure of association after propensity score matching represents the value that most simulates a randomised clinical trial [5], and it was not observed in this analysis to have problems caused by collinearity. However, adjusted measures of association from the propensity score analysis are also reported (table 4.3). These were generally concordant with the crude measure. While the results of the two propensity scoring techniques (matching and weighting) at first appear contradictory, the 95% confidence intervals from the weighted analysis fit within the confidence intervals estimated from the matching procedure, even though the point estimates contrast.

Some factors were linked with starting statin treatment. The total: high density lipoprotein cholesterol ratio was strongly associated with statin initiation. The risk from this factor of going on to treatment was higher than variables associated with greater increments in absolute risk: such as age, diabetes or smoking status. This suggests that treatment decisions for statins tended to be guided by cholesterol values, rather than absolute risk calculations.

Strengths of the analysis included that a large database of subjects at risk were used, with largely complete data available. The use of routine health data obviates issues of both selection bias, and loss to follow-up, other than from external migration.

A limitation of the analysis was possible measurement error created by the use of CVD as an outcome, compared to others, which may be less error-prone, such as total mortality. Another limitation was that the analysis did not address changes in treatment during follow-up. Time-varying effects were not

adjusted for, as time-varying confounding is likely to bias this type of analysis [35], and patients were known to, in general, have a high probability of staying on treatment if they received the drug in the six month period after assessment [77].

If generally poor prognosis patients, classified in this analysis as untreated, are likely to go onto treatment during follow-up, this may lead to underestimation of the drugs' beneficial effects. The lack of information available about change in subjects' prognosis during follow-up, which were likely to influence the decision to start drug treatment during the study, precluded an analysis that included the time-varying nature of exposure. Although propensity score matching, to some extent, overcomes bias associated with selection on treatment, this threat to the study validity remains.

A further threat to these findings is the uncertainty of which criteria should be used for estimating the propensity score. Commonly measured risk factors were used, however, risk factors are not necessarily confounders, as will be discussed further in chapter 5 on page 66. In addition, the width of the adjusted confidence intervals for the propensity-matched estimate was larger than the unadjusted estimate, which was likely to relate to collinearity present in the adjusted model.

This analysis differs from most other studies which aim to predict CVD risk, as it separates temporally baseline clinical and demographic characteristics from drug treatment and CVD events. Most other analyses measure all predictor variables, including drug treatment, at the same point: study enrolment [98, 97]. In preliminary analyses, such a study design lead to analyses which suggested increased rather than reduced risk from treatment.

The association from the unadjusted, propensity-weighted, analysis differs from some of those from meta-analyses of studies in randomised trials of statins, which generally show beneficial associations between statin use and incident CVD events, in populations free of disease at baseline [18, 114, 51]. The results of meta-analyses of trials, however, vary with one study which con-

sidered the effects of statins on all-cause mortality showing no effect on this outcome [95]. It is noted, however, that meta-analyses of trial results balance measured and unmeasured confounders, unlike propensity score methods, so unmeasured confounders present in this cohort may account for the suggestion of increased risk in the weighted analysis.

In summary, the results of this chapter indicate increased to a modest decrease in risk in the statin treated group, depending on which of the various estimates are chosen. This provides some evidence that not accounting for statins in risk-prediction studies is unlikely to seriously bias risk calculations.

Chapter 5

Analysis 3: Using learning Bayesian networks

5.1 Background

A potential problem of the previous two analyses is that the structure of the model assumes that all covariates in the model, apart from treatment status, are confounders. That is, these variables are a shared common cause of the exposure and disease. Most risk prediction and causal models in epidemiology are based on additive combinations of risk factors in a regression model framework, and the additive structure implies that variables typically act, unless interaction effects are introduced, without influence on the other variables, to yield a risk of developing disease. Since they are simply mathematical constructs, the models do not necessarily provide a plausible causal representation of how disease develops.

One way to more explicitly consider causality is to attempt to describe the influence of variables on a particular disease outcome, accounting for causal pathways that are, at least, plausible, in the form of a directed acyclic graph (DAG). These can be built using learning Bayesian network algorithms [63].

The aim of this chapter is to better understand the likely causal relationships between variables which are commonly used to predict CVD risk.

Directed acyclic graphs (DAGs) encode a structure of conditional independence between variables, represented by nodes of a graph. Connections between nodes imply causal influence, observed in the data as statistical dependence. These connections are often directed, to indicate which variable influences the other (referred to as *directed edges*). In this way, DAGs represent a set of conditional dependence and independence properties associated with epidemiological variables [63].

In a DAG, no distinction is made between ‘independent’ and ‘dependent’ variables in the sense used in regression modelling. The idea underlying their use is to fuse domain knowledge with information from the collected data into a model which mimics a network of causal influences of how the observed data were generated.

DAGs are therefore useful for elucidating possible causal pathways and have been applied in epidemiology for this purpose [41]. However, they also have a role in forming sensible judgements about variables to be included in regression models. For example, a key idea of Pearl, who has been a proponent of DAG ideas, is that variables often act as ‘colliders’ [90]. That is, on a causal path between exposure and outcome, a variable on the path is entered and exited through arrowheads, which indicate more than one influence (collision of influences) on the variable (figure 5.1). Here, the terms ‘cause’ and ‘influence’ are used interchangeably to indicate directional conditional dependence, or a link between variables.

This idea of including an explicit causal understanding is absent from much statistical analysis. Including colliders as regressors in prediction models can result in unpredictable behaviour, biasing other measures of association. Pearl shows that bias may increase, by introducing dependence from unobserved or other variables, rather than reduce bias from their inclusion. Further, in certain instances, adjusting for colliders, or their ‘descendants’, that is, variables

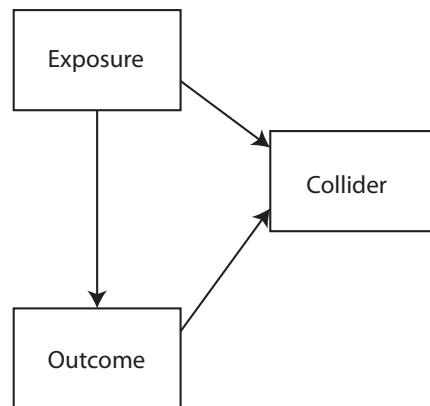


Figure 5.1: A collider variable is influenced by more than one other variable: here the exposure and outcome. The collider is also the *descendant* of both the exposure and outcome variable.

which appear to be causally influenced by colliders, may indicate no causal influence, when in fact a causal relationship does exist [90]. DAGs, derived from data, may help us identify such variables, so that they can be omitted, rather than included in regression models. To develop prediction models, a causal understanding seems more likely to lead to more accurate and reliable predictions than those developed using standard statistical methods alone [120, 64]. In this analysis, the aim was to explore a database of CVD and associated risk factors using learning Bayesian networks to inform variable selection for risk prediction models, and, it is hoped, to better explain the incidence of CVD.

5.2 Study Design

5.2.1 Sample

A group of individuals was enrolled between the 1st of Jan 2006 to the 31st of December 2006, from the PREDICT cohort, so that all subjects had at least two

years follow-up in which it would be possible for them to suffer from a CVD event, and recording of drug exposures was likely to be consistent. Two years after follow-up, it was determined if they had been admitted to hospital with CVD, or died from CVD, or other causes, by consulting hospital diagnosis and cause-of-death records. It was necessary to select a sample with known disease status at a standardised period of follow-up as the selected algorithm is not able to simply handle survival data.

5.2.2 Statistical analysis

The variables which were considered as candidates in the DAG included: age-at-enrolment, sex, diabetes, smoking, ethnic group, family history of premature CVD, statin use, antihypertensive drug use, systolic blood pressure, total: HDL-cholesterol ratio. The continuous variables, age and total: HDL-cholesterol ratio, were divided mostly into deciles. Categorical variables were used as this format is required for the particular algorithm (see below) that was selected. The outcome was fatal and non-fatal CVD.

The R package `bnlearn` drew the DAG, using the *growshrink* algorithm, first developed by Pearl [102]. Tests of independence were applied to determine conditional and marginal independence between adjacent variables, with the false-positive proportion (alpha level) set to 5%. Monte Carlo permutation tests [39] were used, because they had performed better in simulations in which the causal structure of the data were known, compared to standard chi-square tests [102].

The algorithm first learns the Markov blanket for a particular variable (X) using tests of conditional independence with other variables available in the given dataset. The *Markov blanket* is the set of variables that, once conditioned upon, render the variable independent of all other variables included in the data, but not the blanket. The blanket consists of the immediate causes of the variable (parents), those variables directly influenced by X (children) and other direct causes of variables that are directly influenced by X (spouses).

The *growing* phase tests a candidate variable Y for inclusion in the Markov blanket of X , by determining whether Y is conditionally dependent on X , given the set of already included Markov blanket variables. If X and Y are conditionally dependent, then Y is added, otherwise it is excluded. Other variables are then sequentially tested for inclusion in the blanket. Later, as the blanket grows, Y is then tested for conditional dependence with X , given a now potentially updated Markov blanket, and excluded if independent.

For a more complete and accessible illustration of the algorithm, the reader is referred to the following thesis [74].

An estimate of link influence was calculated in the final DAG, by calculating the beta-coefficient for a regression for each potential causal effect in which the variable at the base of the arrow ('cause') was considered a covariate, and the variable at the head of the arrow ('effect') was considered the outcome or dependent variable. Other variables which opened 'back door paths' (Pearl's terminology for confounding, in epidemiological terms) between the cause and effect variables were included as covariates in the regression. Either linear or logistic regression were used, depending on whether the 'effect' variable was continuous or categorical. Although age was not directly linked, in fact 'banned' (see below) as a predictor of ethnic group and sex, age was thought to be an effect modifier of ethnic group and sex in the biological, additive sense [100, 132]. This effect is not captured in the DAG, which does not explicitly account for effect modification. Hence, it was included as an (additive) covariate in estimates of the strength of links between ethnic group and sex and other variables.

The `bnlearn` algorithm allows implausible causal influences to be 'banned'. For example, for age, even if the algorithm indicates a possible cause of this variable, a link will not be drawn on the final graph. The following rules generated the banned list:

1. Sex, ethnic group and age must not be caused by any other variable.

2. Family history must not be caused by drug treatment.
3. The outcomes, fatal and nonfatal CVD, must not cause any other variable.

5.3 Results

After the age and date selection criteria were applied, 6256 subjects were available for analysis, 101 (1.6%) of whom experienced a CVD event during follow-up with 35 (0.6%) dying of causes other than CVD. From table 5.1, age-at-enrolment, ethnic group, smoking status, antihypertensive drug use, systolic blood pressure and diabetes status were significantly associated with CVD. Among ethnic groups, Māori were at highest risk of CVD (odds ratio: 1.87; 95% CI: 1.09 to 3.10). In the univariable analyses, those who used either statins or antihypertensive agents were at higher risk of CVD than non-users.

The derived DAG is depicted in figure 5.2. Directed arrows indicate the direction of causal influence between variables. Only two direct influences on CVD are detected: age-at-enrolment and cigarette smoking. Ethnic group influences risk of CVD, but it does so mediated through the effect of smoking. Age influences several other variables, such as family history of disease and the risk of taking preventive drug treatment. Ethnic group influences three variables: family history, smoking and diabetes status. The ratio of total: HDL-cholesterol concentration is influenced by two variables: sex and cigarette smoking.

There was no link between anti-hypertensive or statin therapy and CVD. Also, commonly accepted causal associations, such as systolic blood pressure and total: HDL-cholesterol ratio did not show a causal link to CVD events. This contrasts with a strong crude association between systolic blood pressure and CVD (from table 5.1). This analysis also did not causally link statins with the cholesterol ratio variable.

Indices of 'link influence' are given in table 5.2. These are beta-coefficients derived from regressing the effect (arrowhead) on the cause (tail of arrow),

Table 5.1: Sample characteristics by cardiovascular disease status (learning Bayesian network analysis): numbers (% of column sample unless otherwise stated).

Factor	CVD	No CVD	Total	Test stat.	<i>P</i> -value
Total	101	6155	6256		
Gender				Chisq. (1 df)	0.343
Men	61 (60.4)	3395 (55.16)	3456 (55.24)		
Age at enrolment				t-test	< 0.001
Mean (SD)	61.73 (10.2)	54.1 (10.45)	54.22 (10.49)		
Ethnic group				Chisq. (3 df)	0.031
Other	62 (61.4)	4348 (70.6)	4410 (70.5)		
Māori	22 (21.8)	826 (13.4)	848 (13.6)		
Pacific	16 (15.8)	773 (12.6)	789 (12.6)		
Indian	1 (1.0)	208 (3.4)	209 (3.3)		
Smoking status				Chisq. (1 df)	0.012
Yes	28 (27.7)	1082 (17.6)	1110 (17.7)		
Statin treatment at baseline?				Chisq. (1 df)	0.127
Yes	20 (19.8)	860 (14.0)	880 (14.1)		
Antihypertensive treatment at baseline?				Chisq. (1 df)	< 0.001
Yes	48 (47.5)	1637 (26.6)	1685 (26.9)		
Systolic blood pressure (mmHg)				Rank sum test	< 0.001
Median (IQR)	140 (130, 150)	130 (120, 142)	130 (120, 143)		
Diagnosis of diabetes?				Chisq. (1 df)	0.0143
Yes	24 (23.8)	896 (14.6)	920 (14.7)		
Total to HDL-cholesterol ratio				Rank sum test	0.744
Median (IQR)	3.71 (3.1, 4.8)	3.8 (3.1, 4.7)	3.8 (3.1, 4.7)		
Premature Family history?				Rank sum test	0.450
Yes	31 (30.7)	1681 (27.3)	1712 (27.4)		
Other death?				Chisq. (1 df)	0.930
Yes	0 (0.0)	35 (0.6)	35 (0.6)		

IQR: Interquartile range.

HDL: high density lipoprotein.

Stat: statistic.

Chisq.: chi-square test of independence.

SD: standard deviation

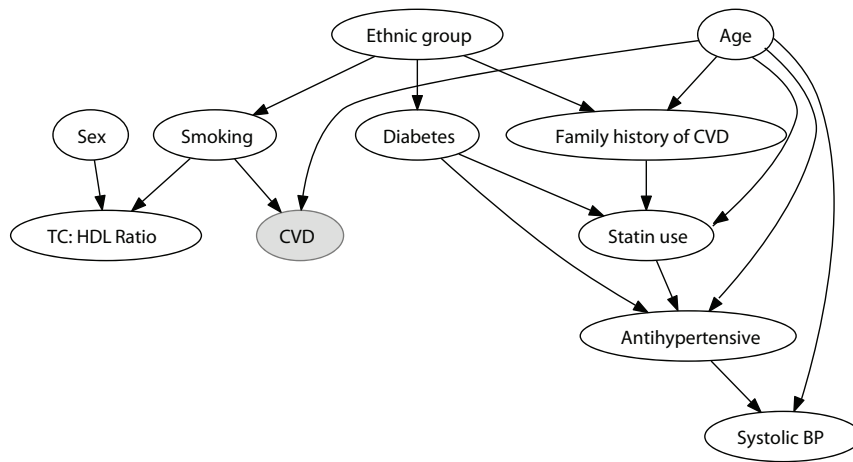


Figure 5.2: DAG, derived from the *grow-shrink* algorithm. The grey box indicates the outcome variable. CVD: Cardiovascular disease. HDL: high-density lipoprotein cholesterol concentration. TC: total cholesterol concentration. BP: blood pressure.

using either linear or logistic regression, adjusting for other immediately adjacent influences on the effect variable. All links between age and other variables show strong evidence of association, along with ethnic group, male sex and diabetes and their causal links. Strong associations were noted between diabetes status and the use of preventive drugs.

From the logistic regression analyses, the greatest odds ratios were between ethnic group and diabetes status. Pacific people were 4.4 times more likely than ‘Others’ to be diagnosed with diabetes (estimated OR: 6.44, 95% CI: 5.39, 7.70; prevalence of diabetes among ‘Others’: 8.6%) and Indian people were almost four times more likely than ‘Others’ to have the diagnosis in this cohort (estimated OR: 5.14, 95% CI: 3.78 to 7.00). For continuous outcome measures, those who used anti-hypertensive drugs had an average systolic blood pressure 7.30 mmHg (95% CI: 6.28 to 8.33) higher than people who did not use these drugs.

5.4 Discussion

This analysis showed that a DAG learning algorithm generated a plausible graph explaining the occurrence of CVD. The DAG captured some known re-

Table 5.2: Lists of causal links and estimated beta-coefficients from linear or logistic regression, adjusted for variables, as indicated in the DAG. Also odds ratios for logistic regression models.

Cause	Effect	Low ⁺	High ⁺	Beta-coeff.(95% CI)	Odds ratio (95% CI)
Age (no adj.)	Other death	43.4	65.2	1.25 (0.54 to 1.95)	3.49 (1.72 to 7.06)
Age	CVD	43.4	65.2	1.54 (1.11 to 1.97)	4.65 (3.03 to 7.14)
Age	Statin use	43.4	65.2	0.84 (0.69 to 0.99)	2.31 (1.99 to 2.69)
Age	Anti-hypertensive	43.4	65.2	1.44 (1.31 to 1.57)	4.23 (3.72 to 4.82)
Age	Family history of CVD	43.4	65.2	-0.31 (-0.43 to -0.20)	0.73 (0.65 to 0.82)
Age	Systolic blood pressure	43.4	65.2	10.42 (9.5 to 11.34)	N/A
Age	Sex (men)	43.4	65.2	-0.73 (-0.83 to -0.62)	0.48 (0.43 to 0.54)
Ethnic group (no adj.)	Diabetes	Other	Indian	1.64 (1.33 to 1.95)	5.14 (3.78 to 7.00)
		Other	Māori	1.03 (0.84 to 1.23)	2.81 (2.31 to 3.42)
		Other	Pacific	1.86 (1.68 to 2.04)	6.44 (5.39 to 7.70)
Ethnic group (no adj.)	Smoker	Other	Indian	-0.50 (-0.99 to -0.01)	0.60 (0.37 to 0.99)
		Other	Māori	1.28 (1.12 to 1.45)	3.60 (3.05 to 4.25)
		Other	Pacific	0.72 (0.54 to 0.91)	2.06 (1.72 to 2.48)
Ethnic group (adj. for age)	Family history of CVD	Other	Indian	0.02 (-0.28 to 0.32)	1.02 (0.75 to 1.37)
		Other	Māori	-0.24 (-0.41 to -0.07)	0.79 (0.67 to 0.93)
		Other	Pacific	-1.03 (-1.24 to -0.82)	0.36 (0.29 to 0.44)
Diabetes (adj. for age)	Statin use	No	Yes	1.94 (1.77 to 2.10)	6.94 (5.90 to 8.16)
Diabetes (adj. for age)	Antihypertensive use	No	Yes	1.68 (1.53 to 1.84)	5.38 (4.60 to 6.28)
Diabetes (adj. for age)	Other death	No	Yes	1.23 (0.54 to 1.91)	3.42 (1.72 to 6.79)
Statin use (adj. for age)	Antihypertensive use	No	Yes	1.70 (1.55 to 1.86)	5.49 (4.69 to 6.42)
Anti-hypertensive (adj. for age)	Systolic blood pressure	No	Yes	7.30 (6.28 to 8.33)	N/A
Smoker (no adj.)	CVD	No	Yes	0.59 (0.15 to 1.03)	1.80 (1.16 to 2.79)
Smoker (no adj.)	Other death	No	Yes	1.02 (0.33 to 1.70)	2.76 (1.39 to 5.50)
Smoker (no adj.)	Total: HDL-cholesterol ratio	No	Yes	0.51 (0.43 to 0.59)	N/A
Family history (adj. for ethnic group and age)	Statin Use	No	Yes	0.42 (0.26 to 0.58)	1.52 (1.30 to 1.79)
Sex (no adj.)	Total: HDL-cholesterol ratio	Female	Male	0.61 (0.55 to 0.67)	N/A

CVD: Cardiovascular disease.
HDL: high density lipoprotein.
adj.: adjustment.

N/A: not applicable.

⁺: for age, comparisons were made at the 84th and 16th centiles.

This allowed comparison with measures of effects from binary variables.

relationships, such as the influence of age and smoking on CVD. Age influenced other variables such as systolic blood pressure, preventive drug use and family history. Positive or higher values of these variables increased in probability with advancing age, except family history of premature CVD which is probably under-reported by older people.

The DAG may help inform variable selection when assessing the influence of ethnic group on CVD. In the DAG produced here, variables such as diabetes, cigarette smoking, and family history of premature CVD were 'caused' by ethnic group. These 'causal' relationships indicate that in trying to assess the effects of ethnic group on CVD, adjusting for any of these effects of ethnic group will likely underestimate this term's causal effect. Since it is only plausible that these variables are influenced by ethnic group, rather than the converse, adjusting for smoking status will result in an underestimate of the effect of ethnic group on risk of CVD event. In contrast, if one were trying, instead, to *predict* disease risk for an individual of a particular ethnic group, the distinction between confounder and mediator is not as important, so that including smoking status, a cause of CVD, in the risk equation may be logical if it improved prediction.

The DAG presented here also may help identify variables which Pearl terms *barren proxies* when conducting causal analyses. For example, consider a scenario in which one was to investigate the statistical evidence for a causal link between sex and CVD. In this case, including the cholesterol ratio variable as a covariate, which, in this dataset, is caused by sex, but does not show convincing evidence of influencing disease status, would increase (rather than reduce) bias in estimating the strength of association between sex and CVD. The value of excluding the cholesterol ratio in a causal analysis is distinct from the value which the variable may play in predicting disease onset.

If, on the other hand, predicting CVD was the aim, and the term 'statin use' was included, from this DAG, many influences on this variable were detected, such as diabetes, age, family history of disease. With so many observed influ-

ences on such variables, many more unobserved influences are likely to exist. So, when adjusting for statin use in a regression model which aims to predict, collider bias may be introduced from such unobserved variables.

If the DAG is a possible representation of causality, then assessing the influence of potential confounding factors on CVD incidence is simplified considerably. For example, for ethnic group, none of the variables shown in the DAG confounds the relationship between this variable and CVD. Thus, in assessing the causal effect of ethnic group, it may only be necessary to adjust for age. Again, if this DAG is valid, it would suggest that very few variables actually cause CVD, so in assessing the effect of various exposures, some adjustment may cause more harm than good. DAG considerations also counter the common practice in clinical research of reporting ‘independent risk factors’, after adjusting for a number of other variables [11]. From a DAG perspective, this approach assumes that all variables are not just associated with disease, but actually cause or are an antecedent to a cause of the disease in question, and similarly influence exposure (unless blocked by a collider). This approach allows variables to have differing causal relationships between each other variable, rather than assuming they only influence the outcome, as in the use of traditional regression methods.

Some unexpected links emerged from the analysis, such as the link between cigarette smoking and serum lipids. After reviewing publications about the topic, the statistical link between cigarette smoking and low HDL-cholesterol levels had been long described [103, 36].

These analyses are not intended to give the definitive view of the causes of CVD, since the dataset has limitations. These include type-2 errors (there were only 101 CVD events), information bias and unmeasured confounding. Rather, the data used in these analyses demonstrate how DAGs can provide an alternative view of the relationships between variables which may not be appreciated from traditional statistical methods.

Few other studies have used learning Bayesian networks to explore similar

datasets. The closest is Twardy [131] who used Bayesian network algorithms, based on minimisation of information metrics, to determine the causal structure of the data in two cohort studies of CVD. The authors did not exclude implausible relationships, as in this chapter. Also, their study was limited by a high proportion of cases in which some covariates were missing. In their 'final' model, several implausible relationships were present, such as diabetes and weight influencing age. Their model described age as the only influence on CHD and had some similar findings to this chapter, of age influencing many risk factors: total cholesterol, triglycerides, systolic blood pressure, smoking status and height. Unlike this analysis, causal links were drawn between diabetes and systolic blood pressure. This is a likely link not drawn in the DAG returned here.

In the future, it would be useful to apply this algorithm to a study which has collected similar information, in terms of outcomes and exposures. If the algorithm were to draw a similar DAG, the validity of the inferences from this chapter would be enhanced.

To summarise the implications of this analysis for statistical modelling, these results suggest that when assessing the causal influence of an exposure on CVD, at least one analysis should estimate the conditional association between disease status, only adjusted for age and the variables showing a causal association in the DAG. If other variables are included in the regression model, researchers must think carefully about whether they are likely to act as confounders, that is, causally influence both the exposure of interest and the outcome, rather than act as 'barren proxies'. The Bradford-Hill criteria [55] may be used to guide such decisions.

For prediction, many variables used in statistical models may be associated, but not causally related to disease. Variables with many causes (here, drug treatment) are likely to provide unreliable information, if further unobserved causes are present and are not included in analyses. If environmental conditions change, along with the nature of statistical associations with disease

status, then the model is likely to become unreliable. This suggests the need to update prediction models, along with prioritising the inclusion of variables which have strong evidence for causal influence.

Conclusions

The derived graph provides useful information to aid variable selection when assessing causal relationships with disease, and since they are related concepts [120], the DAG is also useful in the development of models used for prediction.

Chapter 6

Analysis 4: The clinical utility of using drug treatment variables to predict CVD

6.1 Background

In previous chapters, the strength of association between drug treatment and CVD has been estimated, conditional on other risk factors. The association between drug use or initiation and the disease is small, and generally shows increased rather than reduced risk. Other authors of risk prediction studies have reported positive associations between preventive drug use and CVD events. Some popular risk equations, such as those based on the Framingham study, do not include treatment [1]. The aim of this chapter is to assess the utility of drug use at baseline through determining if this extra data renders a model more 'clinically useful' than a simpler model that excludes drug use.

Aside from the strength of the association between the term and the outcome, it seems logical to use a metric of clinical utility to evaluate the relative merits of differing models, with or without drug use. This contrasts to the rel-

actively widespread use of primarily statistical indices used to evaluate these models [53].

One approach which requires little more information than that available from an epidemiological dataset uses *net-benefit* as the index by which the clinical usefulness of a model may be judged. This approach weighs the benefits of treating people with disease (true-positives), adjudicated by a positive ‘test’ from the model and disease developing during follow-up, against the harms of treating false-positives. A positive test is derived from deciding on a cut-point, above which model-based predictions are considered to indicate disease, or at least the need for treatment, and a negative test, which indicates no disease, and thus no need for treatment.

The model is judged by its ability to classify cases and non-cases at varying cut-points, based on model predictions of disease. The cut-point is considered to be that of clinical equipoise: the point at which the benefits of treating cases equal the harm from treating non-cases. Since there is uncertainty about this point, the threshold for treatment is varied and model performance assessed at differing thresholds. This performance is then contrasted with default alternative policies of ‘treat-all’ or ‘treat-none’.

Here, the performance of a model with commonly recorded risk factors for CVD was compared with a model which also included baseline preventive treatment status: either statins or anti-hypertensives or both. The ‘clinical utility’ of these models was compared to assess whether drug information usefully improves disease classification.

6.2 Study design

6.2.1 Sample

This chapter compares model predicted with observed outcomes, people with at least two years follow-up were selected. Subjects enrolled between the 1st of Jan 2006 to the 31st of December 2006 were chosen (resulting in the same

cohort described in chapter 5). Two years after follow-up, it was determined if they had been admitted to hospital with CVD or died from CVD or other causes by consulting hospital diagnosis and cause of death information.

6.2.2 Statistical analysis

Analysis centred on building a multivariable Cox regression model for CVD events, with commonly measured indices of risk included as covariates (age, sex, ethnic group, diabetes and smoking status, total: HDL-cholesterol ratio, and systolic blood pressure). Model comparison was based on the method described by Vickers and Elkin, comparing the performance of classifying individuals using the model with policies of 'treat all' or 'treat none' [133]. This analysis assumes that model predictions assign treatment to individuals based on cut-points or thresholds of risk. If it is further assumed that at the extremes of risk patients will opt to not be treated (low-risk) or to be treated (high risk), then a threshold is reached in which the benefits of treatment of the true-positives are thought to equal the risks of treating the false-positives. In New Zealand, at the time of writing, for example, the threshold cut point is 10% risk 5 year risk [116]. This threshold was not determined by considering clinical utility, but was instead a result of drug rationing. Economic constraints are ignored in this analysis.

This threshold, expressed as an odds $\frac{1-p_t}{p_t}$, from a *net-benefit* point of view, is the ratio of the harm associated with a false-negative result (compared to a true-positive), to the harm from being treated unnecessarily due to a false positive result (compared to a true-negative). A model is then used to assign treatment based on the predicted probability of disease occurring during follow up. Those with a predicted probability above the threshold p_t are placed on treatment, while those below are not. A high p_t close to 1 carries the assumption of little benefit from treatment, resulting in a high threshold for treatment, and a low p_t near 0 conversely carries the assumption of little or no adverse effects from treatment.

The *net-benefit* is, then, a weighted difference of the proportion of true-positives identified by the model, using p_t to assign treatment, and the proportion of false-positives, as follows:

$$\text{Net benefit} = \frac{\text{true-positive count}}{n} - \frac{\text{false-positive count}}{n} \left(\frac{p_t}{1 - p_t} \right)$$

Where the true- and false-positive counts are the number of patients categorised by the model accordingly, with n the total number of subjects in the analysis, and p_t is the threshold probability (where the benefits of treating those with disease are thought to equal the harms of treating those without disease).

Under the assumptions of this analysis, if p_t equals zero, this scenario represents both a *treat-all* scenario (all individuals will have a predicted risk > 0), so that no harm occurs to those who are false-positives at the end of follow-up. If p_t is at the other extreme, 1, then this means that no subjects will be treated, and that no benefit from treatment is conferred to those who are true-positives at the end of follow-up.

Net-benefit is calculated at a range of clinically plausible treatment thresholds p_t as follows:

1. Choose a value for p_t .
2. Calculate the number of true and false-positives resulting from categorising patients according to the rule in (1).
3. Calculate the *net-benefit* from the application of the rule.
4. Vary p_t over an appropriate range and repeat steps (2) and (3).
5. Plot the results of the first four steps on a graph with p_t represented on the x -axis and *net-benefit* on the y -axis.
6. Repeat the first 5 steps for each model being considered as clinically useful.

7. Repeat the first 5 steps under the assumption of *treat all* (all patients are considered positive).
8. Draw a straight line parallel to the x -axis at $y=0$, which represents the strategy of *treat none*.

These values may then be compared with default policies of *treat-all* or *treat-none*, which do not require a model to guide treatment decisions. The main comparisons in this chapter are the differences between performance of the model with and without drug treatment status (statin use, blood pressure lowering drug use, or both), as well as the range of probability thresholds p_t in which use of the model was likely to be clinically useful. That is, use of the model resulted in an increase in *net-benefit* over *treat-all* or *treat-none* policies.

Once the *net-benefit* had been calculated, a threshold was chosen where the model was useful. The classification of a model with commonly measured risk factors, with or without drug treatment, was compared using *net reclassification index* (NRI) at that chosen threshold [92]. This index compares the number of people, appropriately reclassified as treatment positive or negative, comparing a candidate with an established model. The formula for this metric of comparative model utility is given by:

$$\begin{aligned} \text{NRI} = & P(\text{change from low to high risk} | \text{disease}) - \\ & P(\text{change from high to low risk} | \text{disease}) + \\ & P(\text{change from high to low risk} | \text{no disease}) - \\ & P(\text{change from low to high risk} | \text{no disease}) \end{aligned}$$

Where NRI represents ‘net reclassification improvement’, $P(\text{change from low to high risk} | \text{disease})$ is the probability of changing to a higher risk category (in the new model, whereas the old model classified the subject as low risk), given the subject has disease, and so on. R software (version 2.15), including the `rms` and `epicalc` packages, was used for statistical analysis [94, 45, 19]. The depiction of the pre-

dictive accuracy of the models with and without treatment status was carried out using SPAN software [75].

6.3 Results

After the selection criteria were applied, 6256 subjects were available, 101 (1.6%) of whom experienced a CVD event during follow-up, and 35 (0.6%) of whom died of causes other than CVD. This was the same sample studied in chapter 5. From table 6.1, age-at-enrolment, ethnic group, smoking status, antihypertensive drug use, systolic blood pressure and diabetes status were significantly associated with event status. People who developed CVD were an average of 7.6 years older than those who remained disease free. Among ethnic groups, Māori were at highest risk of a CVD event (odds ratio: 1.87; 95% CI: 1.09 to 3.10).

The results of the Cox models are presented in table 6.2. Age-at-assessment had the strongest association with disease in both the crude and adjusted analyses. People who smoked at baseline were about twice as likely to experience a CVD event during follow-up compared to non-smokers in both crude and adjusted models. The association between systolic blood pressure and CVD was progressively attenuated with increasing levels of adjustment, with a two standard deviation change resulting in a 26 to 32% relative increase in hazard in the two multivariable models in which this term was included.

Total: HDL-cholesterol showed little association with CVD in both crude and adjusted models. Treatment with blood pressure lowering drugs, either alone, or in combination with statins, was strongly associated with CVD, however, these relationships were reduced after adjustment for commonly measured risk factors. The association between diabetes and CVD was attenuated by the inclusion of drug treatment, indicating that some of the excess risk from drug treatment may be due to the complications and excess risk of diabetes which is, itself, strongly associated with preventive drug treatment. Very lit-

Table 6.1: Sample characteristics by CVD status: count (% of column sample unless otherwise stated).

	CVD	No CVD	Total	Test stat.	P-value
Total	101	6155	6256		
Gender				Chisq. (1 df)	0.343
Men	61 (60.4)	3395 (55.16)	3456 (55.24)		
Age at enrolment mean(SD)	61.73 (10.2)	54.10 (10.45)	54.22 (10.49)	t-test	<0.001
Ethnic group				Fisher's exact test (3 df)	0.036
Other	62 (61.4)	4348 (70.6)	4410 (70.5)		
Māori	22 (21.8)	826 (13.4)	848 (13.6)		
Pacific	16 (15.8)	773 (12.6)	789 (12.6)		
Indian	1 (1.0)	208 (3.4)	209 (3.3)		
Smoking status				Chisq. (1 df)	0.012
Yes	28 (27.7)	1082 (17.6)	1110 (17.7)		
Statin treatment at baseline?				Chisq. (1 df)	0.127
Yes	20 (19.8)	860 (14.0)	880 (14.1)		
Antihypertensive treatment at baseline?				Chisq. (1 df)	< 0.001
Yes	48 (47.5)	1637 (26.6)	1685 (26.9)		
Systolic blood pressure (mmHg)				Rank sum test	< 0.001
Median (IQR)	140 (130, 150)	130 (120, 142)	130 (120, 143)		
Diagnosis of diabetes?				Chisq. (1 df)	0.0143
Yes	24 (23.8)	896 (14.6)	920 (14.7)		
Total to HDL-cholesterol ratio				Rank sum test	0.744
Median (IQR)	3.7 (3.1, 4.8)	3.8 (3.1, 4.7)	3.8 (3.1, 4.7)		
Premature Family history?				Rank sum test	0.450
Yes	31 (30.7)	1681 (27.3)	1712 (27.4)		
Other death?				Chisq. (1 df)	0.930
Yes 0 (0.0)	35 (0.6)	35 (0.6)			

CVD: Cardiovascular disease.
HDL: high density lipoprotein.
IQR: Interquartile range.
Stat: statistic.
df: degrees of freedom.
Chisq.: chi-square test of independence.

Table 6.2: Crude and adjusted hazard ratios derived from Cox models of CVD events, which included commonly measured risk factors as covariates with and without drug treatment variables.

Characteristic	Low risk	High risk	Crude Odds ratio (95% CI)	Adj. hazard ratio (95% CI)†	Adj.* hazard ratio (95% CI)	Adj.# hazard ratio (95% CI)
Age at assessment (years)	43.4+	65.2++	4.60 (3.02 to 7.01)	5.21 (3.39 to 7.99)	5.50 (3.49 to 8.67)	5.11 (3.21 to 8.12)
Sex	Women	Men	1.24 (0.83 to 1.84)		1.56 (1.04 to 2.34)	1.59 (1.05 to 2.39)
Ethnic group	Other	Indian	0.66 (0.25 to 1.80)		0.49 (0.07 to 3.60)	0.48 (0.07 to 3.49)
	Other	Māori	1.80 (1.27 to 2.56)		2.07 (1.24 to 3.45)	1.99 (1.19 to 3.33)
	Other	Pacific	1.54 (1.05 to 2.25)		1.62 (0.90 to 2.90)	1.57 (0.87 to 2.81)
Smoking?	No	Yes	1.79 (1.16 to 2.76)	2.42 (1.55 to 3.77)	2.01 (1.28 to 3.17)	2.00 (1.27 to 3.15)
Diabetes?	No	Yes	1.82 (1.15 to 2.89)		1.29 (0.79 to 2.10)	1.16 (0.69 to 1.95)
Systolic Blood Pressure (mmHg)	116+	150++	1.87 (1.37 to 2.56)		1.32 (0.94 to 1.86)	1.26 (0.89 to 1.79)
Total to HDL cholesterol ratio	2.80+	5.13++	1.03 (0.73 to 1.45)		1.12 (0.78 to 1.61)	1.15 (0.80 to 1.66)
Statin treatment only	None	Statin use	1.32 (0.41 to 4.22)			1.63 (0.51 to 5.27)
Blood pressure treatment only	None	BP drug use	2.33 (1.49 to 3.65)			1.39 (0.86 to 2.24)
Both drugs	None	Both drugs	2.67 (1.54 to 4.63)			1.48 (0.80 to 2.73)
<i>c</i> -index				0.7286	0.7596	0.7635

c-index: Harrell's concordance statistic.

HDL: high density lipoprotein. Adj.: adjusted.

†Adjusted for age at assessment and smoking status.

*Adjusted for age at assessment, sex, ethnic group, smoking, diabetes, systolic blood pressure and total: high density lipoprotein cholesterol concentration.

#Adjusted for age at assessment, sex, ethnic group, smoking, diabetes, systolic blood pressure and total: high density lipoprotein cholesterol concentration

and both statin and antihypertensive drug use.

+Measured at the 16th centile. ++Measured at the 84th centile.

the difference in the discrimination indices (*c*-statistic) were noted between the models which included standard risk factors, compared to that which also included drug information. A simplified model, which only included age and cigarette smoking status had a *c*-statistic about 0.03 lower than the more complex models.

The difference between the three models was investigated by plotting *net-benefit* at different (p_t) thresholds used to guide treatment (figure 6.1). The three models included one with commonly measured risk factors without drugs, and one with this variable included. The third was simplified, and only included age and smoking status, and was included as a reference. The beta-coefficients and the *c*-statistic of the model were similar, comparing the model which added drug treatment to the commonly measured variables, to that which excluded the drug information.

The figure 6.1 shows the change in *net-benefit* from the use of a Cox model with differing levels of covariate information included. As expected, the maximum *net-benefit* is at the point at which the predicted probability threshold p_t is 0. This point on the figure corresponds to an assumption that the drug confers only benefit to those who become cases at the end of follow-up, with no harm incurred to false-positives. Similarly, the threshold probability (p_t) of about 3 times the cumulative incidence of disease (5%) is associated with almost no treatment, since almost all subjects are below this risk level. This point on the plot corresponds to a minimum *net-benefit* (0) as no subjects are treated. By comparing the plot of the *net-benefit* derived from the model with a policy of *treat-all* (dashed line, that is a line with slope -0.958 and coinciding with maximum *net-benefit* of 0.016), and *treat-none* (*net-benefit*=0), the model only usefully informs clinical decision making (resulting in positive *net-benefit*) between the probability thresholds of 2 and 5% 2-year risk of CVD.

The *treat-all* scenario assumes that all cases and non-cases are treated, with a roughly linear decrease observed in *net-benefit* as the ratio of harms to benefit increases. This negative slope corresponds to increasing harm from treating

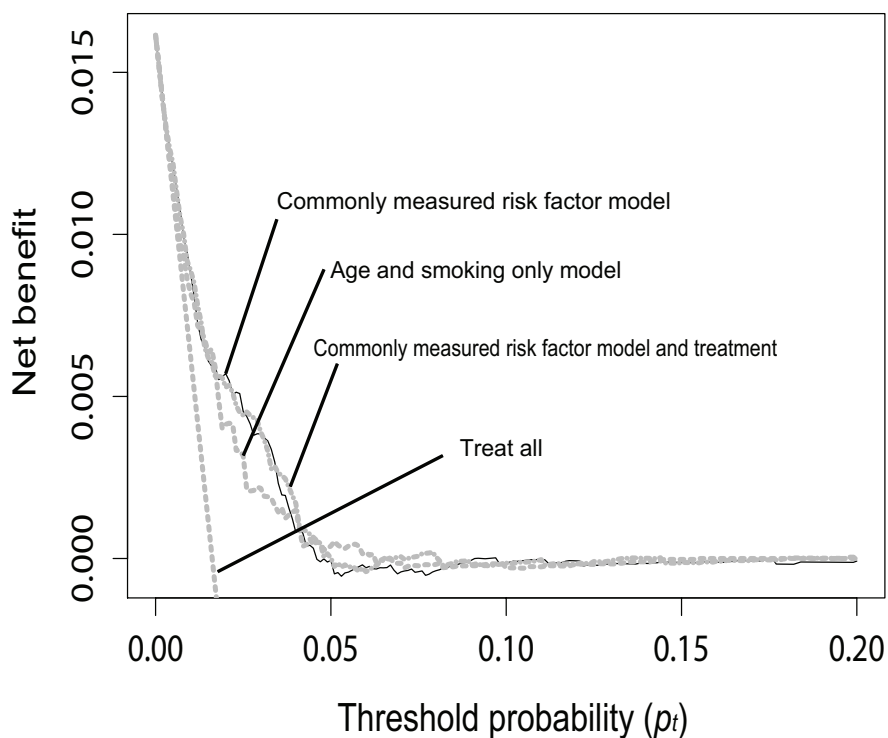


Figure 6.1: *Net-benefit* associated with the use of different CVD prediction models for varying threshold probabilities (p_t) for 2-year CVD risk.

all the non-cases. The figure 6.2 also illustrates that the full model only outperforms a simplified model (that includes only age and smoking status as predictors) between 2 and 4 percent 2-year risk.

Also, comparing the benefit of a model with common risk factors, with or without preventive drug treatment added, there is little or no overall *net-benefit* gain from including the extra treatment information. The difference between the two indices is less than a *net-benefit* of 0.0001, or 1 more true positives identified, for every 10,000 weighted false positives, or 0.67% of the maximum possible *net-benefit* (0.015).

The incremental change of adding statin or blood pressure lowering drug

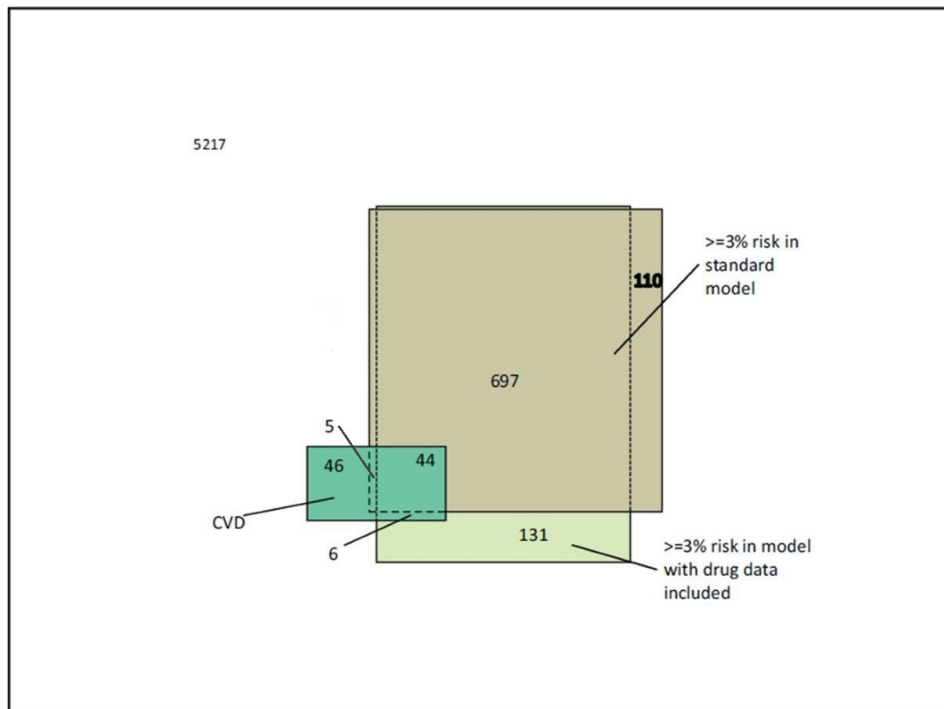


Figure 6.2: Scaled rectangle diagram [75] depicting model predicted high risk patients ($\geq 3\%$ 2 year risk) and comparing them with their disease status at 2 years follow-up. One model had standard risk factors, and the other had these variables as well as preventive drug use data at enrolment.

treatment status to a model with commonly measured risk variables was considered, to classification of treatment status at a threshold, which was indicated in the *net-benefit* analysis to be clinically useful (3% 2-year risk), compared to ‘treating all’ or ‘treating none’. The discrimination of the models at this threshold are depicted in a scaled rectangle diagram [75] (figure 6.2) which contrasts model prediction (high risk) and disease status at the end of follow-up. The outer rectangle represents the total sample analysed in this chapter, the smallest rectangle is the cases of CVD at 2-years follow-up, and the medium sized rectangles represent the two model predicted high risk groups, at the 3% two year threshold.

Of the 101 CVD cases, 11 were re-classified: 6 correctly upward with the treatment model, and 5 downward. Of the 6155 without disease at the end

of follow-up, 110 were correctly downward classified by the treatment model, whereas 131 were incorrectly upwardly reclassified. At this 3% threshold, the net reclassification improvement resulted in a modest improvement for the model which included drug treatment (0.65%; 95% CI: -5.8% to 7.1%); however, this figure was not significant.

6.4 Discussion

In this chapter, a model for CVD risk prediction had a clinical utility between the two year risk thresholds of two and five per cent, compared with other policies which do not require a model: *treat all* or *treat none*. Little difference was found between models which included commonly measured risk factors, with or without information about subjects' blood pressure lowering or statin treatment status at study enrolment.

A strength of this analysis is that it explicitly considers the likely clinical benefit of a model. That is, this method weighs the benefits conferred to those who go on to get disease against the harms from side effects from treatment of people who would remain disease free with or without treatment.

A limitation of this analysis is that it did not consider the problem of overfitting [44], which may lead to poorer performance of this model if it were applied to independently collected data. Since all models investigated had relatively few parameters, compared to the number of outcomes and subjects, this issue is not likely to pose a serious threat to the study validity.

In addition, this method assumes that the model is derived from a cohort study of the natural history of disease, however, in this cohort, almost 30% were treated with either a statin or antihypertensive drug or both, when assessed at baseline. This analysis, therefore, assumes that the treatment assigned will bring benefits not observed in these data. From the analysis carried out in chapter 7 on page 94, and from the overall picture, it is unlikely that this is a serious threat to the validity of this analysis, as the effects of drugs in this cohort

are likely to be small.

Compared to other methods, *net-benefit* does not consider the costs of negative tests, either true-negatives or false-negatives. This is justified in that subjects who are test negative (not treated) can not receive either benefit or harm from the drug, since they are not exposed to it. It may be argued that a false-negative may not undertake lifestyle change that would be otherwise beneficial, if they thought they were at high risk of disease. This analysis only considers the relative harms and benefits of those who would be treated with drugs.

Another limitation of the analysis is the short follow-up time, so that predictions are relatively short-term, compared to other major CVD risk equations [51, 25]. This also limited the statistical power of the study, due to small number of disease events. However, it could be envisaged that the clinical management of CVD risk entails patients being given short term risk prediction (two year), with risk status reassessed at two yearly intervals.

The next question which follows is what is the likely risk-benefit probability of treatment threshold (p_t)? Also, does this threshold correspond to the region from which this analysis indicates that the model is clinically useful? National CVD guidelines may inform where this threshold has been estimated by large scientific bodies. When considering an appropriate threshold, simplifying assumptions are made, that those who get disease receive only benefit from the drug, whereas those who remain free of disease receive only side effects.

In the United States, the equivalent of a 2% 2-year risk of CVD is recommended as the threshold for starting statins, assuming that the hazard of event over 10 years is constant, although the authors of the guidelines admit the cost-effectiveness of this approach is marginal at this threshold, and above the threshold of 4% 2-year risk (20% 10-year risk), treatment is more easily justified [20]. In the U. S., people judged to be at less than 2% 2-year risk are generally considered too low risk to treat with drugs. In New Zealand, at the time of writing, predicted risk greater than 10% after 5-years is considered the thresh-

old for treatment [116], which corresponds roughly to a 4% two year risk, if the hazard of disease throughout study follow-up is assumed constant. If treatment thresholds are set where the benefit of treating those who get disease equal the harm from treating those who remain disease free, these guidelines carry the implicit assumption that failing to give a patient who ultimately gets disease preventive drug treatment is 50 (reciprocal of 2%; $1/0.02$) to 24 times ($1/0.04$) worse than unnecessarily treating an individual with drug therapy who does not go on to get disease, throughout the 2 years of follow-up.

No document, to the author's knowledge, explicitly addresses the issue of the risks and benefits of statin and antihypertensive drug treatment, among those who would become cases and those who remain disease free, from which a sensible value for p_t could be derived. If, on the basis of trial data, it is assumed that the relative reduction in CVD event rates for true-positives from combined treatment with statins and anti-hypertensive drugs is about 30% relative reduction [115] (this assumption is questioned in chapter 7), and that the relative harm from treatment for false-positives is a relative increase in serious adverse events of 0.1% (rhabdomyolysis) over three years follow-up [105], then the cut-point p_t of 0.3% ($0.001/0.3$) may be justified. From this analysis, however, the use of a model could not be justified at this cut-point, since it would be more sensible to *treat all*. Budgetary constraints are ignored here.

If instead, it is accepted that more minor, but still clinically important, adverse events from statin use score as harm, then a higher threshold may be justified. These include, for statins, myalgias and liver enzyme rises which occur at an increased incidence of about 5% over a median follow-up of three years [105], so a more modest probability threshold may be used of 16.7% ($5\%/30\%$). At this threshold, however, a model could not be justified, since the *net-benefit* from using the model would be similar to a policy of *treat none*. If, instead, one considered that the harm to benefit ratio was somewhere in between these two extreme figures, then a 3% two year threshold (p_t) may be justified.

6.4.1 Conclusion

This is the first known use of this method to examine the potential benefits of a risk prediction model to guide drug treatment for CVD. Information about drug use at enrolment in the study did not significantly improve the proportion of individuals who were correctly classified into a high risk group by the model.

Chapter 7

Analysis 5: Publication bias in studies of statin use to prevent CVD

7.1 Background

The analyses which have been conducted so far have focused on the use of observational data to assess the strength of association between statin and blood pressure lowering drugs and incident CVD, in populations free of diagnosed disease when enrolled. An argument is often put forward that data from observational studies should be discarded if randomised trial results are available. A potential problem, however, with randomised controlled trials is that of publication bias. This is underlined, in the case of statins, with the report that a large phase 3, primary prevention, statin trial undertaken in five countries was stopped prematurely. The sponsoring drug company justified this by stating that the trial did not fit with the company's marketing interests [68].

For the use of statins, for example, almost all studies of these drugs are supported by drug company funding. This situation makes the presence of pub-

lication bias more likely [9]. Despite this threat to the validity of information from industry funded studies, statins are usually justified by meta-analyses of randomised trials. Of the most highly cited study summaries for the use of statins, two support their use [114, 12], while another does not [95]. An assumption of meta-analyses and their summary measures of effect is that the constituent studies are an unbiased sample of potential trials. However, when this is not the case, publication bias may occur. Standard assessments of publication bias, such as the Egger test and inspection of funnel plots, have been developed [31, 113].

For statins, these tests do not appear to show any publication bias in meta-analyses that have been reported [114, 12, 95]. However, the tests have limitations; they are only suitable when the number of trials in a meta-analysis is fairly large and are dependent on a relationship between the reported odds ratio and its variance [31]. These limitations are less problematic for an exploratory method developed by Ioannidis and Trikalinos [57], which is a Monte Carlo method with fewer assumptions. Here this technique is applied to reported meta-analyses of statins for the prevention of CVD in high risk patients, with the aim of detecting publication bias in these studies, if it is present.

7.2 Method

7.2.1 Identification of studies

The online databases Google Scholar and Medline were queried with the terms 'statins', together with 'meta-analysis' or 'systematic reviews', along with the term 'primary prevention' to locate the most often cited meta-analyses relevant to the topic of use of statins to prevent CVD. Analyses had to report the total number of participants and events in the treatment and control groups. Studies which adjusted for variables collected after randomisation or those that did not report the trial's 2x2 table were excluded. It was also required that studies report either total CVD or CHD events (death or hospital treatment for such a

condition), or CVD death, or total mortality.

7.2.2 Methods to examine publication bias

We adapted a method designed by Ioannidis and Trikalinos [57], which compares the observed number of positive trials in a meta-analysis, with the expected, if the summary measure of effect, averaged over individual trials, is true. Excess in the observed number of positive trials, compared to the expected, is evidence of publication bias.

The observed number of positive trials in each meta-analysis, O_α , at a given level of statistical significance, α , was calculated by applying Fisher's exact test to the reported 2x2 table data of each trial. The Fisher one-sided P -value was doubled to make a two-sided test.

The corresponding expected number of positive trials, E_α was obtained by summing the statistical powers of each study. The statistical power depended on a given measure of effect which, here, was the pooled odds ratio of the meta-analysis. To obtain the power of a trial, it was assumed that the pooled odds ratio of the meta-analysis was the true measure of effect. By simulating each trial, with the given pooled odds ratio, and the same number of treated and non-treated as in the real trial, the power of the trial, for a given α was estimated as the proportion of simulated trials that were statistically significant, again by a Fisher's exact test (two-sided). The simulated number of events in the treated and untreated groups was done with binomial sampling. By simulating each constituent trial, with the given odds ratio, and the same number of treated and non-treated as in the real trial, the power of the trial was estimated as the proportion of simulated studies that were positive, again by a Fisher's exact test. In the untreated group, the binomial proportion was the percentage of actual events reported in the study and, in the treated group, the binomial sampling proportion was the untreated percentage multiplied by the risk ratio which was derived from the assumed odds ratio. We used 10 000 simulations for each trial and obtained the observed and expected counts for different lev-

els of statistical significance α . The difference between O_α and E_α was tested by a two-sided chi-square with one degree of freedom. Plots of O_α and E_α against α were drawn.

Trikalinos and Ioannidis distinguish between two patterns of results. One was when O_α and E_α were not statistically different for any $\alpha < 0.1$, suggesting no evidence of bias, and another when for some $\alpha < 0.1$ statistical significance was achieved, suggesting the presence of publication bias.

In these tests, a liberal 10% level for statistical significance of the O_α and E_α difference was adopted. A less stringent level of 0.1 is often used in studies of publication bias, as the number of published trials is usually small [31]. R software was used, including the `rmeta` and `metafor` packages, for all study calculations.

The seven reported meta-analyses were separately examined for publication bias, and then all individual studies were combined from the available meta-analyses into a ‘meta-meta-analysis’. For this overall summary, when the same trial appeared in two or more of the meta-analyses, possibly using data published at different times, only data from the last publication was used. The commands used in this chapter have been compiled into a library called `PubBias version: 1.0` and made available on the R-software online repository, CRAN.

7.3 Results

Seven meta-analyses were obtained that fitted our criteria [12, 114, 95, 13, 81, 130, 28]. Another study, the Cholesterol Treatment Trialists’ collaboration (CTT), was identified, but the trials consisted of a mix of primary and secondary prevention that compared statin with placebo, as well as trials that compared one statin regime with another [17]. There were 26 CTT trials and we omitted the 5 that were statin versus statin, at different dosages, leaving 21 placebo versus statin trials. Among these 21 trials there was a mix of patients; not all were

Table 7.1: Summary of recent meta-analyses of randomised-trials which assess the effect of statin use on risk of CVD event, conducted in people at raised risk of a CVD event.

Paper	Included studies	Total	Period	Outcome	Pooled effect estimate (95% CI)	Random or fixed effect?	Publication bias?	Comments
Brugts (2009)	9	70388	1990 to 2008	All-cause mortality	OR: 0.88 (0.81 to 0.96)	Fixed	Egger test and funnel plot negative	No heterogeneity of the treatment effect in clinical subgroups.
Bukkapatnam (2010)	6	11404	1985 to 2009	Coronary heart disease Cerebrovas-cular disease All-cause mortality	OR: 0.70 (0.61 to 0.81) OR: 0.81 (0.71 to 0.93) RR: 0.90 (0.60 to 1.35)	Random	No publication bias (Egger test)	Women only.
Cholesterol Treatment Trialists (2010)	21	129526	Up to 2009	Any coronary disease RR: 0.78 (0.76 to 0.81)	Not stated	No mention	8/21 studies conducted in people with history of CVD.	
De Vries (2012)	4	10187	1966 to 2011	Major CVD	RR: 0.75 (0.51 to 0.92)	Both random and fixed effect	No publication bias (visual inspection of funnel plot)	Diabetic patients only
Naci (2013)	18	68335	1985 to 2013	Fatal and non-fatal stroke Fatal and non-fatal MI All-cause mortality	RR 0.69 (0.51 to 0.92) RR: 0.70 (0.54 to 0.90) RR: 0.84 (0.65 to 1.09)	Random	No mention	Network meta-analysis; primary prevention subset reported here
Ray (2010)	11	65229	1970 to 2009	Major coronary events All-cause mortality.	RR 0.73 (0.68 to 0.80) RR: 0.93 (0.86 to 1.00)	Fixed (random also re-reported)	No evidence	No relationship between difference in LDL-C and mortality.
Taylor (2013)	18	56934	1994 to 2012	All-cause mortality.	RR: 0.86 (0.79 to 0.94)	Fixed	Funnel plot negative	Included studies with $\geq 10\%$ of participants with CVD at enrolment.
Tonelli (2011)	29	80711	1950 to early 2011	CVD (fatal and non-fatal) All-cause mortality. Non-fatal myocardial infarction Non-fatal stroke	RR: 0.75 (0.70 to 0.81) RR: 0.90 (0.79 to 1.03) RR: 0.64 (0.49 to 0.84) RR: 0.81 (0.68 to 0.96)	Random	Not investigated	Restricted to low risk patients.

CVD: cardiovascular disease. CHD: coronary heart disease.
OR: odds ratio. RR: relative risk. LDL-C: low density lipoprotein cholesterol.

CVD free at start, and so they were not strictly preventive trials. Nevertheless, since the majority of patient in the trials were apparently disease free at the outset, the subset of 21 trials was included as a meta-analysis. One further difficulty, however, is that the numerators, but not denominators, in each arm of these trials was not reported. Since the treated and control groups of these trials were apparently equal, denominators in each arm were assumed to be half the total size of each trial.

Table 7.1 shows some differences in the effect of statins, as judged from pooled effect estimates, from meta-analyses. Of the eight studies which reported summary measures of effect, five supported the use of statins to reduce all-cause mortality, and of the seven that reported the incidence of CVD, all reported a significant decrease in disease events in the treated, compared to the control, group. Of those which supported statin use (Brugts, Taylor, Bukkapatnam, CTT, De Vries, Naci, Taylor and Tonelli), the summary risk or odds ratios were relatively small (0.64 to 0.75), indicating between a 25 to 36% relative reduction in CVD event rates. The Ray study only reported all-cause mortality. Benefits to total mortality were less than those reported for CVD or CHD outcomes.

Summary numbers of participants, events, odds ratios and comparisons of observed to expected number of positive trials are displayed in table 7.2. Overall, the proportion of mortality in the study population was 5% (2047/ 40831) in the untreated groups, compared to 4% (1874/ 47305) for overall mortality outcomes. No evidence of publication bias was found at the 5% significance level in any of the eight studies when analysed separately, although there was consistently an excess of observed over expected number of significant trials (except the CTT and one Tonelli end-point). When all unique constituent trials were combined into a 'meta meta-analysis', some evidence of an excess of observed over expected positive trials for both endpoints was found, with 2.4 positive trials expected, when four were reported positive for the all-cause mortality outcome. This difference was not significant at the 0.10 level ($P =$

Table 7.2: Evidence for publication bias of trials reported as significant at the 5% significance level.

Study	Number of studies in meta-analysis (end point)	Untreated; events/ <i>n</i>	Treated; events/ <i>n</i>	Summ. odds ratio (95% CI)	Number observed sign. at 5% level	Number expected sign. at 5% level	Chi-square <i>P</i> -value
Brugts (2009)	9 (all-cause mortality)	1925/ 33793	1725/ 33683	0.89 (0.81 to 0.96)	2	1.7	0.81
Brugts (2009)	7† (CHD events)	1266/ 23946	966/ 23823	0.75† (0.69 to 0.82)	5†	4.1	0.49
Brugts (2009)	9 (stroke events)	767/ 33683	627/ 33683	0.88 (0.71 to 0.93)	3	2.0	0.43
Bukapatnam (2010)	2 (all-cause mortality)	46/ 3216	33/ 3137	0.73 (0.47 to 1.15)	1	0.3	0.15
Bukapatnam (2010)	2 (CHD events)	206/ 2363	163/ 2392	0.75 (0.60 to 0.94)	1	0.8	0.85
CTT (2010)	21 (major vascular events)	7136/ 64758	8934/ 64758	0.76 (0.74 to 0.79)	11	15.5	N/A
de Vries (2012)	4 (major CVD events)	576/ 5087	454/ 5100	0.73 (0.64 to 0.83)	3	2.6	0.65
Naci (2013)	18 (all-cause mortality)	1484/ 33884	1369/34451	0.91 (0.85 to 0.98)	2	1.2	0.43
Naci (2013)	15 (major coronary events)	1392/ 35150	1037/ 35470	0.73 (0.68 to 0.80)	2	0.9	0.21
Tonelli (2011)	23 (all-cause mortality)	1518/ 36608	1419/ 42887	0.90 (0.83 to 0.97)	2	1.4	0.75
Tonelli (2011)	12 (non-fatal MI)	437/ 24285	288/ 25327	0.63 (0.54 to 0.73)	3	3.2	N/A
Tonelli (2011)	9 (non-fatal stroke)	288/ 18440	231/18893	0.80 (0.67 to 0.95)	1	0.9	0.93
Ray (2010)	11 (all-cause mortality)	1447/ 32606	1346/ 32623	0.93 (0.86 to 1.00)	2	0.9	0.23
Taylor (2013)	9 (total CVD events)	1455/ 11913	1103/ 11892	0.72 (0.66 to 0.78)‡	6	3.6	0.11
Taylor (2013)	13 (all-cause mortality)	1223/ 23652	1077/ 24408	0.86 (0.79 to 0.94)‡	2	1.4	0.59
All studies	31 (all-cause mortality)	2047/ 40831	1874/ 47305	0.88 (0.83 to 0.94)	4	2.4	0.29
All studies	24 (cardiovascular disease)	3158/ 41052	2443/ 42162	0.74 (0.70 to 0.78)	9	8.3	0.74

† In the original paper, an effect estimate is given for the HPS trial, but the numbers of outcomes were unavailable.

Thus, the summary odds ratio was recalculated, using the Mantel-Haenszel method for the 7 studies for which numbers of people and numbers of events in each arm were available.

‡ Recalculated summary measure of effect, as odds rather than reported risk ratio.

CVD: cardiovascular disease. CHD: coronary heart disease. MI: myocardial infarction
OR: odds ratio. CI: confidence interval. Sign: significant.

0.29).

Figure 7.1 shows a plot of the expected number of positive studies, compared to the observed, for different levels of significance for the meta-meta-analyses. The greater the difference between the observed (solid black line with stepped increments) and expected number (curved dotted black line) of significant constituent studies, the greater the evidence of publication bias (grey dashed line indicates P -value for the difference). In both plots, the number of observed significant studies is nearly always greater than the expected number. The upper plot showing the CVD meta meta-analysis does not show convincing evidence of publication bias, however the lower plot which pools the mortality end-points does show some evidence of bias in the range of significance between 0.01 and 0.10.

7.4 Conclusion

In the meta-analyses of the effects of statins in primary prevention populations, some evidence of publication bias has been found for effects on overall mortality. Since the summary effects on mortality are not substantial, between a 7 and 16% reduction in events, it is likely that, if no bias were present, little or no survival benefit of the drugs would be observed. For CVD outcomes in the individual meta-analyses and overall 'metameta-analyses', the comparison of the observed to expected number of positive trials was not significant at the 5% level, however, there was a consistent excess in the number of positive trials compared to the expected.

Some meta-analyses reported that there was no evidence of publication bias using funnel plots and the Egger test. Funnel plots are typically difficult to interpret and are sensitive to the choice of statistical measure of precision and choice of measure of effect used to compare studies [113]. The Egger test derives a measure of association between the effect estimate (usually $\log(\text{odds ratio})$) and its variance. The Egger test is only recommended when at least 10 individual studies are summarised into a pooled effect [61], while the Ioannidis

method used here has no such restriction. The latter has, however, been only considered an exploratory technique [57]. As meta-analyses of drug trials are increasingly popular and widely used to evaluate the effects of drug treatment, an assessment of the likelihood of publication bias using this technique is often warranted.

Few other studies address the issue of publication bias in trials of statin use directly, outside of meta-analyses. One report showed that in 192 randomised trials which compared one statin to another, the sponsor's drug was greater than 20 times more likely to show a favourable comparison to the competitor's product [9]. This study suggests that publication bias in reporting drug trial results is relatively common.

Ideally, it would be useful to identify all studies that have been carried out on the effect of statins for the primary prevention of CVD, if indeed, such studies exist. This type of analysis has been completed at least once before, estimating the effect of selective serotonin reuptake inhibitors on symptoms of depression in all studies, including those which were unpublished but submitted when the drugs were considered for approval by the Food and Drug Administration. The analysis showed no effect of treatment on the total sample, and the findings differed substantially to other meta-analyses which had been carried out on published works only [62]. Similar differences between meta-analyses which include all studies, including those which are unpublished, and those which only include published studies have been found when the effect of rosiglitazone on CVD has been examined [86, 73]. The drug was finally removed from the market due to safety concerns.

The findings of this chapter are important, since they cast doubt on the utility of clinical guidelines, such as those published by the National Institute for Health and Care Excellence for preventive use of statins, which recommends treatment with statins at a risk threshold of 20% 10 year risk, after assessment using a predictive model [23]. Similarly, in New Zealand, at the time of writing (2014), statins are recommended at the threshold of 10% 5 year risk [116].

Use of statins is widespread, and is a significant cost to the health system. In 2009, a Belgian team of researchers estimated that 8% of their country's annual health budget was spent on statins [85]. In Australia, a recent nationally representative survey showed that 30% of adult participants, over the age of 50 years, had taken this class of drug in the last twenty four hours [80]. In New Zealand, in 2012, statin expenditure was reported at NZ\$70 million per year, with almost 2 million prescriptions written annually [93]. If, according to the New Zealand treasury, health spending was 14.5 billion a year [117], the proportion of New Zealand's health budget for statins corresponds to 0.5%. While some of the use of statins is in people diagnosed with CVD, a large proportion is likely to be taken by people without disease who have been identified as being at risk using NZ national guidelines [83]. In the PREDICT sample, which was reported earlier in section 3 on page 41, who were risk assessed between January 2006 and October 2009 and were enrolled into this analysis, 20% were taking statins.

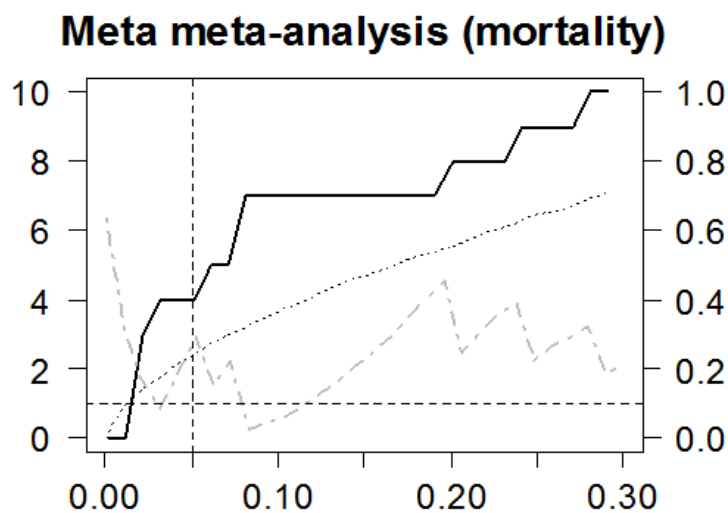
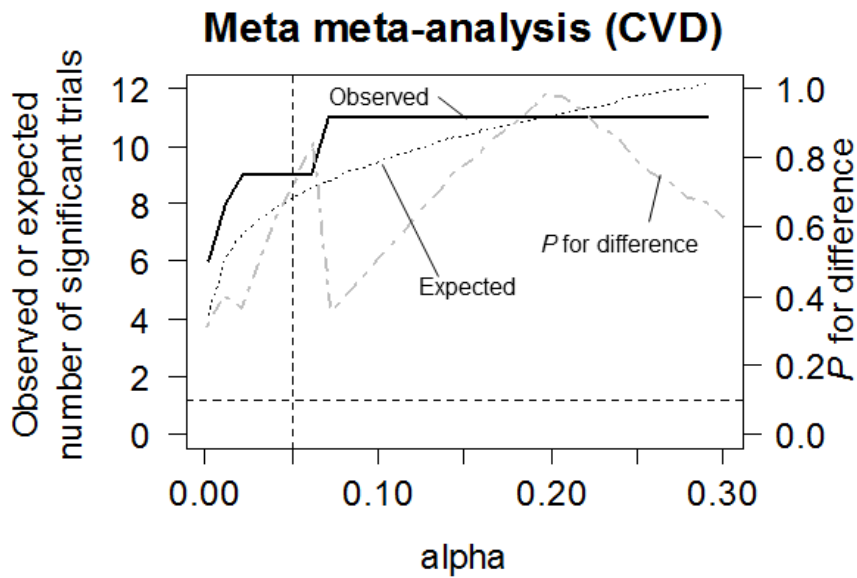


Figure 7.1: *P*-value evaluates the whether the difference between the observed and expected number of statistically significant trials of summary meta meta-analyses is likely to be due to chance, at different levels of significance (α). (Solid black line: observed number of significant studies; black dotted line: expected number of significant studies; grey dotted and dashed line: *P*-value derived from chi-square test for difference between observed and expected numbers of significant trials).

Chapter 8

Analysis 6: Serum urate and the risk of CVD: the use of DAGs to estimate causal effects

8.1 Background

Although this is not directly related to assessing the issue of drug treatment in the primary prevention of CVD, this topic is of relevance to the prediction of CVD in general. It also illustrates how DAGs may be used to decide the structure of a regression model. Unlike chapter 5, where DAGs were drawn by computer algorithms, this analysis uses DAGs created from scientific knowledge, generated by the author. This approach was used as this analysis contains much missing data, and at present, it is unclear how to go about using learning Bayesian networks with large proportions of missing data present.

In the scientific domain, uncertainty is commonly expressed over the role

of serum urate as a predictor of CVD, in primary prevention populations. For example, two research groups came to opposing conclusions about the usefulness of serum urate as a predictor of CVD:

‘Measurement of serum uric acid levels is unlikely to enhance usefully the prediction of CHD [coronary heart disease], and this factor is unlikely to be a major determinant of the disease in general populations.’ [136].

‘These results showed that hyperuricemia has a strong association with ... death in all causes, coronary heart disease, ... and indicated that serum uric acid seems to be a considerable risk factor for reduced life expectancy.’ [129].

Meta-analyses of observational studies, have reported some association between hyperuricaemia (serum urate $400 \mu\text{mol/L}$) and incident CVD. One study [60], for example, reported an increase in risk of CVD (pooled relative risk (RR) 1.46; 95% CI 1.20 to 1.73), but the association diminished when adjustment for ‘established risk factors’ was done (pooled RR 1.09; 95% CI 1.03 to 1.16). These factors included: age, systolic blood pressure, cholesterol, smoking, and either a diagnosis or other laboratory indices of diabetes. However, generally, it is not clear which factors to adjust for. Some, such as systolic blood pressure, may be on the causal pathway between serum urate and CVD, thus mediating the relationship, rather than acting as a confounder, since there is evidence that high levels of serum urate causally influence the onset of hypertension [112, 34]. Thus, whether serum urate is causally associated with CVD remains uncertain.

The prevalence of gout is extremely high, by international standards, among Māori (about 6%) and Pacific people (about 8%) living in New Zealand [139]. Together, these ethnic groups comprise between 20 to 25% of the total population, so that serum urate is measured frequently in New Zealand. Here, the relationship between urate levels and the incidence of CVD is explored.

8.2 Study design

Enrolees for the PREDICT study between January 2006 to 15 October 2009 were selected for this analysis since full dispensing and coded mortality records were available for this period. The urate concentration of enrolees were obtained by anonymized linkage to community laboratory data, in the Diagnostic Medlab (DML) database. DML was the sole provider of laboratory tests, outside of hospital, to the greater Auckland region until late 2009. If a person had had many serum urate measurements, the one immediately prior to their first PREDICT assessment was included. Serum urate and other laboratory variables (HbA1c and HDL cholesterol) were only recorded if a patient had had a laboratory test in the five year window period before their PREDICT assessment, otherwise, data for these variables were missing.

8.2.1 Statistical analysis

Statistical analysis centred on developing a survival model for incidence of any CVD event, taking serum urate as a potential predictor. A Cox regression model was used, with age-at-event as the time variable, rather than time-to-event from baseline enrolment. This approach is increasingly recommended for observational epidemiological analysis of CVD [119]. Time, therefore, was considered as left-truncated (patients were observed conditional on survival up until that point) and right-censored. Proportional hazards assumptions were checked using scaled-Schoenfeld residuals [44]. Restricted cubic splines were used to investigate the relationship between continuous variables and time-to-event as a check on the modelling assumption of linearity. Hazard ratios for continuous variables are reported by comparing the relative hazard between the 16th and 84th centiles of the variable (one standard deviation either side of the mean for a normal distribution) [124]. This allows direct comparison to relative hazards of binary variables, since the 16th and 84th centiles are roughly equivalent to the one-unit difference between 0 and 1 of a binary vari-

able. That is, two standard deviations on a binary scale, with a mean of about 0.5, is one.

To help decide model structure and which variables to include in a model, a diagram of likely causal mechanisms and pathways (DAG) was sketched, which is an idea proposed by Pearl to examine causality. The DAG that was constructed included measured and unmeasured influences on cardiovascular disease risk that were felt to be important (figure 8.1). Lines with an end arrow represent a causal link acting in the direction of the arrow. Those lines that are solid represent a link that can potentially be observed from these data, dashed lines are causal mechanisms believed to hold but cannot be observed in these data.

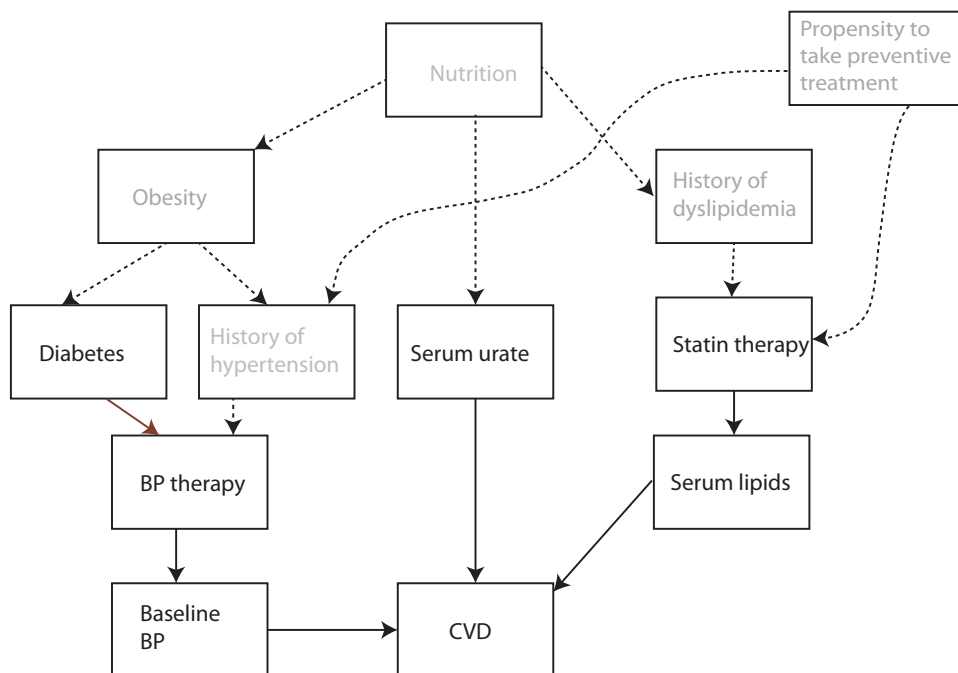


Figure 8.1: Directed acyclic graph, showing causal relationships between the exposure serum urate, and cardiovascular disease (CVD) outcomes, with observed (black font) and unobserved (grey font) variables. From the PREDICT study. BP: blood pressure.

From figure 8.1, blood pressure (BP) was considered a mediator of the influ-

ence of urate on CVD risk and it was therefore not controlled for in the model. Although not shown in the figure, similarly, a diagnosis of gout (derived from dispensing data) was likely to be a mediator of the influence of urate on CVD risk. Blood pressure, measured before baseline ('history of hypertension' in Figure 8.1), and measured at baseline (BP), were considered mediators of the influence of urate on CVD risk and were therefore not controlled for in the model. Blood pressure lowering and statin therapies were considered to represent 'colliders' [90], which may introduce bias from an unobserved variable, that would otherwise not influence this analysis, so these variables were excluded. Although certain blood pressure medications, such as thiazide diuretics, are known to raise serum urate, it was believed that this is a relatively small influence on the causal paths, compared to those otherwise identified in the DAG.

In view of these considerations, to 'block backdoor paths', using Pearl's terminology (adjust for confounding variables), the regression analysis adjusted for: gender, HbA1c, ethnic group, smoking status and lipoprotein concentrations (HDL cholesterol). Because post-treatment variables are likely to be strongly influenced by baseline recordings, and other unobserved influences (colliders), it was believed that including such variables is more likely to result in biased effect estimates than excluding them from the analysis [38].

Interactions were tested for, in particular, between serum urate and gender, because gout is more prevalent in males than females, suggesting biological differences in the handling of serum urate. From other studies of CVD, using similar risk factor profiles, effect estimates have been observed to change with age [49]. These were tested using the likelihood ratio ($p < 0.05$). When using age-at-event as the time variable, such interactions may manifest as violations of the proportional hazards assumption. Violations were tested for using the `cox.zph` function which calculates tests of the proportional-hazards assumption for each covariate, by correlating the corresponding set of scaled Schoenfeld residuals with a suitable transformation of time, based on the Kaplan-

Meier estimate of the survival function [118]. Stratification allowed different baseline hazards for exposure groups, where the proportional hazards assumption was likely to be contravened.

Because many serum urate and other laboratory values were missing, multiple imputation was carried out to account for missing covariates. Ten imputations were used, so that the modelling results reported are based on averaging model parameters over 10 random imputations, as recommended by Harrell [44]. Variables that were used in the multiple imputation modelling were: gender, serum metabolic markers (creatinine, HDL-cholesterol, triglycerides, HbA1c, urate), age-at-enrolment, systolic blood pressure, smoking status, diabetes, ethnic group, death and CVD event along with an interaction between serum urate and sex.

All analyses were done using R software (version 2.14.1) [94]. The package *rms* [45] was used for the Cox regression analysis and the functions `aregImpute` and `fit.mult.impute` for multiple imputation. The `aregImpute` function finds transformations that optimise how each variable may be predicted from every other variable, using additive semiparametric models. `fit.mult.impute` was then used to average sets of regression coefficients and compute variance and covariance, adjusted for the error derived from the uncertainty from imputation of missing data.

8.3 Results

Table 8.1 shows the baseline characteristics of the PREDICT cohort, that is, when first enrolled in the PREDICT database. Men (56%) outnumbered women, and the mean age was 55 years, with women rather older. Pacific people accounted for 23% of the cohort, Māori 16%, and Indian 7%. The remainder of the cohort and majority ethnic group, ‘Other’, was 86% European, 6% Chinese and 8% composed of a large variety of ethnic groups. More women than men were diagnosed with diabetes; however, HbA1c levels were equivalent. Men

Table 8.1: Urate and CVD analysis: baseline characteristics, by sex.

Characteristic	Men	Women	Total
Total	43815	34892	78707
Age at baseline (years)			
Mean (SD)	52.9 (10.5)	57.1 (9.8)	54.8 (10.4)
Ethnic group			
Other	23971 (54.7)	19311 (55.4)	43282 (55.0)
Māori	6578 (15.0)	5696 (16.3)	12274 (15.6)
Pacific	9967 (22.8)	7766 (22.3)	17733 (22.5)
Indian	3299 (7.5)	2119 (6.1)	5418 (6.9)
Serum urate (mmol/L); <i>n</i> = 34008 (43%)			
Mean(SD)	0.39 (0.09)	0.32 (0.09)	0.36 (0.09)
HbA1c (%); <i>n</i> = 33 075 (42%)			
Mean (SD)	6.66 (1.71)	6.68 (1.67)	6.67 (1.69)
Diagnosis of diabetes	7709 (17.6)	7170 (20.6)	14879 (18.9)
Current smoker	8630 (19.7)	5370 (15.4)	14000 (17.8)
Systolic blood pressure (mmHg)			
Mean (SD)	130.3 (17.0)	131.3 (18.4)	130.8 (17.7)
Total cholesterol/HDL ratio			
Mean (SD)	4.31 (1.35)	3.72 (1.17)	4.04 (1.31)
HDL cholesterol (mmol/L); <i>n</i> = 68224 (87%)			
Mean(SD)	1.30 (0.35)	1.55 (0.43)	1.41 (0.41)

SD: standard deviation;
mmol/L: millimoles per litre.

had higher levels of serum urate, higher lipid ratios, and lower levels of HDL. Women were less likely than men to smoke cigarettes.

Urate values were approximately normally distributed, with a lower mean in women, compared to men. There were, however, 57% missing serum urate levels, 58% missing HbA1c, and 13% missing HDL cholesterol. No other data items had missing values. Mean observed urate values varied substantially by ethnic group, but varied little by diabetes or smoking status (table 8.2). Māori and Pacific ethnic groups were highest, whereas mean levels among the Other and Indian ethnic groups were about half to two thirds of a standard deviation lower.

A total of 1328 CVD events occurred during follow-up, 167 of whom died during this period. Median follow-up time was 538 days, with a maximum of

Table 8.2: Observed mean serum urate, by sex, ethnic group, diabetes and smoking status.

Characteristic	Serum urate, mean (SD) mmol/L	
	Men	Women
Ethnic group		
Other	0.37 (0.08)	0.30 (0.08)
Māori	0.41 (0.09)	0.34 (0.09)
Pacific	0.42 (0.09)	0.36 (0.09)
Indian	0.35 (0.08)	0.29 (0.07)
Diabetes		
Yes	0.37 (0.09)	0.34 (0.09)
No	0.39 (0.09)	0.32 (0.08)
Current smoker		
Yes	0.39 (0.09)	0.33 (0.09)
No	0.39 (0.09)	0.32 (0.09)

SD: standard deviation.

1424 and minimum of one. The distributions of observed and imputed urate values (figure 8.2), showed that the variance is reduced among imputed compared to observed, because imputed values were regressed to the mean.

Initial model checking revealed that the proportional hazard assumption was not supported for gender and ethnicity. For this reason, the model was stratified on these factors, meaning that separate baseline hazards were assumed for all permutations of gender and ethnic group. A gender by urate and ethnic group by urate interaction was tested for, but neither was statistically significant.

Table 8.3 shows the results of the Cox regression analysis. In the model, comparing measures at the 16th and 84th centile, revealed a hazard ratio of 1.56 (95% CI 1.32 to 1.84) for serum urate. The relative hazard is linear (on the logarithmic scale) which indicates that an increase in serum urate of 0.29 mmol/L [= $\ln(2)/(\beta_{\text{urate}})$] doubles the relative hazard of CVD across the distribution of observed urate values. In the study population, the association between a two standard deviation difference in urate level (56% relative increase) on incident CVD risk was higher than the adjusted association of the equivalent change in the distribution of HbA1c (41% relative increase) and HDL cholesterol (22%

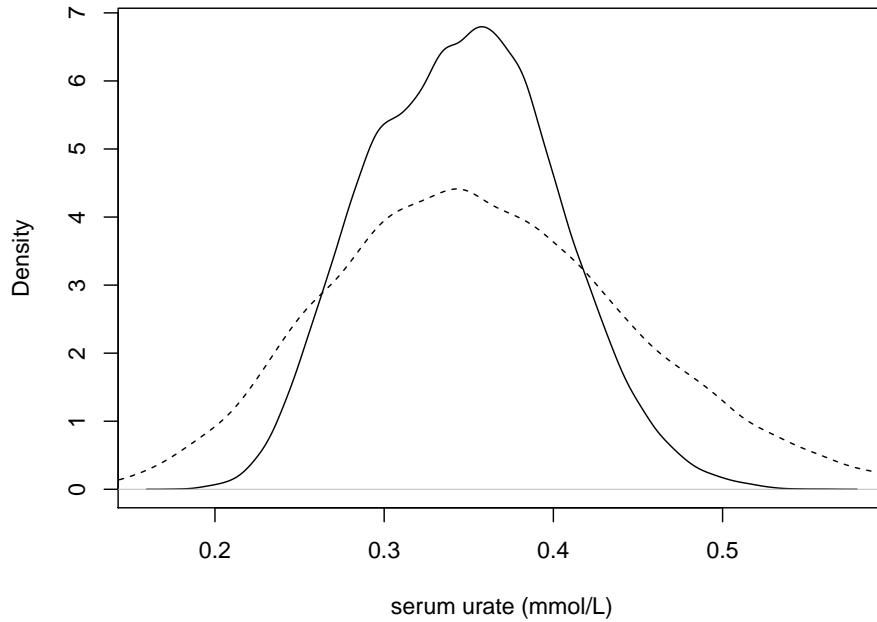


Figure 8.2: Density plot comparing the distributions of observed (dashed line) and imputed urate values (solid line).

Table 8.3: Measures of association with CVD from a Cox proportional hazard model. $n = 78707$.

Factor	Comparison*		Crude hazard ratio [†] (95% CI)	Adjusted hazard ratio [‡] (95% CI)
	High	Low		
Serum urate* (mmol/L)	0.45	0.27	1.76 (1.56 to 1.99)	1.56 (1.32 to 1.84)
Smoker: non-smoker	1.00	0.00	1.71 (1.50 to 1.95)	1.63 (1.43 to 1.86)
HbA1c* (%)	8.00	5.50	1.41 (1.29 to 1.54)	1.41 (1.29 to 1.55)
Serum HDL* (mmol/L)	1.03	1.79	1.75 (1.49 to 2.04)	1.22 (1.03 to 1.45)

* Comparisons for the continuous variables are taken at the 84th and 16th centiles ($z = 1$), for comparability with binary variables.

The binary variables are, by definition, set at 0 and 1.

[†] Stratified by sex and ethnic group.

[‡] Adjusted for all variables in table and stratified by gender and ethnic group.

relative increase).

An illustration of the extent to which change in serum urate affects survival is shown in Figure 8.3. It gives the Cox-modelled survival plot for a man in the 'Other' ethnic group, with HDL cholesterol = 1.34 mmol/L, HbA1c = 6%, and a non-smoker, showing the average change associated with a two standard deviation difference in serum urate on survival probability. This association (change in urate on survival probability) increases with age.

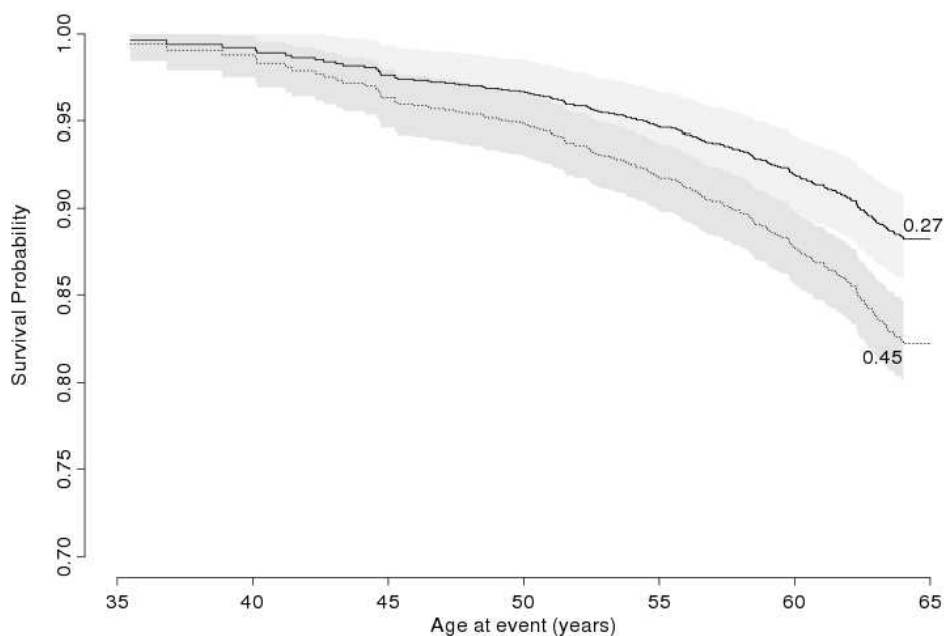


Figure 8.3: Cox-modelled, survival estimates (with 95% confidence bands) by time for serum urate at the 16th centile (0.27 mmol/L), compared to the 84th centile (0.45 mmol/L) for a non-smoking male in the 'Other' ethnic group, with HDL cholesterol = 1.34 mmol/L, HbA1c = 6%.

From the model, the cumulative incidence over a specified period of time, for example the 5-year risk, may be derived using the formula:

$$1 - \frac{S(t + \delta)}{S(t)}$$

where: t is age of an individual, δ is the interval over which the cumulative incidence is calculated (typically 5 or 10 years) and S is the survival probability, estimated by the Cox model.

For a sixty year-old male smoker in the 'Other' ethnic group with a serum urate level of 0.27 mmol/L, HbA1c of 6%, and serum HDL cholesterol 1.34 mmol/L; his five-year-risk of CVD equates to $1-(0.813/0.871) = 6.7\%$. If his serum urate is about two standard deviations greater, 0.45 mmol/L, his 5 year cumulative incidence of CVD will increase to 10.1% (about a one third increase). This calculation is based on an estimate of the baseline survival (S_0) of 88.9% at 60 years, and 83.8% at 65 years.

8.4 Discussion

Raised serum urate is likely to have a substantial causal effect on incidence of CVD, of a similar or greater magnitude to the effects of high HbA1c or low HDL cholesterol levels.

A strength of the study was that it was based on a large cohort, $n = 79707$, in which the information had been collected in a standardised way. An associated weakness however, is the incompleteness of the laboratory data, and in particular of the urate values. To deal with these missing values, multiple imputation was used, which is recommended to reduce bias in effect estimates, when compared to complete-case analysis [44]. In a sensitivity analysis, using complete-case analysis, the effect estimates were similar to those derived from multiple imputation (the adjusted hazard ratio of the effect of urate was 1.59 (95% CI: 1.32 to 1.90), compared to 1.56 in table 8.3). If serum urate had been measured in everyone, the effect estimates would probably not have differed substantially.

The multiple imputation analysis is dependent on the assumption that the mechanism which leads to absent data, given the observed information, is independent of unobserved variables ('missing at random' assumption). If data

were missing not at random, then the relationship between observed urate values and other covariates would be different from that between unobserved urate values and observed covariates. Sensitivity analyses which include a term for the influence of unobserved data in the imputation model are one way of exploring the impact of such an assumption, however, this method has not been used here.

Sensitivity analyses of the multivariate effect estimates were carried out (see figures A.1 through to A.4). These include separate analyses by gender, with and without imputation of missing values, varying the time scale chosen, including those who had used loop diuretics at baseline assessment, and dividing urate values by quintile. Generally, these analyses were congruent with the effect estimates reported in table 8.3.

A positive aspect of the study was that causal paths were explicitly considered to determine variables to adjust for in the model [91, 63]. This process is conceptually appealing and less likely to underestimate the effect of serum urate, or to introduce inappropriate confounder adjustments. Whether this DAG is an accurate reflection of the causal interaction between the variables in this analysis may be debated. For example, whether systolic blood pressure is a mediator of the effect of urate, or whether it should be included as a confounder in regression equations. To explore the effects of variable selection based on the DAG, a sensitivity analysis was carried out, altering the model by (1) adding a diagnosis of diabetes in the model, (2) interchanging HDL cholesterol with the total cholesterol/HDL ratio, and (3) adding blood pressure as a potential confounder. The first two changes to the model resulted in a less than 2% change in the adjusted hazard ratio for serum urate. After inclusion of diabetes, the effect of HbA1c was reduced from 1.41 to 1.36 (comparing individuals with baseline measures of 8.00% and 5.50%). If systolic blood pressure was included, the adjusted effect of urate diminished slightly (for 0.45 compared to 0.27 mmol/L; the HR reduced from 1.56 (95% CI 1.32 to 1.84) to HR = 1.48; (95% CI 1.25 to 1.75)).

Family history of ischaemic heart disease may also be considered a confounder of the relationship between urate and CVD. For this study, family history of CVD was recorded as a history of coronary heart disease or ischaemic stroke in a first-degree relative (father or brother <55 years, mother or sister <65 years). When this variable was added, however, the point and interval estimate for serum urate remained unchanged.

The nature of the cohort, composed of a generally high CVD risk population, rather than a designed sample is a potential weakness. The high prevalence of diabetes (about 20%) reported among the cohort, along with the gender distribution favouring men, was expected. Nevertheless, there is substantial heterogeneity in the risk profile of the study population, so there is no reason to believe the lack of representativeness was an important problem for these analyses.

Using age as the outcome variable in the Cox model imposes an assumption that entry into the study is independent of age. This is unlikely to be true here, given that national CVD screening guidelines recommend initiation of screening at specified ages which differ for varying ethnic groups. However, use of age in survival modelling is now the recommended approach for cohort studies which estimate CVD incidence [119]. This choice of time scale is justified because use of age-at-event results in less bias of effect estimates compared to time-to-event [119]. Traditional methods also lead to paradoxes in assigning risks to individuals [109]. This method is increasingly recommended if the period of observation is time from birth to event, as it is when modelling CVD risk; rather than time from an intervention to event, as it is in a randomised trial [119].

There are some differences and similarities between these results and other investigations of the effect of serum urate on CVD. Some studies report the association between raised serum urate in a binary fashion (presence or absence of hyperuricemia) [60][24]. In the present analysis, evidence of a log-linear relationship between urate concentration and CVD risk was found, so studies

which employed a cut-off are likely to underestimate the effect of the exposure. Some studies have reported effects of serum urate on CVD as a continuous variable. In another study, the Framingham study [24] ($n = 6763$) found no significant association between serum urate and coronary heart disease incidence in their adjusted Cox model among men, but a significant increase in risk with women. The analysis presented here found a statistically significant association between urate and CVD, without a gender*urate interaction, and so, the measures of association were aggregated over the whole population.

This analysis reached similar conclusions to the NHANES-I [33] cohort study ($n = 5926$) which found that serum uric acid levels were significantly associated with CVD mortality, after adjustment for known confounders. The NHANES study used CVD mortality as their outcome event while this study used hospital admission or death from CVD. Its advantages were that it had long follow-up time (16.4 years) and included both white and black participants (not present in the Framingham cohort). The authors reported that age-adjusted hazard ratios, in participants aged 45 to 54 years, for cardiovascular mortality that were similar to those presented here: 1.28 (95% CI: 1.08 to 1.52) in men and 1.43 (95% CI: 1.16 to 1.37) in women, for each 0.06 mmol/L rise in serum urate level. This study, using age as the time-to-event variable, is likely to adjust for the effect of age more accurately than traditional methods used in the NHANES study, which aggregated age into 15 year intervals. Despite these differences in methods, this study was concordant with the findings of this chapter.

Another method for assessing the evidence for causal effects of urate on CVD involves the use of Mendelian randomisation and instrumental variable analysis [88]. One study, based on a Danish cohort, showed a positive association between urate and incident CVD after adjustment for age, sex, smoking and income (0.09 mmol/L change associated with hazard ratio of 1.41; 95% CI 1.32 to 1.51). For a change in urate of 0.18mmol/L, the adjusted hazard ratio from the Danish study is 1.99, which is stronger than the equivalent adjusted hazard ratio in the PREDICT cohort (1.56).

Instrumental variable analysis is based on the assumption that a genetic influence on urate is independent of risk factors, due to the random nature of how genes are distributed during reproduction. The analysis consists of regressing genotype on mean urate to yield a predicted genetic change in the serum marker (confounders need not be accounted for, since they are assumed independent of the genotype). Then predicted genetic urate is then regressed on the outcome to yield the causal effect on the outcome, given the assumptions that underlie the method. In the Danish study, however, instrumental variable analysis showed no evidence for causal influence of urate on either blood pressure or ischaemic heart disease. Urate was, however, associated in this study with increased body size, after accounting for genetic variation. It may be, however, that the variation in urate explained by the genetic variant used in the study (about 2%) was inadequate to allow detection of these causal effects [88].

The findings of this chapter provides statistical evidence that elevated serum urate is likely to cause CVD. Biological evidence also supports these findings, highlighting the role of serum urate in the pathogenesis of endothelial injury [47]. If this causal relationship is true, attention turns to what determines serum urate, and how urate may be lowered, either by change in lifestyle, or through drug treatment. As well as the role of purine rich foods, fructose intake is likely to lead to hyperuricaemia through increased hepatic synthesis of purines [32, 82].

This laboratory-based evidence gives impetus for public health action to limit intake of sugar, the dominant source of fructose in the diet. In addition, this finding raises the question of whether pharmaceutical reduction of urate levels, with urate-lowering therapy such as allopurinol, can reduce CVD risk. Such an idea is strengthened by the results of a small ($n = 65$) cross-over trial of allopurinol therapy in patients with symptomatic, effort dependent angina. Treatment with 600mg of allopurinol per day for six weeks increased time-to-ST depression by 43 seconds (95% CI: 31 to 58), compared to placebo [87].

Māori and Pacific people in this study had mean levels of serum urate about 15% higher than 'Others', 'Indians' and 'Other Asians'. It is speculated that dietary rather than genetic differences are more likely to account for these differences, because New Zealand Māori almost all share significant ancestry with Europeans due to widespread intermarriage, yet their mean urate values vary widely.

In this regression analysis, to obtain a good fitting model, the regression was stratified by ethnicity, rather than evaluating the effects of this variable. However, because CVD incidence and prevalence is significantly higher in Māori and Pacific peoples compared to Europeans [122, 16, 128], future work will evaluate whether variation in serum urate levels account for ethnic variations in the prevalence of CVD.

In conclusion, serum urate is associated with incident CVD and is likely to be a significant factor on the disease's causal pathway.

Chapter 9

Conclusion

9.1 Statement of principal findings

The author concludes that the effects of drugs are likely to be overstated in randomised trials, due to publication bias. From the analyses presented, since the magnitude of the association between preventive drug use and CVD is not large (adjusted hazard ratios ≤ 2 or ≥ 0.5), not accounting for drug use among treated participants is unlikely to seriously distort the accuracy of models derived from cohort studies which aim to predict CVD risk. A contrary argument may be made, however, that the magnitude of the association between drug use and CVD incidence is of similar magnitude to other risk factors, such as total cholesterol:HDL ratio. It is argued, however, that drug treatment, unlike laboratory factors are likely to change with treatment guidelines and are likely to introduce unreliability into prediction models if they are used on populations substantially different from those which they are derived from.

First, the observational study information will be briefly reviewed (see summary table 9.1). Several methods of analysis investigated the nature of the association between drug use in the primary care population in PREDICT sample, with CVD. These included traditional regression, propensity score methods, and learning Bayesian networks. These measures, generally, showed small in-

creases or reductions in risk associated with the use of statins or antihypertensive agents, when other commonly measured variables were adjusted for. From chapter 4 on page 53, the association between statin drug use after baseline and incident CVD was not significant, after propensity score matching. An alternative analysis which used all the available subjects, with stabilized inverse probability of treatment weights, showed a small increased risk of disease in the statin treated group, compared to the group who remained untreated after baseline. Conversely, propensity score matching showed a small decrease in risk in the statin initiated group, compared to those who remained untreated.

A feature of statistical modelling to assess causal effects is the uncertainty of selecting covariates to adjust for [40]. A learning Bayesian network analysis provided a causal diagram compatible with the joint distribution of the baseline and CVD outcome data collected in PREDICT. This analysis was limited by a relatively small number of CVD outcomes (101), but showed no convincing influence of preventive drug treatment on CVD. Similarly, indices of CVD risk, such as systolic blood pressure and the ratio of total to LDL cholesterol were not linked by the algorithm with CVD. Drug treatment was, however, causally linked to adverse prognostic indicators, such as diabetes status and older age. These findings were concordant with the results of the traditional regression analyses, with and without propensity score matching (chapters 3 and 4, which generally showed no strong association between drug treatment and disease).

Clinical utility analysis (chapter 6 on page 79) showed little or no benefit to classification metrics when models were compared, with or without drug treatment. Although it may be argued that this analysis was carried out on a 'treatment contaminated' population.

Chapter 7 highlighted a possible inflation in the benefits reported in meta-analyses of the effect of statins in primary prevention populations. This is the first study to show positive evidence of publication bias in these summaries of

Table 9.1: Summary of the analyses.

Chapter	Why important?	Question	Method (CVD events /n)	Finding	Weaknesses
3	Weak association unlikely to distort risk prediction	Strength of association (starting drug and CVD)	Cox model (412 / 43366)	Null to increased risk	Covariate imbalance (type 1 and 2 errors and unmeasured confounding)
4	See above, addresses covariate imbalance	Strength of association (starting drug and CVD)	Cox model (655 / 56051)	Null association	Variable selection (type 2 error and unmeasured confounding)
5	See above, not necessary to assume causality	Strength of association (starting drug and CVD)	Learning Bayesian network (101 / 6256)	Conditional independence between either drug use and CVD	Exploratory, no hidden or latent causes, reverse causation, type 1 and 2 error.
6	Other analyses ignore clinical use of prediction model	Does drug use improve disease classification based on model risk score?	Clinical utility (101 / 6256)	Classification not improved	Over-fitting, untreated not considered, uncertain treatment threshold
7	Summary of trial results of statins may be inflated	Are statin meta-analyses influenced by publication bias?	Simulation study (5319 / 74296)	Some evidence of bias	Exploratory
8	Importance of other candidate markers relative to drugs	Is serum urate likely to be on causal path to CVD?	Multiple imputation; Cox model (1328 / 78707)	Moderate positive association	Imputation model accurate, variable selection, type-1 error

statins effects.

The last chapter addressed the magnitude of association between urate and CVD. Unlike the associations between drug use and CVD, the adjusted hazard ratio relating risk of urate with CVD was relatively large and statistically significant, after potential confounding variables were adjusted for. The uncertainty of which variables to adjust for was raised as a possible explanation for the conflicting conclusions of whether serum urate is an important influence on CVD risk.

9.2 Strengths and weaknesses of the thesis

A strength of this thesis is the variety of methods which were used. They include traditional Cox models (chapters 3 and 4), the use of clinical utility

analysis (chapter 6), Bayesian learning methods (chapter 5), and exploratory methods to detect publication bias (chapter 7).

Propensity score methods (chapter 4) accounted for the threat to validity posed by covariate imbalances between the treated and untreated groups, which makes the findings from a traditional regression more dependent on the structure of the model. The separation of the risk factors, collected at baseline, and treatment status, obviated the problem of reverse causation, in which baseline covariates would have otherwise been affected by treatment. It also ensured that the duration of drug exposure was relatively uniform, in that individuals were started on the drug at a similar time during follow-up.

The varying, but generally small, measures of association observed in the cohort analyses lead to further interrogation of meta-analyses of drug trials. Some inconsistencies of meta-analyses which support preventive drug treatment were identified. One study of the effect of aspirin inappropriately, in the author's view, justified the use of drugs, against the statistical evidence reported [127]. Others indicated that their findings related to populations without CVD, yet include participants with disease at enrolment [79, 114].

A strength of the thesis was the use of alternative methods to examine the validity of trial results. Traditional tests of publication bias may be insensitive to detection of this threat to the validity of meta-analyses.

This thesis has also considered statistical risk prediction, contrasting and comparing this method with causal concepts provided by the use of DAGs. Causal analysis and statistical prediction are often considered separately, however, it has been argued here, and in an accompanying publication [120], that they are related. Further, causal considerations are likely to explain some of the reduction in accuracy that is observed when applying prediction models to new populations. From the learning Bayesian network analysis, the use of drug status is likely to lead to reduced accuracy of prediction models, due to little evidence for causal influence.

The contrast between the results of chapters 5 and 8 also illustrate some of

the underlying assumptions of predictive modelling which may be questioned. The researcher-drawn DAG presented in chapter 8 contains many more causal pathways from risk factor variables than that depicted in chapter 5. The diagram drawn by the computer algorithm attributes many of the risk factors as caused by age, rather than as direct causes of the disease itself. This apparent discrepancy may be due to the lack of statistical power present from the restricted sample size, used in the learning Bayesian network analysis.

Weaknesses of the thesis include the sometimes low sample size present in analyses, which raise the possibility of type-2 errors. This applies to chapter 5, in which the causal influence between variables was assessed using learning Bayesian network algorithms. Further limitations include features of the study design, so that some baseline variables (systolic blood pressure, for example) are drawn from routine clinical practice, rather than collected with standardised protocols. This increases the likelihood of measurement error, both in baseline variables and CVD outcomes, which are derived from coded hospital admission and mortality records. Since follow-up status was dependent on death and hospital diagnosis records, no information was collected about external migration of patients, and loss to follow-up may have created bias in regression analyses.

9.3 Strengths and weaknesses of the thesis in relation to other studies

No other study, so far identified, has raised the issue of how best to deal with the issue of treatment received during follow-up when predicting CVD incidence. In the review, Glasziou and Liew have raised this issue, yet offer few solutions. Also, few other studies have sought to reconcile the varying results derived from observational and randomised studies of the effects of drugs. Chapter 7 page 94 provides a synthesis of the two sets of results. The analysis suggests that the beneficial effects of drug treatment may be inflated by

publication bias.

The findings presented here, which question the use of drug treatment for primary prevention, concur with others. For example, Brindle [10], in 2006, reviewed the results of four randomised controlled trials undertaken in populations with either hypertension or diabetes, and found no statistical evidence that a CVD risk assessment used to assign treatment improved risk factor profiles after follow-up, comparing those risk assessed with those assigned to usual care (treating individual risk factors, rather than a global CVD risk assessment, by a general practitioner or physician).

Similarly, a meta-analysis of multiple risk factor interventions (principally behavioural, dietary and pharmaceutical) to reduce the incidence of CVD returned a summary odds ratio of total and CHD mortality of 1.00 (95% CI 0.96 to 1.05) and 0.99 (95% CI 0.92 to 1.07). These were calculated from a total of 3507 events from a sample of 67520 subjects in the treated versus 4672 events from a sample of 71712 people followed up in the controls [30]. Many of the studies included drug treatment, and an analysis within classes of drug treatment showed inconsistent results. For example, those studies which defined their intervention as treatment with anti-hypertensive or statins showed a small mortality benefit, whereas those studies that treated with both agents did not. The analyses presented here differ from the multiple risk factor interventions which involved randomisation of the intervention.

The findings of chapters 4 and 3 of the association between preventive drug use and CVD may be compared with one study which investigated and compared the benefits of statins, using a variety of study designs. The meta-analysis addressed the effectiveness of statins [26]. The study addressed both primary and secondary prevention of CVD with statins, and compared observational study results of both 'prevalent' and 'incident' users with the results of randomised controlled trials. *Prevalent* users are those subjects who have already started the drug sometime before study enrolment, while *incident* users include those who start the drug after enrolment at baseline assessment.

The authors of the meta-analysis [26] reported a significant benefit of randomised studies in populations without disease (pooled hazard ratio for CHD death or non-fatal myocardial infarction: 0.69; 95% confidence interval 0.60 to 0.79). Only two observational studies addressed the issue of incident use of statins and risk of CHD, and the pooled adjusted hazard ratio was 0.80 (95% CI 0.63 to 1.02), comparing those initiated on treatment to those who remained untreated. The point estimate is similar to the analysis carried out in chapter 3 (on page 41), which reported an adjusted hazard ratio of 0.83 (95% CI: 0.55 to 1.27), comparing subjects initiated onto statins with those who remained untreated with either statins or anti-hypertensive agents. In the analysis and comparison of secondary prevention studies, the authors concluded that analyses which included prevalent users of drugs were more likely to exaggerate associations between drug use and reduced risk of disease, due to the presence of selection bias. This bias was argued to be due to prevalent users with poorer prognosis who would have otherwise died, taking part in the study because they have survived under treatment until study enrolment. So, if drugs reduce risk of the outcome, the treated prevalent group will become eventually enriched with patients of generally poorer prognosis [26].

Chapter 6, which examined net-benefit of CVD models, reported the comparative risk of prevalent statin users, compared to those who used neither statins nor anti-hypertensive drugs (adjusted hazard ratio: 1.63; 95% CI: 0.51 to 5.27). The findings of this chapter indicated increased risk among prevalent statin users, compared to the analyses of incident users. The selection bias mechanism, proposed by Danaei, suggests that the increase in risk observed among prevalent users, compared to the analyses of incident users is due to selection bias. This explanation does not, however, explain the generally increased risk observed in both prevalent and incident users of anti-hypertensive drugs (chapters 6 and 3), observed in this thesis.

Another possibility, which explains the generally increased risk seen in users of preventive drugs, is unmeasured confounding, or 'indication bias'.

That is, that individuals who are initiated onto treatment are at higher risk, after adjustment for other risk factors, than those who are not treated. There is some support for this idea in that chapters 3 and 4 show that those initiated onto statin drugs were more likely to have higher prevalence of poor prognostic factors (particularly higher cholesterol ratios and prevalence of diabetes).

These unmeasured prognostic factors which explain the increased risk of CVD must also outweigh any beneficial effects of drug treatment. If it is assumed that indication bias explains the beneficial effects of drugs, it implies that the doctor who initiates the drug is capable of predicting adverse risk factors for disease, better than an analyst is able to accomplish, after the events have occurred. This is because treatment is decided by the clinician, it is not an intrinsic property of the subject taking part in the study. From what is known of clinicians' ability to predict risk in other spheres, that it is generally poor [67], indication bias is considered less likely to account for the modestly increased risk of CVD among the treated, compared to the untreated.

A related explanation for the finding of modestly increased risk among the treated, compared to the untreated, is measurement error. It remains possible that treated individuals have a higher prevalence of misclassified unrecorded adverse prognostic factors, such as the presence of CVD, when they are thought not to have the disease.

Apart from biases, random error may play a role in the similar risk observed among the treated, compared to the untreated. If it is assumed that drugs convey a small beneficial effect on CVD risk, then this influence may only be detected in large studies with large numbers of events. In many of the analyses presented in this thesis linking drug use with CVD, the findings were not statistically significant, raising the possibility of type-2 error. This suggests that the assumed beneficial effects of the drugs were not sufficiently strong to consistently demonstrate significant associations between use or initiation of the drug and disease.

Meta-analyses of the effects of drugs on CVD or mortality generally have

large samples and high numbers of events, compared to the PREDICT analyses presented here. The study by Ray and colleagues, for example, fails to show a convincing benefit of statins on survival [95]. The overall population in this study consisted of 65229 patients with 2793 deaths during follow-up. This compares to the overall analysis undertaken in chapter 4, which has 56053 individuals with 655 CVD events occurring during follow-up. Comparing the two samples, and findings, it is not surprising that the observational analyses undertaken here, showed inconclusive results at the traditional 5% alpha level. Other studies, such as that undertaken by Taylor [114], which reported a significant reduction in CVD events among the treated, involved 23805 subjects with 2558 CVD outcomes. The Taylor study included three times the number of CVD events than the largest sample analysed in this thesis.

The learning Bayesian network analysis (chapter 5) challenged the assumption that some risk factors for CVD, that are often assumed to be causal, may not be, but, instead, associated with other causal factors (cholesterol influenced by cigarette smoking, for example). A Mendelian randomisation study has similarly concluded that HDL cholesterol, for example, is an unlikely cause of coronary heart disease, despite its association with disease [134]. Since genetic variation in HDL levels is independent of traditional risk factors for CVD, the causal effect of HDL may be estimated by predicting average HDL levels differences due to genetic variation, and then regressing these expected values on the disease outcome. The study concluded that HDL level was unlikely to cause disease events, in line with the findings of chapter 5.

9.4 Implications of the thesis

Liew and Glasziou [69] presented three approaches to deal with the issue of drug treatment during follow-up and distorted CVD risk. They included:

1. favour old studies and discard new ones
2. monitor treatment uptake and use a penalised Cox model to account for

study treatment, or

3. study cohorts over short time periods with larger numbers.

Of these options, from this thesis, treatment, either at baseline or initiated during follow-up, is unlikely to distort effect estimates from modern studies, as the effect of drug use is unlikely to be large enough to distort predictions. Also, it is likely that the very nature of risk prediction equations, based on cohort studies, means that it is unlikely that a single model will have enduring accurate prediction over time, unless the model only uses risk factors which are themselves causally linked to the disease. Chapter 5 which used learning Bayesian network algorithms, suggests that age and smoking status were causal, but not serum cholesterol ratios, diabetes status or systolic blood pressure.

This study suggests that, when the focus of research is on the prediction of CVD, the effects of drug treatment are unlikely to bias the effects of analysis. It also suggests, from chapter 5 on page 66, that drug treatment variables will not supply reliable information to predict CVD risk. This is because drug use, or initiation, is not convincingly associated with the disease outcome in a variety of observational analyses.

The findings of chapter 7, on page 94, along with a critique of one meta-analysis [127] have some relevance to the management of CVD risk in primary care. The results show possible evidence of publication bias in statin trials. Since these drugs are commonly used (by 20% of the PREDICT cohort without CVD), guidelines which routinely recommend treatment above a certain risk threshold may need to be reconsidered.

Population strategies to prevent CVD using combined treatments into one pill (the 'polypill') [135] may not deliver on their initial promise. This strategy is usually considered to improve adherence to treatment, and the benefits of each class of drugs will be additive. With some of the drug effects possibly inflated by bias, it seems probable that trials involving polypills will return

disappointing results.

Propensity score methods were explored to attempt to overcome the potential problem of bias due to indication, to estimate drug effects (chapter 4). A problem of adjusted analyses, after matching was identified: collinearity. It is recommended that in propensity score analyses, the difference in standard error between crude and adjusted estimates be monitored carefully, together with variance inflation factors associated with these models. Propensity score weighting methods do not appear to similarly suffer from this cause of model instability. This issue has seemed to receive little attention in the statistical literature relating to the use of propensity score methods.

The use of *net-benefit* is also of interest to population planning of thresholds at which to assign treatment. This type of utility analysis has not been included in current treatment guidelines (for example: [83, 20]). The method provides a framework for considering the threshold for treatment as a trade-off of the harms of treating false-positives to the benefits of treating true-positives, and may usefully indicate where a threshold for treatment should be set.

Other methodological techniques used in this thesis include learning Bayesian networks. The use of 'artificial intelligence' algorithms provided an alternative way of visualising the relationships between CVD risk factors, using DAGs (chapter 5). The method employed here of estimating the influence of arcs, although simple, is likely to be useful in other settings.

Some of the analyses presented here may be considered relatively exploratory. For example, those from chapters 5 and 6 were carried out on a relatively small, limited cohort size with only 101 CVD events occurring during follow-up. Replicating the results of this study in another study population would support the validity of the causal, and other relationships which were found. Similarly, the publication bias findings (chapter 7) were based on an exploratory technique.

9.5 Unanswered questions and future research

This work raises questions for the field of CVD risk prediction, as well as for other fields of epidemiological research. The author suggests that meta-analysts incorporate the exploratory test of publication bias used in chapter 7 more widely, due to the limits of traditional methods, such as the Egger test. Trial registers to identify all possible trials of statins, for example, should also be investigated for supporting evidence of bias.

The accuracy and reliability of the findings of the learning Bayesian network analysis are unknown. Re-examining the findings of this analysis on independently collected data would substantiate the findings of chapter 5 if concordant findings were reported.

With regard to the PREDICT research programme, the author recommends that preventive drug treatment not be included in regression models which calculate risk of first CVD event. The author also recommends that serum urate be considered for incorporation into future risk equations.

Chapter 8 explored the link between serum urate and incident CVD. During the analysis, it was noted that major differences in mean observed urate levels were detected by ethnic group (table 8.2 on page 112), which are generally concordant with ethnic differences in disease incidence. If it is accepted that serum urate causes CVD, then future work may explore the population attributable risk associated with raised levels of the marker.

In this thesis, any possible beneficial effects of blood pressure or lipid lowering drugs were likely to be more than compensated for by unmeasured adverse prognostic factors. This work highlights the generally weak nature of the adjusted association between drug use and incident CVD. These findings suggest that risk prediction models will not be unduly affected by excluding preventive drug use status in treated subjects.

Appendix A

Sensitivity analyses of the effect of serum urate on CVD incidence

Tables of sensitivity analyses of the results of multivariate Cox models of differing complexity and structure.

All hazard ratios adjusted for all other variables in table.

Comparisons for the continuous variables are taken at the 84th and 16th centiles ($z=1$), for comparability with binary variables. The binary variables are, by definition, set at 0 and 1.

Table 1. Measures of association with cardiovascular disease from a Cox proportional hazard model. Men only; stratified by ethnic group, urate measured as a continuous variable; multiple imputation; time variable: age at event, $n=43\ 815$.

Factor	Low	High	Adj. haz. ratio (95% CI)
Serum urate (mmol/L)	0.27	0.45	1.45 (1.18 to 1.77)
Hba1c (%)	5.5	8	1.37 (1.21 to 1.54)
HDL cholesterol (mmol/L)	1.79	1.03	1.27 (1.56 to 1.03)
Current smoker	No	Yes	1.60 (1.36 to 1.88)

Table 2. Measures of association with cardiovascular disease from a Cox proportional hazard model. Women only; stratified by ethnic group, urate measured as a continuous variable; multiple imputation; time variable: age-at-event, $n=34\ 892$.

Factor	Low	High	Adj. haz. ratio (95% CI)
Serum urate (mmol/L)	0.27	0.45	1.75 (1.41 to 2.18)
Hba1c (%)	5.5	8	1.47 (1.28 to 1.69)
HDL cholesterol (mmol/L)	1.79	1.03	1.20 (0.93 to 1.56)
Current smoker	No	Yes	1.69 (1.34 to 2.14)

Table 3. Measures of association with cardiovascular disease from a Cox proportional hazard model. Men only; stratified by ethnic group, urate continuous variable; multiple imputation; time variable: time-on-study, $n=43\ 815$.

Factor	Low	High	Adj. haz. ratio (95% CI)
Baseline age (years)	44.23	65.969	5.02 (4.26 to 5.91)
Serum urate (mmol/L)	0.27	0.45	1.44 (1.17 to 1.76)
Hba1c (%)	5.5	8	1.38 (1.22 to 1.55)
HDL cholesterol (mmol/L)	1.79	1.03	1.27 (1.56 to 1.03)
Current smoker	No	Yes	1.60 (1.37 to 1.88)

Figure A.1:

Table 4. Measures of association with cardiovascular disease from a Cox proportional hazard model. Women only; stratified by ethnic group, urate continuous variable; multiple imputation; time variable: time-on-study, n=34 892.

Factor	Low	High	Adj. haz. ratio (95% CI)
Baseline age (years)	44.23	65.969	5.46 (3.92 to 7.60)
Serum urate (mmol/L)	0.27	0.45	1.75 (1.41 to 2.18)
Hba1c (%)	5.5	8	1.49 (1.30 to 1.72)
HDL cholesterol (mmol/L)	1.79	1.03	1.22 (0.94 to 1.59)
Current smoker	No	Yes	1.71 (1.36 to 2.16)

Table 5. Measures of association with cardiovascular disease from a Cox proportional hazard model. Men only; stratified by ethnic group, urate continuous variable; complete case analysis; time variable: time on study, n=11 753.

Factor	Low	High	Adj. haz. ratio (95% CI)
Baseline age (years)	44.23	65.969	3.76 (2.88 to 4.92)
Serum urate (mmol/L)	0.27	0.45	1.36 (1.07 to 1.72)
Hba1c (%)	5.5	8	1.24 (1.06 to 1.46)
HDL cholesterol (mmol/L)	1.79	1.03	1.20 (0.87 to 1.67)
Current smoker	No	Yes	1.39 (1.05 to 1.83)

Table 6. Measures of association with cardiovascular disease from a Cox proportional hazard model. Women only; stratified by ethnic group, urate continuous variable; complete case analysis; time variable: time on study, n=9 243.

Factor	Low	High	Adj. haz. ratio (95% CI)
Baseline age (years)	44.23	65.969	3.30 (2.08 to 5.22)
Serum urate (mmol/L)	0.27	0.45	1.93 (1.46 to 2.54)
Hba1c (%)	5.5	8	1.44 (1.20 to 1.73)
HDL cholesterol (mmol/L)	1.79	1.03	1.08 (0.73 to 1.59)
Current smoker	No	Yes	1.76 (1.23 to 2.53)

Figure A.2:

Table 7. Measures of association with cardiovascular disease from a Cox proportional hazard model. Men only; stratified by ethnic group, urate measured as quintile; complete case analysis; time variable: time on study, n=11 753.

Factor	Low	High	Adj. haz. ratio (95% CI)
Baseline age (years)	44.23	65.969	3.75 (2.87 to 4.89)
Hba1c (%)	5.5	8	1.22 (1.04 to 1.43)
HDL cholesterol (mmol/L)	1.79	1.03	1.22 (0.88 to 1.69)
Current smoker	No	Yes	1.38 (1.05 to 1.82)
Serum urate (mmol/L; quintile over population)	0.01 to 0.27 (lowest)	0.28 to 0.32	0.80 (0.51 to 1.26)
		0.33 to 0.37	0.91 (0.60 to 1.39)
		0.38 to 0.42	0.98 (0.63 to 1.53)
		0.43 to 0.91	
		(highest)	1.19 (0.80 to 1.77)

Table 8. Measures of association with cardiovascular disease from a Cox proportional hazard model. Women only; stratified by ethnic group, urate measured as quintile; complete case analysis; time variable: time on study, n=9 243.

Factor	Low	High	Adj. haz. ratio (95% CI)
Baseline age (years)	44.23	65.969	3.34 (2.11 to 5.29)
Hba1c (%)	5.5	8	1.42 (1.19 to 1.70)
HDL cholesterol (mmol/L)	1.79	1.03	1.11 (0.75 to 1.64)
Current smoker	No	Yes	1.71 (1.19 to 2.45)
Serum urate (mmol/L; quintile over population)	0.01 to 0.27 (lowest)	0.28 to 0.32	1.08 (0.69 to 1.67)
		0.33 to 0.37	0.98 (0.61 to 1.56)
		0.38 to 0.42	1.74 (1.09 to 2.78)
		0.43 to 0.91	
		(highest)	2.00 (1.30 to 3.06)

Table 9. Measures of association with cardiovascular disease from a Cox proportional hazard model. Men and women; stratified by gender and ethnic group; urate measured as continuous variable; multiple imputation; time variable: age at event, n=78 707

Factor	Low	High	Adj. haz. ratio (95% CI)
Serum urate (mmol/L)	0.27	0.45	1.56 (1.32 to 1.84)
HDL cholesterol (mmol/L)	1.03	1.79	1.25 (1.10 to 1.43)
HbA1c (%)	5.5	8	1.42 (1.30 to 1.56)
Family history of CVD	0	1	1.30 (1.12 to 1.51)
Current smoker	0	1	1.62 (1.42 to 1.86)

Figure A.3:

Table 10. Measures of association with cardiovascular disease from a Cox proportional hazard model.

Men and women; stratified by gender and ethnic group; urate measured as continuous variable; multiple imputation; time variable: age at event, n=80 249, (includes 1549 additional people who were excluded from the main analysis due to loop diuretic use).

Factor	Low	High	Adj. haz. ratio (95% CI)
Serum urate (mmol/L)	0.27	0.45	1.38 (1.18 to 1.61)
HDL cholesterol (mmol/L)	1.03	1.79	1.25 (1.41 to 1.10)
HbA1c (%)	5.5	8	1.36 (1.24 to 1.49)
Current smoker	0	1	1.58 (1.39 to 1.79)
Loop diuretic use at enrolment	0	1	3.45 (2.85 to 4.18)

Bibliography

- [1] K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B. Kannel. Cardiovascular disease risk profiles. *American Heart Journal*, 121(1):293–298, 1991.
- [2] K. M. Anderson, P. W. Wilson, P. M. Odell, and W. B. Kannel. An updated coronary risk profile. A statement for health professionals. *Circulation*, 83(1):356–362, Jan 1991.
- [3] N. R. Anthonisen, M. A. Skeans, R. A. Wise, J. Manfreda, R. E. Kanner, and J. E. Connett. The effects of a smoking cessation intervention on 14.5-year mortality. A randomized clinical trial. *Annals of Internal Medicine*, 142(4):233–239, 2005.
- [4] G. Assmann, H. Schulte, P. Cullen, and U. Seedorf. Assessing risk of myocardial infarction and stroke: new data from the prospective cardiovascular mnster (procam) study. *European Journal of Clinical Investigation*, 37(12):925–932, Dec 2007.
- [5] P. C. Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3):399–424, 2011.
- [6] P. C. Austin. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10(2):150–161, 2011.

- [7] C. Baigent, L. Blackwell, R. Collins, J. Emberson, J. Godwin, R. Peto, J. Buring, C. Hennekens, P. Kearney, T. Meade, et al. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. *The Lancet*, 373(9678):1849, 2009.
- [8] C. Baigent, A. Keech, P. M. Kearney, L. Blackwell, G. Buck, C. Pollicino, A. Kirby, T. Sourjina, R. Peto, R. Collins, R. Simes, and C. Cholesterol Treatment Trialists. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins.[see comment][erratum appears in the lancet. 2005 oct 15-21;366(9494):1358]. *The Lancet*, 366(9493):1267–78, 2005. (Ctt) Baigent, C Keech, A Kearney, P M Blackwell, L Buck, G Pollicino, C Kirby, A Sourjina, T Peto, R Collins, R Simes, R Cholesterol Treatment Trialists' (CTT) Collaborators Comment in: *ACP J Club*. 2006 May-Jun;144(3):62; PMID: 16646606, Comment in: *Curr Atheroscler Rep*. 2007 Jan;9(1):8-9; PMID: 17169240, Comment in: *Evid Based Cardiovasc Med*. 2006 Mar;10(1):8-10; PMID: 16530655, Comment in: *The Lancet*. 2006 Feb 11;367(9509):469-70; author reply 470-1; PMID: 16473115, Comment in: *Lancet*. 2006 Feb 11;367(9509):469; author reply 470-1; PMID: 16473116.
- [9] L. Bero, F. Oostvogel, P. Bacchetti, and K. Lee. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. *PLoS Medicine*, 4(6):e184, Jun 2007.
- [10] P. Brindle, A. Beswick, T. Fahey, and S. Ebrahim. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart*, 92(12):1752–1759, 2006.
- [11] D. J. Brotman, E. Walker, M. S. Lauer, and R. G. O'Brien. In search of fewer independent risk factors. *Archives of Internal Medicine*, 165(2):138, 2005.

- [12] J. Brugts, T. Yetgin, S. Hoeks, A. Gotto, J. Shepherd, R. Westendorp, A. de Craen, R. Knopp, H. Nakamura, P. Ridker, et al. The benefits of statins in people without established cardiovascular disease but with cardiovascular risk factors: meta-analysis of randomised controlled trials. *BMJ: British Medical Journal*, 338, 2009.
- [13] R. N. Bukkapatnam, N. B. Gabler, and W. R. Lewis. Statins for primary prevention of cardiovascular mortality in women: a systematic review and meta-analysis. *Preventive Cardiology*, 13(2):84–90, 2010.
- [14] J. B. Cavender, W. J. Rogers, L. D. Fisher, B. J. Gersh, C. J. Coggin, and W. O. Myers. Effect of smoking on survival and morbidity in patients randomized to medical or surgical therapy in the coronary artery surgery study (CASS): 10-year follow-up. *Journal of the American College of Cardiology*, 20(2):287–294, 1992.
- [15] L. E. Chambless, A. R. Folsom, A. R. Sharrett, P. Sorlie, D. Couper, M. Szklo, and F. J. Nieto. Coronary heart disease risk prediction in the atherosclerosis risk in communities (aric) study. *Journal of Clinical Epidemiology*, 56(9):880–890, Sep 2003.
- [16] W. Chan, C. Wright, T. Riddell, S. Wells, A. J. Kerr, G. Gala, and R. Jackson. Ethnic and socioeconomic disparities in the prevalence of cardiovascular disease in new zealand. *New Zealand Medical Journal*, 2008.
- [17] Cholesterol Treatment Trialists, and others. Efficacy and safety of more intensive lowering of ldl cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised trials. *The Lancet*, 376(9753):1670–1681, 2010.
- [18] Cholesterol Treatment Trialists, and others. The effects of lowering ldl cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *The Lancet*, 380(9841):581–590, 2012.

- [19] V. Chongsuivatwong. *Analysis of epidemiological data using R and Epicalc*. Prince of Songkla University, Mar. 2009.
- [20] J. Cleeman, S. Grundy, D. Becker, L. Clark, et al. Expert panel on detection, evaluation and treatment of high blood cholesterol in adults. executive summary of the third report of the national cholesterol education program (NCEP) adult treatment panel (ATP III). *JAMA: the Journal of the American Medical Association*, 285(19):2486–2497, 2001.
- [21] R. Conroy, K. Pyörälä, A. e. Fitzgerald, S. Sans, A. Menotti, G. De Backer, D. De Bacquer, P. Ducimetiere, P. Jousilahti, U. Keil, et al. Estimation of ten-year risk of fatal cardiovascular disease in europe: the score project. *European Heart Journal*, 24(11):987–1003, 2003.
- [22] N. R. Cook. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical Chemistry*, 54(1):17–23, 2008.
- [23] A. Cooper, L. Nherera, N. Calvert, N. O’Flynn, N. Turnbull, J. Robson, J. Camosso-Stefinovic, C. Rule, N. Browne, G. Ritchie, et al. Clinical guidelines and evidence review for lipid modification: cardiovascular risk assessment and the primary and secondary prevention of cardiovascular disease. *Lipid modification: Full Guideline May*, page 3, 2008.
- [24] B. F. Culleton, M. G. Larson, W. B. Kannel, and D. Levy. Serum uric acid and risk for cardiovascular disease and death: the Framingham Heart Study. *Annals of Internal Medicine*, 131(1):7–13, 1999.
- [25] R. B. D’Agostino, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, and W. B. Kannel. General cardiovascular risk profile for use in primary care the framingham heart study. *Circulation*, 117(6):743–753, 2008.
- [26] G. Danaei, M. Tavakkoli, and M. A. Hernán. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a

- meta-analysis of statins. *American Journal of Epidemiology*, 175(4):250–262, 2012.
- [27] B. de la Iglesia, J. F. Potter, N. R. Poulter, M. M. Robins, and J. Skinner. Performance of the assign cardiovascular disease risk score on a uk cohort of patients from general practice. *Heart*, 97(6):491–499, Mar 2011.
- [28] F. M. de Vries, P. Denig, K. B. Pouwels, M. J. Postma, and E. Hak. Primary prevention of major cardiovascular and cerebrovascular events with statins in diabetic patients: a meta-analysis. *Drugs*, 72(18):2365–2373, Dec 2012.
- [29] D. Diao, J. M. Wright, D. K. Cundiff, and F. Gueyffier. Pharmacotherapy for mild hypertension. *The Cochrane Database of Systematic Reviews*, 8, 2012.
- [30] S. Ebrahim, F. Taylor, K. Ward, A. Beswick, M. Burke, and G. Davey Smith. Multiple risk factor interventions for primary prevention of coronary heart disease. *The Cochrane Database of Systematic Reviews*, 1(1), 2011.
- [31] M. Egger, G. D. Smith, M. Schneider, and C. Minder. Bias in meta-analysis detected by a simple, graphical test. *BMJ: British Medical Journal*, 315(7109):629–634, 1997.
- [32] B. Emmerson. Effect of oral fructose on urate production. *Annals of the Rheumatic Diseases*, 33(3):276, 1974.
- [33] J. Fang and M. H. Alderman. Serum uric acid and cardiovascular mortality. *JAMA: the Journal of the American Medical Association*, 283(18):2404–2410, 2000.
- [34] D. I. Feig, B. Soletsky, and R. J. Johnson. Effect of allopurinol on blood pressure of adolescents with newly diagnosed essential hypertension.

JAMA: the Journal of the American Medical Association, 300(8):924–932, 2008.

- [35] L. D. Fisher and D. Y. Lin. Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, 20:145–157, 1999. PMID: 10352854.
- [36] R. Garrison, W. Kannel, M. Feinleib, W. Castelli, P. McNamara, and S. Padgett. Cigarette smoking and hdl cholesterol the framingham offspring study. *Atherosclerosis*, 30(1):17–25, 1978.
- [37] T. A. Gaziano, C. R. Young, G. Fitzmaurice, S. Atwood, and J. M. Gaziano. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the nhanes i follow-up study cohort. *Lancet*, 371(9616):923–931, Mar 2008.
- [38] A. Gelman and J. Hill. *Data analysis using regression and multi-level/hierarchical models*. Cambridge University Press, 2007.
- [39] P. Good. *Permutation, parametric and bootstrap tests of hypotheses*. Springer Science+ Business Media, 2005.
- [40] S. Greenland. Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health*, 79(3):340–349, 1989.
- [41] S. Greenland, J. Pearl, and J. M. Robins. Causal diagrams for epidemiologic research. *Epidemiology*, 10(1):37–48, 1999.
- [42] S. Guo and M. W. Fraser. *Propensity score analysis*. 2010.
- [43] G. H. Guyatt, R. B. Haynes, R. Z. Jaeschke, D. J. Cook, L. Green, C. D. Naylor, M. C. Wilson, W. S. Richardson, et al. Users’ guides to the medical literature XXV. evidence-based medicine: Principles for applying the users’ guides to patient care. *JAMA: the Journal of the American Heart Association*, 284(10):1290–1296, 2000.

- [44] F. Harrell. *Regression modelling strategies*. Springer, 2010.
- [45] F. E. Harrell. *rms: Regression Modeling Strategies*, 2013. R package version 3.6-3.
- [46] F. E. Harrell, K. L. Lee, and D. B. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4):361–387, Feb 1996.
- [47] M. R. Hayden and S. C. Tyagi. Uric acid: A new look at an old risk marker for cardiovascular disease, metabolic syndrome, and type 2 diabetes mellitus: The urate redox shuttle. *Nutrition & Metabolism*, 1(1):10, 2004.
- [48] R. Henderson and N. Keiding. Individual survival time prediction using statistical models. *Journal of Medical Ethics*, 31(12):703–706, 2005.
- [49] J. Hippisley-Cox, C. Coupland, J. Robson, and P. Brindle. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. *BMJ: the British Medical Journal*, 341:c6624, 2010.
- [50] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, and P. Brindle. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. *Heart*, 94(1):34–39, Jan 2008.
- [51] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, M. May, and P. Brindle. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ: the British Medical Journal*, 335(7611):136, Jul 2007.
- [52] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, R. Minhas, A. Sheikh, and P. Brindle. Predicting cardiovascular risk in England and

Wales: prospective derivation and validation of QRISK2. *BMJ: the British Medical Journal*, 336(7659):1475–1482, Jun 2008.

- [53] M. A. Hlatky, P. Greenland, D. K. Arnett, C. M. Ballantyne, M. H. Criqui, M. S. Elkind, A. S. Go, F. E. Harrell, Y. Hong, B. V. Howard, et al. Criteria for evaluation of novel markers of cardiovascular risk: A scientific statement from the American Heart Association. *Circulation*, 119(17):2408–2416, 2009.
- [54] D. E. Ho, K. Imai, G. King, and E. Stuart. Matchit: Nonparametric pre-processing for parametric causal inference. *Journal of Statistical Software*, 42:1–28, 2011.
- [55] M. Höfler. The Bradford-Hill considerations on causality: a counterfactual perspective. *Emerging Themes in Epidemiology*, 2(1):11, 2005.
- [56] S. Hopewell, K. Loudon, M. J. Clarke, A. D. Oxman, and K. Dickersin. Publication bias in clinical trials due to statistical significance or direction of trial results. *The Cochrane Database of Systematic Reviews*, (1):MR000006, 2009.
- [57] J. P. Ioannidis and T. A. Trikalinos. An exploratory test for an excess of significant findings. *Clinical Trials*, 4(3):245–253, 2007.
- [58] W. B. Kannel, W. P. Castelli, T. Gordon, and P. M. McNamara. Serum cholesterol, lipoproteins, and the risk of coronary heart disease. the Framingham study. *Annals of Internal Medicine*, 74(1):1–12, 1971.
- [59] W. B. Kannel, D. McGee, and T. Gordon. A general cardiovascular risk profile: the Framingham study. *The American Journal of Cardiology*, 38(1):46–51, 1976.
- [60] S. Y. Kim, J. P. Guevara, K. M. Kim, H. K. Choi, D. F. Heitjan, and D. A. Albert. Hyperuricemia and coronary heart disease: A systematic review and meta-analysis. *Arthritis Care & Research*, 62(2):170–180, 2010.

- [61] B. R. Kirkwood, J. A. C. Sterne, and B. R. Kirkwood. *Essential medical statistics*. Blackwell Science, Malden, Mass., 2003.
- [62] I. Kirsch, B. J. Deacon, T. B. Huedo-Medina, A. Scoboria, T. J. Moore, and B. T. Johnson. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. *PLoS Medicine*, 5(2):e45, 2008.
- [63] U. B. Kjaerulff and A. L. Madsen. *Bayesian networks and influence diagrams: a guide to construction and analysis*, volume 22. Springer, 2013.
- [64] S. Kleinberg and G. Hripcsak. A review of causal inference for biomedical informatics. *Journal of Biomedical Informatics*, 44(6):1102–1112, 2011.
- [65] M. Law, J. Morris, and N. Wald. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *BMJ: British Medical Journal (Overseas & Retired Doctors Edition)*, 338(7705), 2009.
- [66] E. T. Lee, B. V. Howard, W. Wang, T. K. Welty, J. M. Galloway, L. G. Best, R. R. Fabsitz, Y. Zhang, J. Yeh, and R. B. Devereux. Prediction of coronary heart disease in a population with high prevalence of diabetes and albuminuria: the strong heart study. *Circulation*, 113(25):2897–2905, Jun 2006.
- [67] L. Liao and D. B. Mark. Clinical prediction models. Are we building better mousetraps? *Journal of the American College of Cardiology*, 42(5):851–853, 2003.
- [68] M. Lièvre, J. Ménard, E. Bruckert, J. Cogneau, F. Delahaye, P. Giral, E. Leitersdorf, G. Luc, L. Masana, P. Moulin, P. Passa, D. Pouchain, and G. Siest. Premature discontinuation of clinical trial for reasons not related to efficacy, safety, or feasibility. *BMJ*, 322(7286):603–605, Mar 2001.

- [69] S. M. Liew, J. Doust, and P. Glasziou. Cardiovascular risk scores do not account for the effect of treatment: a review. *Heart*, 97(9):689–697, 2011.
- [70] S. M. Liew and P. Glasziou. Risk prediction continue to ignore treatment effects. *BMJ: British Medical Journal; Rapid Response*, 2012.
- [71] E. Lonn and S. Yusuf. Polypill: the evidence and the promise. *Current Opinion in Lipidology*, 20(6):453–459, 2009.
- [72] A. G. Mainous, 3rd, R. J. Koopman, V. A. Diaz, C. J. Everett, P. W. F. Wilson, and B. C. Tilley. A coronary heart disease risk score based on patient-reported information. *Am J Cardiol*, 99(9):1236–1241, May 2007.
- [73] E. Mannucci, M. Monami, C. Lamanna, F. Gori, and N. Marchionni. Prevention of cardiovascular disease through glycemic control in type 2 diabetes: a meta-analysis of randomized clinical trials. *Nutrition, Metabolism and Cardiovascular Diseases*, 19(9):604–612, 2009.
- [74] D. Margaritis. *Learning Bayesian network model structure from data*. PhD thesis, University of Pittsburgh, 2003.
- [75] R. J. Marshall. Scaled rectangle diagrams can be used to visualize clinical and epidemiological data. *Journal of Clinical Epidemiology*, 58(10):974–81, 2005.
- [76] S. I. McFarlane, R. Muniyappa, R. Francisco, and J. R. Sowers. Pleiotropic effects of statins: Lipid reduction and beyond. *The Journal of Clinical Endocrinology & Metabolism*, 87(4):1451–1458, 2002. PMID: 11932263.
- [77] S. Mehta, S. Wells, C. Grey, T. Riddell, A. Kerr, R. Marshall, S. Ameratunga, J. Harrison, T. Kenealy, D. Bramley, et al. Initiation and maintenance of cardiovascular medications following cardiovascular risk assessment in a large primary care cohort: PREDICT CVD-16. *European Journal of Preventive Cardiology*, 2012.

- [78] L. Michaels. Aetiology of coronary artery disease: an historical approach. *British Heart Journal*, 28(2):258, 1966.
- [79] E. J. Mills, B. Rachlis, P. Wu, P. J. Devereaux, P. Arora, and D. Perri. Primary prevention of cardiovascular mortality and events with statin treatments: a network meta-analysis involving more than 65,000 patients. *Journal of the American College of Cardiology*, 52(22):1769–1781, 2008.
- [80] T. K. Morgan, M. Williamson, M. Pirotta, K. Stewart, S. P. Myers, and J. Barnes. A national census of medicines use: a 24-hour snapshot of Australians aged 50 years and older. *The Medical Journal of Australia*, 196(1):50–53, 2012.
- [81] H. Naci, J. Brugts, and T. Ades. Comparative tolerability and harms of individual statins: a study-level network meta-analysis of 246 955 participants from 135 randomized, controlled trials. *Circ Cardiovasc Qual Outcomes*, 6(4):390–399, Jul 2013.
- [82] T. Nakagawa, H. Hu, S. Zharikov, K. R. Tuttle, R. A. Short, O. Glushakova, X. Ouyang, D. I. Feig, E. R. Block, J. Herrera-Acosta, et al. A causal role for uric acid in fructose-induced metabolic syndrome. *American Journal of Physiology-Renal Physiology*, 290(3):F625–F631, 2006.
- [83] New Zealand Guidelines Group. The assessment and management of cardiovascular risk. Technical report, New Zealand Guidelines Group, 2003.
- [84] New Zealand Guidelines Group. New Zealand cardiovascular guidelines handbook: a summary resource for primary care practitioners. 2nd ed. Technical report, New Zealand Guidelines Group, 2009.
- [85] M. Neyt, C. De Laet, H. Van Brabandt, O. Franco, and D. Ramaekers. Cost-effectiveness of statins in the primary prevention of cardiovascular disease: a systematic review and economic analysis for Belgium. *Acta Cardiologica*, 64(1):1, 2009.

- [86] S. E. Nissen and K. Wolski. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, 356(24):2457–2471, 2007.
- [87] A. Noman, D. S. Ang, S. Ogston, C. C. Lang, and A. D. Struthers. Effect of high-dose allopurinol on exercise in patients with chronic stable angina: a randomised, placebo controlled crossover trial. *The Lancet*, 375(9732):2161–2167, 2010.
- [88] T. M. Palmer, B. G. Nordestgaard, M. Benn, A. Tybjærg-Hansen, G. D. Smith, D. A. Lawlor, and N. J. Timpson. Association of plasma uric acid with ischaemic heart disease and blood pressure: Mendelian randomisation analysis of two large cohorts. *BMJ: British Medical Journal*, 347, 2013.
- [89] L. Palmieri, S. Panico, D. Vanuzzo, M. Ferrario, L. Pilotto, R. Sega, G. Cesana, S. Giampaoli, and Gruppo di Ricerca del Progetto C. U. O. R. E. [Evaluation of the global cardiovascular absolute risk: the Progetto CUORE individual score]. *Ann Ist Super Sanita*, 40(4):393–399, 2004.
- [90] J. Pearl. *Causality: models, reasoning and inference*, volume 29. Cambridge Univ Press, 2000.
- [91] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Mar. 2000.
- [92] M. J. Pencina, R. B. D’Agostino, and E. W. Steyerberg. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine*, 30(1):11–21, 2011.
- [93] Pharmac. Pharmaceutical management agency annual review 2012. Webpage, 2012.

- [94] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.
- [95] K. K. Ray, S. R. K. Seshasai, S. Erqou, P. Sever, J. W. Jukema, I. Ford, and N. Sattar. Statins and all-cause mortality in high-risk primary prevention: a meta-analysis of 11 randomized controlled trials involving 65 229 participants. *Archives of Internal Medicine*, 170(12):1024, 2010.
- [96] T. Riddell, S. Wells, R. Jackson, A.-W. Lee, S. Crengle, D. Bramley, S. Ameratunga, R. Pylypchuk, J. Broad, R. Marshall, et al. Performance of Framingham cardiovascular risk scores by ethnic groups in New Zealand: Predict cvd-10. *New Zealand Medical Journal*, 123(1309):50–61, 2010.
- [97] P. M. Ridker, J. E. Buring, N. Rifai, and N. R. Cook. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women. *JAMA: the Journal of the American Medical Association*, 297(6):611–619, 2007.
- [98] P. M. Ridker, N. P. Paynter, N. Rifai, J. M. Gaziano, and N. R. Cook. C-reactive protein and parental history improve global cardiovascular risk prediction the Reynolds risk score for men. *Circulation*, 118(22):2243–2251, 2008.
- [99] P. R. Rosenbaum and D. B. Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38, 1985.
- [100] K. J. Rothman, S. Greenland, and A. M. Walker. Concepts of interaction. *American Journal of Epidemiology*, 112(4):467–470, 1980.
- [101] P. Sanmuganathan, P. Ghahramani, P. Jackson, E. Wallis, and L. Ramsay. Aspirin for primary prevention of coronary heart disease: safety

and absolute benefit related to coronary risk derived from meta-analysis of randomised trials. *Heart*, 85(3):265–271, 2001.

- [102] M. Scutari. Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software*, 35(3):1–22, 2010.
- [103] B. J. Shaten, L. H. Kuller, and J. D. Neaton. Association between baseline risk factors, cigarette smoking, and chd mortality after 10.5 years. *Preventive Medicine*, 20(5):655–669, 1991.
- [104] C. Silagy, T. Lancaster, L. Stead, D. Mant, and G. Fowler. Nicotine replacement therapy for smoking cessation. *The Cochrane Database of Systematic Reviews*, 2000.
- [105] M. A. Silva, A. C. Swanson, P. J. Gandhi, and G. R. Tataronis. Statin-related adverse events: A meta-analysis. *Clinical Therapeutics*, 28(1):26–35, 2006. doi: DOI: 10.1016/j.clinthera.2006.01.005.
- [106] J. Simes, M. Voysey, R. O’Connell, P. Glasziou, J. D. Best, R. Scott, C. Pardy, K. Byth, D. R. Sullivan, C. Ehnholm, A. Keech, and F. I. E. L. D. S. I. . A novel method to adjust efficacy estimates for uptake of other active treatments in long-term clinical trials. *PLoS One*, 5(1):e8580, 2010.
- [107] G. Siontis, I. Tzoulaki, K. C. Siontis, and J. Ioannidis. Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ: British Medical Journal*, 344, 2012.
- [108] A. Sniderman, G. Thanassoulis, P. Couture, K. Williams, A. Alam, and C. D. Furberg. Is lower and lower better and better? A re-evaluation of the evidence from the Cholesterol Treatment Trialists’ Collaboration meta-analysis for low-density lipoprotein lowering. *Journal of Clinical Lipidology*, 6(4):303–309, 2012.

- [109] M. Sperrin. Correcting for age at observation in epidemiological survival analysis. In *Annual Conference of the International Society for Clinical Biostatistics*, 2011.
- [110] E. W. Steyerberg. *Clinical prediction models: a practical approach to development, validation, and updating*. Springer, 2009.
- [111] N. J. Stone, J. Robinson, A. H. Lichtenstein, C. N. Bairey Merz, D. M. Lloyd-Jones, C. B. Blum, P. McBride, R. H. Eckel, J. S. Schwartz, A. C. Goldberg, S. T. Shero, D. Gordon, S. C. Smith, D. Levy, K. Watson, and P. W. Wilson. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*, pages –, Nov. 2013.
- [112] J. Sundström, L. Sullivan, R. B. D’Agostino, D. Levy, W. B. Kannel, and R. S. Vasan. Relations of serum uric acid to longitudinal blood pressure tracking and hypertension incidence. *Hypertension*, 45(1):28–33, 2005.
- [113] J.-L. Tang and J. L. Liu. Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology*, 53(5):477–484, 2000.
- [114] F. Taylor, K. Ward, T. Moore, M. Burke, S. Davey, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database of Systematic Reviews*, 1, 2013.
- [115] P. Thavendiranathan, A. Bagai, M. A. Brookhart, and N. K. Choudhry. Primary prevention of cardiovascular diseases with statin therapy: a meta-analysis of randomized controlled trials. *Archives of Internal Medicine*, 166(21):2307, 2006.
- [116] The Ministry of Health. Cardiovascular disease risk assessment: updated 2013. Technical report, The Ministry of Health, 2013.

- [117] The Treasury. Health. Webpage, 2013.
- [118] T. M. Therneau and P. M. Grambsch. *Modeling survival data: extending the Cox model*. Springer Verlag, 2000.
- [119] A. Thiébaud and J. Bénichou. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine*, 23(24):3803–3820, 2004.
- [120] S. Thornley. Causation and statistical prediction: Perfect strangers or bedfellows. *Journal of Biometrics & Biostatistics*, 3:e115, 2012.
- [121] S. Thornley. Are models useful to identify people for preventive treatment for cardiovascular disease? In *Australasian Epidemiology Association Annual Scientific Meeting*. Australasian Epidemiological Association, 2013.
- [122] S. Thornley, W. C. Chan, S. Crengle, T. Riddell, S. Ameratunga, S. Mehta, D. Gentles, S. Wells, R. Marshall, and R. Jackson. Sociodemographic differences in prevalence of diagnosed coronary heart disease in New Zealand estimated from linked national health records. *New Zealand Medical Journal*, 124(1334), 2011.
- [123] S. Thornley, R. Marshall, S. Wells, and R. Jackson. Using directed acyclic graphs for investigating causal paths for cardiovascular disease. *Journal of Biometrics & Biostatistics*, 4(182):2, 2013.
- [124] S. Thornley and R. J. Marshall. Measures of association in epidemiological studies: How best to compare discrete and continuous variables? *Journal of Biometrics & Biostatistics*, 2012.
- [125] S. Thornley, R. J. Marshall, R. Jackson, D. Gentles, N. Dalbeth, S. Crengle, A. Kerr, and S. Wells. Is serum urate causally associated with incident cardiovascular disease? *Rheumatology*, 52(1):135–142, 2013.

- [126] S. Thornley, R. Tayler, and K. Sikaris. Sugar restriction: the evidence for a drug-free intervention to reduce cardiovascular disease risk. *Internal Medicine Journal*, 42(S5):46–58, 2012.
- [127] S. J. Thornley and F. Barzi. Faulty logic justifies aspirin use for primary prevention of cardiovascular disease. *BMJ: British Medical Journal*, 347, 2013.
- [128] M. Tobias, W. C. Chan, C. Wright, R. Jackson, S. Mann, and L.-C. Yeh. Can the incidence and prevalence of coronary heart disease be determined from routinely collected national data? Population-based estimates for New Zealand in 2001-03. *Australian and New Zealand Journal of Public Health*, 32(1):24–27, 2008.
- [129] M. Tomita, S. Mizuno, H. Yamanaka, Y. Hosoda, K. Sakuma, Y. Matuoka, M. Odaka, M. Yamaguchi, H. Yosida, H. Morisawa, et al. Does hyperuricemia affect mortality? A prospective cohort study of Japanese male workers. *Journal of Epidemiology/Japan Epidemiological Association*, 10(6):403, 2000.
- [130] M. Tonelli, A. Lloyd, F. Clement, J. Conly, D. Husereau, B. Hemmelgarn, S. Klarenbach, F. A. McAlister, N. Wiebe, B. Manns, and A. K. D. N. . Efficacy of statins for primary prevention in people at low cardiovascular risk: a meta-analysis. *Canadian Medical Association Journal*, 183(16):E1189–E1202, Nov 2011.
- [131] C. R. Twardy, A. E. Nicholson, K. B. Korb, and J. McNeil. Epidemiological data mining of cardiovascular Bayesian networks. *Electronic Journal of Health Informatics*, 1(1):1–13, 2006.
- [132] M. I. Ulfers. *Is Age Too Automatically Controlled for as a Confounder in Epidemiologic Studies?* ProQuest, 2008.

- [133] A. J. Vickers and E. B. Elkin. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 26(6):565–574, 2006.
- [134] B. F. Voight, G. M. Peloso, M. Orho-Melander, R. Frikke-Schmidt, M. Barbalic, M. K. Jensen, G. Hindy, H. Hólm, E. L. Ding, T. Johnson, et al. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *The Lancet*, 380(9841):572–580, 2012.
- [135] N. J. Wald and M. R. Law. A strategy to reduce cardiovascular disease by more than 80%. *BMJ: British Medical Journal*, 326(7404):1419, 2003.
- [136] J. G. Wheeler, K. D. Juzwishin, G. Eiriksdottir, V. Gudnason, and J. Danesh. Serum uric acid and coronary heart disease in 9,458 incident cases and 155,084 controls: prospective study and meta-analysis. *PLoS Medicine*, 2(3):e76, 2005.
- [137] H. D. White. Improving cardiovascular risk assessment. *New Zealand Medical Journal*, 124(1335):5–9, May 2011.
- [138] P. W. Wilson, R. B. D’Agostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel. Prediction of coronary heart disease using risk factor categories. *Circulation*, 97(18):1837–1847, May 1998.
- [139] D. Winnard, C. Wright, W. J. Taylor, G. Jackson, L. Te Karu, P. J. Gow, B. Arroll, S. Thornley, B. Gribben, and N. Dalbeth. National prevalence of gout derived from administrative health data in Aotearoa New Zealand. *Rheumatology*, 51(5):901–909, 2012.
- [140] C. Wiysonge, H. Bradley, B. Mayosi, R. Maroney, A. Mbewu, L. Opie, and J. Volmink. Beta-blockers for hypertension. *The Cochrane Database of Systematic Reviews*, 1, 2007.
- [141] M. Woodward, P. Brindle, H. Tunstall-Pedoe, and S. I. G. N group on risk estimation. Adding social deprivation and family history to cardiovas-

cular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart*, 93(2):172–176, Feb 2007.

- [142] J. M. Wright and V. M. Musini. First-line drugs for hypertension. *The Cochrane Database of Systematic Reviews*, 3(3), 2009.
- [143] Y. Wu, X. Liu, X. Li, Y. Li, L. Zhao, Z. Chen, Y. Li, X. Rao, B. Zhou, R. De-trano, K. Liu, U. S. A.-P. R. C. C. S. o. C. , C. E. R. Group, and C. M. C. S. o. C. E. R. G. . Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. *Circulation*, 114(21):2217–2225, Nov 2006.
- [144] S. Xu, C. Ross, M. A. Raebel, S. Shetterly, C. Blanchette, and D. Smith. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value in Health*, 13(2):273–277, 2010.