Cardiovascular disease risk prediction modelling in 400,000 primary care patients in New Zealand: the derivation and validation of equations for primary prevention and diabetes populations

Romana Pylypchuk

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Community Health, The University of Auckland, 2018.

ABSTRACT

Most cardiovascular disease (CVD) risk prediction equations used today were developed from data collected decades ago in populations at higher CVD risk, but less ethnic and socioeconomic diversity, than patients to whom they are now applied. These equations commonly overestimate risk in contemporary populations and underperform when applied in sub-populations. Between 2002 and 2015, the PREDICT cohort study collected CVD risk profiles on a large, representative New Zealand primary care population, to develop CVD risk prediction equations for general populations and a range of sub-populations. Data were recorded using a web-based clinical decision support system in routine practice and linked to outcomes using an encrypted national health identifier. The aim of the research presented in this thesis was to develop new sex-specific models for predicting CVD risk, in the general population (PREDICT-1^{o} models) and a type 2 diabetes (T2D) sub-population (PREDICT-1^{o} T2D models). The new models were also compared to determine whether the performance of the T2D-specific models justified their development.

Cox regression models were fitted with pre-specified known CVD predictors, plus measures of ethnicity and socioeconomic status. Discrimination, calibration, and explained variation of the new and the comparison models were assessed. Existing models were used as a comparison: the North American Pooled Cohort Equations (PCEs) for the PREDICT-1^{o} models, and the New Zealand Diabetes Cohort Study (NZDCS) models for the PREDICT-1^{o} T2D models.

There were 401,752 participants who experienced 15,386 first CVD events in the general population cohort and 39,834 participants who experienced 3,295 first CVD events in the T2D sub-cohort. Median predicted CVD risk using PREDICT-1^{o} models was 3.2 % (IQR: 1.8%, 6.0%) in men, and 2.3% (IQR:1.3%, 4.2%) in women. Median predicted CVD risk using PREDICT-1^{o} T2D models was 7.5% (IQR: 5.0%, 11.5%) in men, and 4.5% (IQR: 2.7%, 7.4%) in women. Predicted CVD risk using new models was much lower than risk calculated from either PCEs or NZDCS models and the new models performed better across all performance indicators. Also, the PREDICT-1^{o} T2D models performed significantly better than the PREDICT-1^{o} models, in patients with T2D.

These findings demonstrate that CVD prediction models should be derived from populations that represent contemporary patients eligible for CVD risk assessment. Sub-population-specific models also produced measurably more accurate prediction equations. Finally, adding equity-relevant predictors (i.e. ethnicity and socioeconomic status), identified high risk sub-populations who might otherwise be undertreated.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my main supervisor Professor Rod Jackson, for his support and guidance throughout the research and writing process. Your enthusiasm and extensive knowledge inspired me to embark upon this work, and our thought-provoking discussions have kept me on track. Thank you for your patience and kindness, and for being so encouraging in the pursuit of ideas and learning opportunities. I would also like to thank my co-supervisors, Associate Professor Sue Wells, Associate Professor Roger Marshall and Dr. Katrina Poppe, for their valuable comments and support, and for sharing their immense expertise so generously.

Many thanks to my friends and colleagues in the VIEW group, for the wonderful work environment and for the inspiring conversations we had. You supported me greatly and were always willing to help. In particular, I would like to thank Suneela Mehta, Jinfeng Zhao, Billy Wu and Vanessa Selak for their friendship and kindness.

Finally, I am infinitely grateful to David, my family and friends, for their patience, understanding and support during my work on this thesis.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACS	acute coronary syndrome
ACR	albumin creatinine ratio
ACS/AHA	American College of Cardiology and American Heart Association
AF	atrial fibrillation
AHT	antihypertensive treatment
AIC	Akaike information criterion
ARIC	The Atherosclerosis Risk in Communities Study
ASCVD	atherosclerotic cardiovascular disease
ASSIGN	the cardiovascular risk score of the Scottish Intercollegiate Guidelines Network
ATPIII	Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)
AUROC	area under the Receiver Operator characteristic (ROC) curve
BIC	Bayes information criterion
BMI	body mass index
CABG	coronary artery bypass graft
CHD	coronary heart disease
CI	confidence interval
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CVD	cardiovascular disease
DBP	diastolic blood pressure
DCA	decision curve analysis

DFBETA	delta beta, a measure of standardized differences between regression coefficients when a given observation is included or excluded
DHB	District Health Board
ECG	electrocardiogram
eGFR	estimated glomerular filtration rate
EHR	electronic health record
GP	general practitioner
HbA1c	glycated haemoglobin
HDL-C	high density lipoprotein cholesterol
HF	heart failure
HR	hazard ratio
hsCRP	high sensitivity C-Reactive Protein
ICD	the International Classification of Diseases
ICD-10-AM	The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification
IDI	integrated discrimination improvement
IQR	interquartile range
LDL-C	low density lipoprotein cholesterol
LOWESS	locally weighted scatterplot smoothing
LRT	likelihood ratio test
LVH	left ventricular hypertrophy
MELAA	Middle Eastern, Latin American or African

MFP	multivariable fractional polynomials
MI	myocardial infarction
ML	maximum likelihood
NHI	National Health Index
NICE	National Institute for Health and Care Excellence
NMDS	National Minimum Dataset
NRI	net reclassification improvement
NZ	New Zealand
NZDCS	the New Zealand Diabetes Cohort Study
NZDep	the New Zealand Index of Deprivation
PCEs	the Pooled Cohort Equations
PCI	percutaneous coronary intervention
PHARMS	Pharmaceutical Claims Data Mart
PI	prognostic index
PROCAM	German Prospective Cardiovascular Münster study
PVD	peripheral vascular disease
RCT	randomised controlled trial
RECORD	Reporting of studies Conducted using Observational Routinely-collected Data
SCORE	European Systematic COronary Risk Evaluation
SBP	systolic blood pressure
SES	socioeconomic status

sICAM-1	soluble intercellular adhesion molecule-1
TC:HDL	the ratio of total cholesterol to high density lipoprotein
T2D	type 2 diabetes mellitus
TIA	transient ischaemic attack
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
UK	United Kingdom
UKPDS	UK Prospective Diabetes Study
USA	the United States of America
WHO	World Health Organisation

The major component of this thesis was published in a peer-reviewed publication:

Pylypchuk, R; Wells, S; Kerr, A; Poppe, K; Riddell, T; Harwood, M; Exeter, D; Mehta, S; Grey, C; Wu, BP; Metcalf, P; Warren, J; Harrison, J; Marshall, R; Jackson, R; (2018) Cardiovascular disease risk prediction equations in 400000 primary care patients in New Zealand: a derivation and validation study. The Lancet. 391: 1897–90

Chapter 1. Introduction

Background and objectives of this research

This thesis describes the development of new models for predicting risk of cardiovascular diseases (CVD) using routinely-collected data in New Zealand's primary care. Cardiovascular diseases are a broad group of conditions affecting heart, brain and blood vessels, many of which are caused by atherosclerosis. Atherosclerosis is a condition that develops when plaque builds up in the walls of the arteries. This build-up narrows and stiffens the arteries, which may obstruct the blood flow and, if a blood clot forms, it can stop the blood flow.

A number of risk factors have been recognised that increase the probability of developing an atherosclerotic CVD event:

- Ageing
- Men are at increased risk
- A family history of premature CHD
- Diabetes
- Raised blood pressure
- High levels of LDL cholesterol and low levels of HDL cholesterol are particularly associated with CHD
- Smoking
- Excessive consumption of saturated fat, salt, sugar
- Lack of exercise
- Obesity, largely through its effect on diabetes risk, raised blood pressure and blood lipids.

In this research, the focus was on predicting the risk of a 'global' CVD outcome, including coronary heart disease (myocardial infarction, unstable angina, coronary procedures such as percutaneous coronary intervention and coronary artery bypass graft), cerebral vascular diseases (such as ischaemic stroke and transient ischaemic attack), peripheral vascular diseases and procedures (such as femoral-popliteal artery bypass), haemorrhagic stroke and congestive heart failure (Figure 1.1). This definition

was chosen because it covers the cardiovascular conditions, predominantly caused by atherosclerosis, which can be prevented or delayed through modification of known risk factors.

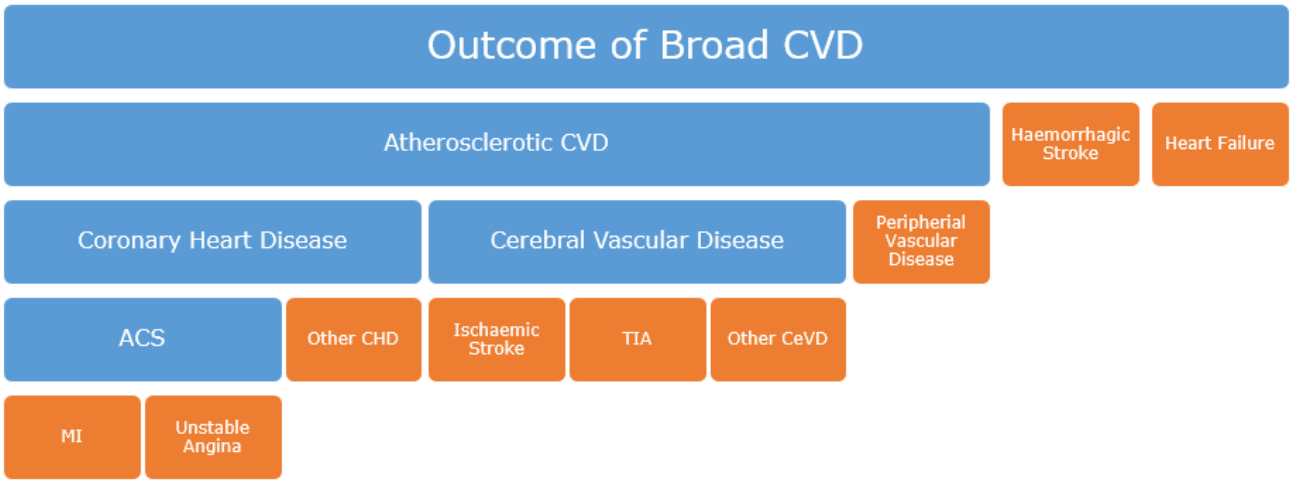


Figure 1.1. Cardiovascular outcome definition in this thesis

Cardiovascular diseases are a leading cause of premature preventable death and disease worldwide (1). It is widely accepted that most premature CVD could be prevented by public health interventions addressing risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity, high blood pressure, diabetes and raised lipids (2). Targeting high-risk patients for personalised treatment can effectively and efficiently complement public health approaches to preventing CVD (2). However, a complex chronic disease such as CVD, with its multifactorial origin and long-term nature, makes accurate targeting of high-risk patient groups difficult in routine clinical practice. Risk prediction models to help target these patients can be developed by combining risk factor and subsequent outcome data from studies in populations at risk. A risk prediction model is a mathematical equation that relates multiple predictors for a particular individual to the probability of or risk for the presence (diagnosis) or future occurrence (prognosis) of a particular outcome (3, 4). The probability estimates are typically based on combining the information from multiple predictors, as information from a single predictor is usually insufficient to provide reliable estimates of risks (3). Because such models are based on accumulated systematic evidence and data aggregated from populations, they are usually more accurate than opinion-based clinical predictions made by individual practitioners (5).

More than 40 years ago, Framingham Heart Study investigators described the clinical application of CVD risk prediction models in two influential papers (6, 7). Using a set of measurements including

age, sex, smoking status, blood pressure, blood cholesterol, glucose intolerance and left ventricular hypertrophy by electrocardiogram, they developed risk equations that predicted the occurrence of first CVD events more accurately than using either blood pressure or lipid levels alone, and proposed that treatment decisions should be informed by a multivariable predicted risk rather than by a single risk factor.

In the 1990s, New Zealand developed the world's first national CVD risk factor assessment and management guidelines based on multivariable predicted risk (8) and recommended using 1991 Framingham Heart Study prediction equations (9) to inform treatment decisions. Most international guidelines for CVD risk factor management now include similar recommendations (10-13). Although more than 360 CVD risk equations have been published since the first Framingham models, most are based on cohort studies established last century (14). Participants in these older studies are very different from patient populations the equations are now applied to, and their applicability is uncertain. Models may not perform well in practice because of deficiencies in studies on which they are based or because the target populations differ markedly from the original study population (15). New models derived from local and contemporary cohort studies is one way to address the later issue.

The candidate is a research fellow with the Vascular Informatics using Epidemiology and the Web (VIEW) cardiovascular research programme, based at the University of Auckland, which has three large cohort studies that were primarily designed to develop CVD risk prediction algorithms and to support quality improvement initiatives. Two of these cohorts are derived from web-based clinical systems: one in primary care to help clinicians assess and manage CVD risk, and one in hospitals to support quality improvement for acute coronary syndrome (ACS). Anonymised patient data are gathered from these two web-based systems, and encrypted national health identifiers are used to link to routinely-collected health data on investigations, interventions and outcomes. The third cohort is based exclusively on linked routine national datasets. This thesis is based on data from the primary care cohort, called PREDICT, which was initiated in 2002, and had comprehensive CVD risk profiles on about 500,000 patients at the time of completion of this thesis.

The VIEW programme has excellent access to regional and national routinely-collected health datasets, defined as health-related data collected for purposes other than research or without specific *a priori* research questions formulated before data collection (16). These are accumulated in the process of clinical management, documentation of clinical care, health system planning, epidemiological surveillance and can be used to answer various research questions. Routinely-collected data play a large role in all three VIEW cohorts. For example, in the PREDICT cohort, several national and

regional datasets are linked with data from primary care patients' profiles in order to obtain information on patient's CVD outcomes, clinical history, biomarker levels, pharmaceutical dispensing and sociodemographic characteristics.

The PREDICT study started in 2002, when a computerised decision-support system was developed to help general practitioners implement the national CVD guidelines while simultaneously generating a cohort of primary care patients assessed for risk of CVD. The candidate, an epidemiologist by background, joined the PREDICT team in 2009, first as data manager responsible for the linkage of the cohort data with the national and regional datasets, and subsequently as the lead research analyst responsible for statistical analyses of PREDICT data. Several years ago, the study had accumulated sufficient follow-up and CVD events to allow not only the validation of existing models, but also the development of new CVD prediction models for use in New Zealand primary care, and the candidate became responsible for developing new CVD prediction models using PREDICT primary care data.

The main objective of this thesis was to develop and validate new general population CVD risk prediction models using contemporary patient data collected in routine primary care in New Zealand, and to compare their performance against existing models, which were currently recommended by clinical guidelines. This involved identifying an explicit and transparent approach to model development and validation, based on current best-practice recommendations. In order to facilitate comparisons with the existing models, the approaches followed by established research groups and published in national and international guidelines informed the methods for assessment of models' performance.

Furthermore, this research investigated the possible advantages of a sub-population-specific approach to prediction modelling. For example, limited work has been done to compare the performance of general population models with models specifically developed in patients with type 2 diabetes mellitus (T2D). Therefore, the second objective of this study was to develop a set of T2D-specific models, using the same methods applied to develop and validate the general population models. The final objective of this study was to compare the performance of the T2D models with the performance of the general population models to determine whether the development of sub-population-specific models could be justified.

The structure of this thesis

This thesis is structured as follows.

Chapter 2 presents a literature review of the published CVD prediction models developed by influential research groups, with the focus on the methodologies of model development and validation. The review provides a summary of how key CVD prediction models have been developed and evaluated, and shows the differences and common features in the approaches to prediction modelling.

In Chapter 3, the statistical concepts and methods behind the approach to analysis, model construction, and, importantly, assessment of model performance are described.

Chapter 4 presents the PREDICT cohort study background and design. It also describes the data collected in routine primary care and from the national and regional sources linked to the final PREDICT dataset. These data are the basis of all new analyses in this thesis, therefore the key strengths and potential weaknesses of the PREDICT dataset are discussed.

Chapter 5 presents the specific methodological procedures applied during development of new CVD prediction models, and during validation of the existing models in the PREDICT dataset reported in Chapters 6-8.

Chapter 6 presents the new general population CVD risk prediction models (PREDICT-1^{o}). These are compared with the established North American Pooled Cohort Equations (PCEs), which were applied to the PREDICT-1^{o} dataset and assessed using the same performance measures as those used to assess the new general population models.

Chapter 7 presents the new CVD prediction models developed for patients with type 2 diabetes mellitus (PREDICT-1^{o} T2D). These models are compared with the New Zealand Diabetes Cohort Study (NZDCS) equations derived from the New Zealand diabetes registry data.

Chapter 8 starts by providing the rationale for and the possible advantages of building the sub-population-specific models, then compares the performance of the PREDICT-1^{o} T2D-specific models with the in PREDICT-1^{o} general population models, in PREDICT study participants with T2D.

Chapter 9 summarises the findings of this research, discusses its practical implications, and concludes by discussing priorities for future research in this field.

Chapter 2. Methodological Approaches to Developing Cardiovascular Risk Prediction Models: Literature Review

Introduction

As discussed in Chapter 1, the VIEW research programme is generating large-scale population-based cohort studies in primary and secondary care settings in New Zealand, in order to improve vascular risk prediction and better targeted vascular risk management. One of the lead projects within the VIEW research programme is to develop new CVD risk models for New Zealand based on a large contemporary primary health care cohort and the candidate was responsible for specifying the methodology and deriving new risk prediction equations for patients without prior CVD. As a first step, the candidate investigated what methodologies have been utilised in CVD prediction modelling to date.

A systematic review of CVD risk prediction models by Damen and colleagues has recently been published (14). It provided an overview of studies describing the development or external validation of models for predicting CVD risk in the general population and covered publications until June 2013. The review identified 363 prediction models from 212 publications, and found that most models were developed in Europe, and predicted risk of coronary heart disease over a 10-year period. The most common predictors were smoking and age, and most models were sex-specific. The study observed substantial heterogeneity in predictor and outcome definitions, important methodological information was often missing, and only 36% of the models were externally validated (14). For 25% of the models, crucial information was missing to allow the model to be used for individual risk prediction (14). The review concluded that, despite the abundance of models predicting incident CVD in the general population, the usefulness of these was not clear due to methodological shortcomings, incomplete presentation, and lack of external validation.

Although comprehensive and fairly recent, this systematic literature review did not focus on methods of CVD prediction model-building to the level of detail required to inform the development of these models. Therefore, a decision was made to review, in more detail, a range of methodological approaches applied when building influential CVD risk prediction models, i.e. those developed by well-established research groups and/or included in national and international clinical guidelines. Thus, an additional search was undertaken, to identify new CVD risk prediction models published after the abovementioned systematic literature review, and to limit the prediction models to those especially influential in clinical practice. These include models that have directly informed clinical guidelines or produced by leading research groups generating models which are being directly implemented for individual risk prediction.

Many reviews have demonstrated poor quality of reporting in published articles describing the development of prediction models in medicine, with insufficient information provided in all aspects of model development and validation (14, 17-20). To address the problem of incomplete reporting of prediction model studies, and to support authors writing reports on these and help editors and peer revivers in reviewing the manuscripts, the TRIPOD statement was published in 2015 by Collins and colleagues (21), along with the explanation and elaboration document by Moons and colleagues (3). The TRIPOD statement includes a checklist of 22 items which are essential for good reporting of prediction models studies and relate to the title and abstract, background and objectives, methods, results, discussion and other information. Of note, the TRIPOD checklist is not a quality assessment tool to judge the quality of a multivariable prediction model and does not prescribe how models should be developed or validated. Nevertheless, there is an implicit expectation that a study is of appropriate design and certain analyses were conducted in order to report all aspects of model development and validation (21). In this chapter, the TRIPOD checklist was used to help structure the relevant information on methods reported in publications selected for literature review.

Aims and objectives

The aim of this chapter was to undertake a review of modelling strategies utilised in development of influential CVD risk prediction models.

To achieve this aim, the objectives of the chapter were to:

- Identify risk prediction equations that have been recommended by national or international guidelines for use in clinical practice.

- Summarise how the models were developed and describe: the sources of data used to develop them (study design and setting, key study dates, duration of follow up); the study populations (types of participants, eligibility criteria, treatments received and sample sizes); the outcomes predicted; predictors used and considered for inclusion; how missing data were handled; model development methods (how predictors were handled, type of statistical model used, model-building procedures including predictor selection).

- Summarise approaches to internal validation of models, i.e. approaches to assessing model performance in the sample that was used to construct the models.

Methods

Study inclusion and exclusion criteria

Studies published from 1990 to present time were included if they developed new CVD prediction models or presented modifications of the existing ones and gave an overall risk estimation combining the predictive information from at least three variables. Only models recommended by national or international guidelines, and published in English language journals were identified. A study was eligible if it fulfilled the following criteria:

(1) it had a population of adults with no prior CVD;

(2) the study design was a randomised controlled trial (RCT) or cohort study;

(3) the main objective was to develop prediction models for estimating absolute cardiovascular risk in general populations, with or without restriction by a disease condition;

(4) models produced risk charts/calculators used to calculate risk for an individual patient;

(5) models predicted event rates for any of the major cardiovascular diseases up to 10 years;

(6) the statistical models used were regression models.

Studies using simulated populations were excluded because the aim of this thesis was to develop models using real patients. Studies published before 1990 were excluded as they have been superseded by others in international CVD guidelines. Long-term/lifetime risk models were not considered as they are not directly clinically relevant (22).

Search methods of identification of studies

A MEDLINE and EMBASE search was undertaken for a period from January 1, 1990 to December 30, 2017 using a mix of MeSH terms and free text for the keywords ‘cardiovascular diseases’, ‘cardiovascular diseases/epidemiology’, ‘cardiovascular diseases/aetiology’, ‘cardiovascular diseases/classification’, ‘risk assessment’, ‘risk factors’, ‘humans’, ‘female’, ‘male’, ‘adult’, ‘middle aged’, ‘prospective studies’, ‘cohort studies’, ‘follow-up studies’, ‘epidemiologic studies’, ‘prognosis’, ‘algorithms’, ‘models statistical’, ‘risk assessment/methods’, ‘prognosis’, ‘statistical models’, ‘proportional hazards models’, ‘regression analysis’, ‘clinical decision-making’, ‘cardiovascular risk guidelines’. The search was restricted to studies published in English language. Additional

publications were identified from the reference lists of publications, and from consultations with the experts from the PREDICT research group.

Data extraction

Full-text articles of the studies that met the inclusion criteria were retrieved. As per the TRIPOD checklist (21), data on the following items were extracted: sources of data (study design and key study dates); participants (study setting, eligibility criteria for participants, treatments received); outcomes (definitions of outcomes, how and when assessed, any actions to blind assessment of the outcome to be predicted); predictors (definitions, how and when measured, any actions to blind assessment of predictors); sample size; missing data (how these were handled, details of any imputation method); statistical analyses methods (how predictors were handled, type of model and model-building procedures, methods of internal validation, measures to assess model performance and whether they were internally or externally validated). This was a descriptive, not a quality assessment exercise in order to synthesise how key CVD prediction models have been developed and evaluated, and show the differences and common features in the approaches to prediction modelling.

Results

Over 100 publications were screened to identify 14 articles, from which 54 CVD risk prediction models were identified. Most studies presented more than one model.

There were 16 models from the Framingham Heart Study, many of which have been recommended in a wide range of national and international guidelines:

- six models for different outcome subcategories using systolic blood pressure and
- six using diastolic blood pressure from the Framingham study published in 1991 (9),
- two sex-specific models from the Framingham study published in 1998,
- and another two in 2008 (23, 24).

The 2013 US American Heart Association/American College of Cardiology (ACC/AHA) guidelines on the assessment of cardiovascular risk included four models from the Pooled Cohort Equations study (PCEs), with separate models by sex and race (for Blacks and Whites) (12).

There were four models from the Atherosclerosis Risk in Communities (ARIC) study (sex- and race-specific) (25), two each from the US Reynolds risk score for men (one using 'traditional predictors' and one with the 'traditional predictors' plus C-reactive protein and family history of premature CVD) (26) and the Reynolds risk score for women (one best-fitting model and one 'clinically simplified' model) (27).

There were four models from the New Zealand Diabetes Cohort Study (two separate models, one with and one without the blood pressure lowering treatment status, for predicting CVD risk, and an equivalent set for predicting the risk of myocardial infarction) (28).

There were eight models from the European Systematic COronary Risk Evaluation (SCORE) study (for CHD and non-CHD outcomes, using total cholesterol and TC:HDL, and for high and low risk regions) (29), two sex-specific each from the UK QRISK1, QRISK2 and Scottish Intercollegiate Guidelines Network scores called ASSIGN (30-32), and one each from the UK Prospective Diabetes Study (UKPDS), and the German Prospective Cardiovascular Münster (PROCAM) study (33, 34).

The QRISK3 study, published after the publication of the review by Damen et al., presented six models: three from men, and three for women, which differed from each other by the predictors included (35).

Sources of data and participants

Of the fourteen included publications, three used data from randomised controlled trials' participants (26, 27, 33). The remaining data were from prospective cohort studies, two of which included occupational cohorts (29, 34) and nine were population-based cohorts. Most of the studies included both men and women, except for two men-only studies (26, 34), and one women-only study (27). Eight studies out of fourteen had available information about ethnicity or race and considered it as a predictor in their models (12, 25, 27, 28, 31, 32, 34, 35). Table 2.1 summarises the characteristics of the included studies: study design, recruitment period, length of follow-up, sample sizes and number of outcomes, settings and geographical locations, basic demographic characteristics of populations and key study dates.

Table 2.1. Sources of data and participants of the studies included in the review

Study	Design	Country	Population	Recruitment Period	Duration of follow-up	Population size/events	Age	Men (%)	Ethnicity	Co-morbidities excluded	Other exclusion criteria
Framingham 1991	Prospective cohort	USA	Population cohort (original + offspring)	1968-1975	12 years	5,573/not reported	30-74	46	Mostly white, ethnicity info not used	CVD, cancer	NR
Framingham 1998	Prospective cohort	USA	Population cohort (original + offspring)	1968-1975	12 years	5,345/610	30-74	47	As above	CHD	NR
UKPDS 2001	RCT	UK	RCT subjects	1977-1991	Median 10.3 29,878 person-years	4,540/NS	25-65	Not reported	White, Afro-Caribbean, Asian-Indian	CVD, malignant hypertension, ketonuria > 3mmol/litre, severe retinopathy	Survival through the first 4 years Missing data on predictors
PROCAM 2002	Prospective cohort, occupational	Germany	Occupational cohort, employees of companies and local government authorities	1979-1985	10 years	5,389/325	35-65	100	NR	MI, stroke, IHD by ECG, angina	NR
ARIC 2003	Prospective cohort	USA	Population cohort	1987-1989	Median 10.2 years	14,054/1,064	45-64	43	White and black	CHD	Other Race. Missing data on history of CHD or predictors
SCORE 2003	Several prospective cohorts	Belgium, Denmark, Finland, France, Germany, Italy, Norway, Russia, Scotland, Spain, Sweden, UK	Pooled population and occupational cohorts	1967-1991	205178 persons, 2.7 mln pers years follow-up	205,178/7,934	19-80, varied by cohort	57	NR	Heart attack	NR
QRISK 2007	Prospective cohort, retrospectively extracted from electronic medical records	UK	Electronic medical database, GP patients	1995-2007	Derivation cohort: 8.2 million person years median 6.5 years (0-12 years)	1.9 mln/96,483	35-74	50	Area measure of ethnicity, considered for inclusion in the model	CVD or diabetes	< One year of complete data in medical record; Temporary residents; Interrupted periods of registration at GP; Missing Townsend score
QRISK2 2008	Prospective cohort, retrospectively extracted from electronic medical records	UK	Electronic medical database, GP patients	1993-2008	2.29 mln patients, 16 mln person years	2.3 mln/40,000	35-74	50	White/not recorded, Indian, Pakistani, Bangladeshi, other Asian, black African, black Caribbean, Chinese.	CVD; on statins at baseline	Temporary residents; Interrupted periods of registration at GP; No valid Townsend score
Reynolds women 2007	Prospective cohort using RCT subjects	USA	RCT subjects, female health professionals	1992-1995	Median 10.2 years	24,558/766	>=45 (med 52)	Women only	Self-reported race/ethnicity: white, black, Hispanic	CVD, cancer	Adequate plasma sample; Complete data on blood covariates

	(Woman's Health Study)								American, Asian American, other.		
ASSIGN 2007	Prospective cohort Scottish Hearth Health Extended Cohort	Scotland	Population cohort	1984-1995	10 to 21 years follow-up	13,297/1,165	30-74	49	Mostly white	CHD, stroke, TIA	Complete data on risk factors
Reynolds men 2008	Prospective occupational cohort, from RCT Physicians Health Study II	USA	RCT subjects, male health professionals	1995-1997	Median 10.8, interquartile range 7.8 to 11.2 years	10,724/1,294	50-80	100	Mostly white, ethnicity info not used	Diabetes mellitus, CVD, cancer	Adequate plasma sample; Complete data on predictors
Framingham 2008	Prospective cohort	USA	Population cohort (original + offspring)	1968-1975 1984-1987	12 years	8,491/1,174	30-74	47	Mostly white, ethnicity info not used	CVD	Complete data on predictors; Attended specified examination cycles
NZDCS 2010	Prospective cohort	New Zealand	Population cohort	2000-2006	Median 3.9, 141,169 person-years	36,127/6,479	50-79	49	55% non-European: Māori, Pacific, Indo-Asian, East Asian, other	CVD	Missing data on predictors at baseline assessment or within 2 years
Pooled Cohort Equations, 2013	Prospective cohorts	USA	Community-based cohorts	1950-1989	12 years or more for all participants	24,626/2,689	40-79	44	African-American and White, other race/ethnic groups excluded	MI, stroke, HF, PCI, coronary artery bypass surgery, AF, otherwise 'apparently healthy'	Other Ethnicity less than 12 years of follow up.; history of AF; complete data
QRISK3, 2017	Prospective cohort, retrospectively extracted from electronic medical records	UK	Electronic medical database, GP patients	1998-2015	Derivation cohort: 50.8 million person-years. Median 4.4 years, in derivation and validation cohorts	7.9 mln/363,565 in derivation and 2.7 mln in validation cohort.	25-84	49	White/not recorded, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean, Black African, Chinese, Other	CVD, on prescribed statins at baseline	Missing data on Townsend score

Outcomes and time

The majority (13 out of 14) studies developed models predicting 10-year risk of a CVD. The NZDCS presented a 5-year risk equation (28) and while the main equations derived from the PCE study were 10-year equations, they also presented 5-year equations in a subsequent publication (36). Six out of fourteen publications chose a single global CVD event as their outcome (23, 27, 30-32), although the components of the outcome varied. The Reynolds score for men used CVD as the endpoint in the main analyses, but also used CHD as the endpoint in their secondary analyses, in accordance with the Third Report of the National Cholesterol Education Programme Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III (ATP-III)) global risk assessment tool. The NZDCS presented models for both global CVD and coronary heart disease (myocardial infarction). Four studies focused on coronary heart disease: both hard and global CHD (24) and hard only (25, 33) and "major coronary events" (34). One study presented separate models for several, not mutually exclusive, cardiovascular outcomes: CHD; CHD death, sudden or not sudden; MI; stroke; global CVD; death from CVD (9). One study predicted hard atherosclerotic CVD which included MI, CHD death, and stroke (fatal and non-fatal) (12). One study predicted fatal atherosclerotic CVD only (29). The composition of the outcomes used in the studies are presented in Table 2.2.

Table 2.2. Composition of outcomes in the reported risk prediction models

	Outcomes predicted								
	Revascularization interventions	**Angina pectoris**	**Unstable angina**	**MI**	**CHD death**	**Stroke**	**Stroke death**	**Cardiac Failure**	**TIA**
Framingham 1991		x	x	x	x	x	x		x
Framingham 1998		x	x	x	X				
UKPDS 2001				x	x				
PROCAM 2002				x	x				
ARIC 2003	x			x	x				
SCORE 2003					x		x		
QRISK1 2007		x	x	x	x	x	x		x
QRISK2 2008		x	x	x	x	x	x		x
Reynolds women 2007	x			x	x	x	x		
ASSIGN 2007	x			x	x	x	x		x
Reynolds men 2008	x			x	x	x	x		
Framingham 2008		x	x	x	x	x	x	x	x
NZDCS 2010	x			x	x	x	x		x
Pooled Cohort Equations, 2013				x	x	x	x		
QRISK3 2017		x	x	x	x	x	x		x

All publications except one (29) provided some information on methods of outcomes ascertainment. Four publications reported a combination of clinical examinations, medical records/national datasets and end-point committees (9, 23, 24, 33). Seven studies reported using medical records and/or national datasets to ascertain outcomes (25, 28, 30-32, 34, 35). Two studies used medical records and end-points committees (26, 27). One study used medical records and clinical examinations (25). Regarding measures to ensure unbiased outcome ascertainment, seven studies reported using ICD codes (28-33, 35), and two reported consensus diagnosis approach which was a review of all suspected new events by a panel of experienced investigators who evaluated relevant records (23, 26).

Standard variables included in all prediction models

A number of variables were common to all of the models: age, sex, systolic blood pressure, smoking and blood lipids (mostly TC, HDL, or ratio of these). Diabetes wasn't included in the SCORE model due to its unavailability in some of the population based cohorts used in the study. The Reynolds risk score for men and QRISK1 included cohort members without diabetes only, while UKPDS and NZDCS data only included people with type 2 diabetes. Table 2.3 summarises how these standard variables were defined and fitted in models reported by the publications.

Table 2.3. Summary of the variables commonly included in reported models, their definitions and transformations

	Variable					
Publication/ name of study or risk score	**Age (in years)**	**Sex**	**Smoking**	**Systolic blood pressure (mm Hg)**	**Blood lipids (mg/dL)**	**Diabetes**
Framingham 1991	Continuous, log transformed, quadratic term if woman for MI and CHD	Included as variable in the model	Binary Persons who smoked regularly during the previous 12 months	Continuous, log transformed	TC:HDL continuous, log transformed	Binary Treatment with insulin or oral hypoglycaemic med; casual glucose >=150 mg/dL or fasting glucose >=140 mg/dL. If missing, coded "no diabetes"
Framingham 1998	Continuous, log transformed. Added quadratic term if women.	Separate equations for men and women	Binary Persons who smoked regularly during the previous 12 months	Categories as in JNC-V definitions of hypertension: <130 130-139 140-159 >=160	Categories of TC (<200, 200-239, >=240), LDL-C (<130, 130-159, >=160) and HDL-C (<35, 35-59, >=60) as per guidelines (NCEP ATP II)	Binary As in Anderson, 1991
UKPDS 2001	Age at diagnosis of diabetes, continuous	Included as a variable in the model.	Binary Current smoker at diagnosis of diabetes	Mean of two measurements taken 1 year apart	TC:HDL, mean of two measurements taken 1 year apart	HBA1C, mean of two measurements taken 1 year apart
PROCAM 2002	Continuous	Separate equations for men and women	Binary	Continuous	LDL and HDL , continuous	Binary T2D
ARIC 2003	Continuous	As above	Binary (definition not specified but measured current, ex and never smoker). Pack/years cigarettes added to some of the models.	As in Wilson 1998	TC and HDL-C categories as in Wilson 1998	Binary T2D Defined as fasting glucose level 126 mg/dL, nonfasting level 200 mg/dL, self-reported physician diagnosis of diabetes, or pharmacologic treatment for diabetes.
SCORE 2003	Fitted as time scale.	As above *	Binary	Continuous	Two different risk charts: TC and TC:HDL	Not used
QRISK 2007	Continuous, centered Log (age/10)	As above	Binary Current vs non-smoker (icl ex-smoker)	Continuous, centered	TC:HDL, continuous, centered	Diabetes-free participants
QRISK2 2008	Continuous, log (age/10), centered	As above	Binary Current vs non-smoker (icl ex-smoker)	Continuous SBP with 20 unit increment, centered	TC:HDL, continuous, centered	Binary T2D Obtained from GP records
Reynolds women 2007	Continuous	As above	Binary Definition not specified: measured current, ex and never.	Continuous, log transformed	HDL-C and TC, both log transformed	Binary T2D HbA1C added if diabetes+
ASSIGN 2007	Continuous	As above	Both binary (ex-smokers coded as smokers for the first year, then as non-smokers) and cigarettes per day for smokers.	Continuous	TC and HDL as separate continuous variables	Diabetes (binary)
Reynolds men 2008	Continuous, log transformed	Men only	Binary (current/noncurrent), self-reported	Continuous, log transformed, self-reported	TC, HDL continuous, log transformed	Diabetes-free participants

Framingham 2008	Continuous, log transformed	As above	Binary Current smoking, self-reported	Continuous, log transformed	TC and HDL continuous, log transformed	Binary Treatment with insulin or oral hypoglycaemic agents; fasting glucose >=140 mg/dL (original cohort) or >=126 mg/dL (offspring)
NZDCS 2010	Age at diagnosis of diabetes, continuous (in years)	Included as variable in the models	Non-smoker, previous, current	Continuous	TC:HDL, continuous	Binary T2D Determined by GP used duration of diabetes, continuous (in years)
PCE 2013	Continuous, log transformed; in women, log transformed and squared	Separate equations for men and women	Current smoker, binary	Continuous, log transformed Treated and untreated SBP fitted as separate terms	TC, HDL continuous, log transformed	Binary
QRISK3 2017	Continuous, centered, transformed using fractional polynomials terms	Separate equations for men and women	Non-smoker, Former smoker, Light smoker, Moderate smoker, Heavy smoker	SBP continuous, centered, standard deviation of SBP	TC:HDL, continuous, centered	Binary T1D and T2D From GP records

* Stratified by cohort and sex to calculate hazard function. The risk factor coefficients were calculated from the whole dataset which assumes their effects do not vary by cohort or sex.

Additional variables

In order to improve the accuracy of risk prediction, additional variables were investigated and included in a number of the reported models. Additional socio-demographic variables were ethnicity or race (these variables are reported here according to how they were used in publications) and socioeconomic status. Ethnicity was included in QRISK2 and QRISK3 models (European, Indian, Pakistani, Bangladeshi, other Asian, black African, black Caribbean, Chinese, other including mixed) and NZDCS models (European, Māori, Pacific, Indo-Asian, East Asian, other). Race was included in three studies and defined as Afro-Caribbean vs Caucasian/Asian-Indian by the UKPDS, and as black/white in ARIC and PCEs. Five publications reported including area-based measures of socioeconomic status among other predictors:

- QRISK, QRISK2 and QRISK3 used the Townsend score,
- ASSIGN used the Scottish index of multiple deprivation,
- NZDCS used NZDep.

Additional clinical variables were family history of premature CVD, renal disease, atrial fibrillation, rheumatoid arthritis, hypertension treatment, body mass index, time since diagnoses of diabetes, duration of diabetes, and a range of blood markers. QRISK3 models, which had the largest number of additional clinical variables, also included migraine, corticosteroid use, systemic lupus erythematosus and atypical antipsychotic use. The authors also considered but did not include HIV/AIDS. The additional variables investigated in the included studies are presented in Table 2.4.

How effect modifications were handled

Effect modification (interaction) occurs when the magnitude of the effect of the primary exposure on an outcome differs depending on the level of a third variable. In this situation, an overall estimate of association would be misleading. A common strategy to accommodate for an interaction is to stratify on the effect modifying variable. One of the assumptions of most regression models is additivity of effects of the predictors (on a relative-risk scale), i.e. absence of interaction. Interactions can be tested in these models by adding crossproduct terms. Interactions can also take the form of a change in shape which might require a non-linear

transformation of a continuous variables (for example, a linear age relationship for men but quadratic transformation of age variable for women) (37).

Studies in this review used different strategies to accommodate effect modification, with stratification and interaction terms being the most common, and nonlinear terms being the least common. For demographic variables, studies chose to stratify the population instead of adding an interaction term to the model. All but three (9, 28, 33) publications analysed men and women separately; three studies also stratified populations by race or country (12, 25, 29). One study reported considering sex-specific models, but after finding no significant interactions between sex and any risk factor, decided not to derive separate equations for men and women (28).

Three studies fitted quadratic terms for age in women, to accommodate a rapid increase in risk at younger ages and little change at older ages (9, 12, 24). To adjust for an interaction of blood pressure with pharmaceutical management, two studies fitted treated and untreated blood pressure (continuous log transformed) as two separate variables (12, 23).

The choice of interaction terms can be guided by prior subject knowledge, or can be made by applying statistical tests in model development. In studies included in this review, most decisions to include interactions were done by testing the significance of an interaction term. Two studies tested interactions and variable transformations by the maximum likelihood method (9, 33). The PCE study, in addition to having a pre-specified statistical significance threshold, used the integrated discrimination improvement and the net reclassification improvement to select interactions (12). In another study, the Reynolds risk score for women, interactions were investigated more extensively via stepwise selection procedures and multiple additive regression trees, partial dependence plots (even in absence of main effects) and further testing the interaction terms in Cox models (27). The QRISK, QRISK2 and QRISK3 studies, as well as the NZDCS, used fractional polynomials to model non-linear relationships with continuous variables (28, 31, 32, 35).

The range of interactions considered for inclusion also varied. Three studies did not report testing for any interactions (25, 26, 29). Seven studies tested the interaction and/or non-linear terms for pre-specified variables: age with all variables (12, 35); age with all variables, as well as diabetes with deprivation (32); blood pressure lowering treatment with systolic blood pressure and smoking with deprivation (31); sex and ethnicity with other predictors (28); interactions between HBA_{1c}, systolic blood pressure and lipid ratio, and between each of these

and age and sex (33); quadratic transformation of age and the interactions of TC and LDL with age (24). The PROCAM study considered interactions between all the variables selected for the final model (34). The Reynolds Risk Score for women study considered all possible interactions with all available predictors, before the final model was chosen, and explored interactions even in the absence of main effects (27). The Framingham 1991 study statistically selected covariates that appeared to model age well, including interactions and quadratic terms, then added other variables (9).

Two studies reported including interaction terms without statistical testing, based on subject knowledge: diabetes with female sex (9), and blood pressure treated and untreated (23). In one case, for clinical reasons, statistically significant interactions (ethnicity with age at diagnosis of diabetes, and duration of diabetes) were not included in final models, nor were separate models fitted for different ethnicities (28). Table 2.4 summarises the interactions and non-linear terms reported in the publications.

Table 2.4. Summary of additional (to those in Table 2.3) variables, interaction and non-linear terms included, or considered for inclusion, in the models

Publication/ name of study or risk score	Additional variables	Interactions terms, non-linear terms etc.
Framingham 1991	LVH by ECG (binary)	Terms selected by model fit: Interaction term between ECG-LVH and male for MI model. Interaction terms between log age and female, in the CVD, CHD, and MI equation. Log age^2 and female in CHD and MI equation. Term included based on previous studies: Diabetes and female for all the six outcomes.
Framingham 1998	None	Quadratic terms for age were considered in the models, and were included in models for women. Interaction terms between TC, LDL and age were considered. Neither interaction terms were found to be significant in either sex.
UKPDS 2001	Included: Race (binary, Afro-Caribbean vs Caucasian/Asian-Indian) Duration of diagnosed diabetes (in years). Considered, but not significant (P>0.5): Triglycerides Ex-smoker Asian-Indian.	TC:HDL log transformed. Tested interactions between HBA1C, SBP, TC:HDL, and between each of these and age and sex. None were significant (p>0.25).
PROCAM 2002	Included: Diabetes (binary) Family history of premature MI (binary) Triglycerides cont. (mg/dL) Considered 57 clinical and laboratory variables not listed fully in the publication.	Considered first-order interactions between the 8 variables selected for final model, none of those exceeded significance threshold of 0.05.
ARIC 2003	Included: Race (black/white, stratified by) BMI (continuous) Waist-hip ratio, sport activity index, forces expiratory volume, plasma fibrinogen, factor VIII, von Willerband factor, Lp(a), heart rate, Keys score, and carotid intima-media thickness – all substantially improved prediction of future CHD for men, less for women. Considered, but didn't improve predictivity: AHT Triglycerides Creatinine LVH by ECG Other non-traditional risk factors and biomarkers of subclinical disease.	NR
SCORE 2003	None	NR
QRISK 2007	Included: Townsend deprivation score from census 2001 (continuous) BMI (continuous, centered) AHT (binary) Family history of CVD in a first degree relative aged less than 60 (bin) Current prescription of at least one AHT: thiazide, blocker, calcium channel blocker, angiotensin converting enzyme inhibitor (binary) Considered: Percentage of South Asian residents from 2001 census (continuous) Left ventricular hypertrophy from clinical records (binary)	Tested for interaction between SBP and AHT, and between smoking and socioeconomic deprivation. Included interaction term for SBP and AHT Used fractional polynomials to model non-linear risk relations with continuous variables.
QRISK2 2008	Ethnicity: White/not recorded, Indian, Pakistani, Bangladeshi, other Asian, black African, black Caribbean, Chinese, other including mixed SES: Townsend deprivation score (output area level 2001 census data) AHT: diagnosis of hypertension and at least one current prescription of at least one AHT agent (binary) Family history: CHD in a first degree relative under 60 Renal disease (binary)	Tested for interactions between each variable and age and between diabetes and deprivation and included significant interactions in the final model: between age (10% increase) and BMI, Townsend score, SBP (20 unit increase), family history, smoking, treated hypertension, type 2 diabetes, AF. These interactions suggested increased hazard ratios for the risk factors among younger patients compared with

	Atrial fibrillation (binary) Rheumatoid arthritis (binary) BMI (continuous, centered)	older patients. Same interaction terms used in equations for men and women. Used fractional polynomials to investigate non-linear risk relations with continuous variables.
Reynolds women 2007	Included: HbA_{1C} (continuous) added if diabetic Family history: parental history of MI before age of 60 Apolipoprotein B-100 and apolippprotein A-1 in best-fitting model instead of cholesterol (both in continuous form). HsCRP in the simplified model, continuous log transformed. Considered, not included: AHT Creatinine Total cholesterol HDL-C LDL-C Lipoprotein (a) Apolipoproteins A-I and B-100 hsCRP sICAM-1 Fibrinogen HbA_{1C} Plasma homocysteine concentration	Considered all potential transformations and interactions between all available exposure variables and all blood biomarkers. Two interaction terms were included in the best overall prediction algorithm: HbA_{1C} if diabetes was present and lipoprotein(a) level if apolipoprotein B-100 was 100 mg/dL or higher. The simplified model included only HbA_{1C} if diabetes present; the other interaction was removed.
ASSIGN 2007	Included: SES: Scottish Index of Multiple Deprivation Family history: "heart disease or stroke before 60 in parents or siblings", if negative – "heart disease or stroke before 60 in two or more uncles, aunts or first cousins". Considered but not included: LVH by ECG (significant only in women)	Identified a significant interaction between sex and SIMD score. Not specified if other interaction terms tested.
Reynolds men 2008	**Main analysis** Included (based on published literature): Parental history of myocardial infarction before 60 years (bin), self-reported HsCRP – continuous, log transformed Considered but not included: Glomerular filtration rate Other biomarkers (not listed) **Secondary analyses** Basic model plus blood pressure lowering treatment Basic model lipid lowering treatment	NR
Framingham 2008	Included: LVH by ECG: in model for MI in men only Considered but not included: BMI Triglycerides	Do not report testing for interactions. To adjust for use of AHT: log of SBP if treated, log of SBP if untreated.
NZDCS 2010	Included: Ethnicity (self-assigned) HbA_{1C} ACR Considered but not included: BMI, SES, DBP, blood pressure lowering and lipid lowering treatment.	Tested interaction between medications ad blood pressure, lipids; also between sex, ethnicity and other risk vars. Interaction between blood pressure-lowering medication and SBP was included in some models. Used fractional polynomials to investigate non-linear risk relations with continuous variables.
Pooled Cohort 2013	Included: Race (black/white, stratified by) Considered but not included: Lipid lowering treatment Diastolic blood pressure (DBP) Family history of ASCVD Moderate or severe chronic kidney disease defined via eGFR BMI (continuous and categorical, modelled separately Would have considered but couldn't due to missing data: High-sensitivity C reactive protein Apolipoprotein B Microalbuminuria Cardiorespiratory fitness Coronary artery calcium score Carotid artery intima-media thickness Ankle-brachial index	Investigated age-covariate interactions. Interactions included: Age x total cholesterol Age x HDL Age x SBP treated for women only Age x SBP untreated models for women only Age x current smoker Non-linear terms: Quadratic transform of age for women only.

QRISK3 2017	Included: Ethnicity (White/NR, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean, Black African, Chinese, other) Family history of CHD in first-degree relative<60 yrs Townsend score BMI Treated hypertension Rheumatoid arthritis Atrial fibrillation CKD Migraine Corticosteroid use Systemic lupus erythematosus Atypical antipsychotic use Severe mental illness Erectile dysfunction or treatment Considered but not included: HIV/AIDS	Investigated age-covariate interactions (method not specified). Interactions included: Age x BMI Age x SBP Age x Townsend score Age x family history of CHD Age x treated hypertension Age x T1D Age x T2D Age x CKD Age x AF Age x smoking status Age x migraine Age x corticosteroid use Age x SLE (women) Age x erectile dysfunction (men) Non-linear terms: Fractional polynomial terms for BMI

Statistical models for prediction

All studies used survival modelling methods, predominantly Cox regression. The Framingham equations by Anderson, 1991 (9) and the SCORE equations (29) were Weibull models: the former used a non-proportional hazards Weibull accelerated failure time model, while the later used the standard Weibull model and cross checked by comparison with a Cox model to ensure that the Weibull assumption about the shape of survival function did not compromise the performance of the equations. The UKPDS (33) used the discrete time version of Gompertz regression model, which was chosen for its close fit to the UKPDS data and was motivated by the approximate multiplicative effect of age on risk observed in the study (33). In this model, the risk of CHD increases with age, and it increases even more per year after diagnosis of diabetes (33).

Survival models have a range of choices for their time scale: time on study, age (attained age, age at risk, age at onset), or calendar time. Most studies included in this review used time on study as their time scale. The only study where age was fitted as the timescale were the SCORE models. The author's motivation for this choice was that using age as the timescale allowed estimations to be made for the entire range of age observed in the study, and survival estimations were not limited by the actual duration of the study's follow up (29).

Eleven publications reported using Cox regression models (12, 23-28, 30-32, 34, 35); all Cox models had time on study as the timescale and fitted age at study entry as a covariate. In the NZDCS and UKPDS, the effect of age was modelled as two variables: 'age at diagnosis of diabetes' and 'duration of diabetes' (28). One publication did not report the type of the model used (26). The summary of model specification approaches is presented in Table 2.5.

Table 2.5. Model specification methods and validation approaches in the reported risk prediction models

Study	How missing data were handled	Predictor selection	Sex-specific model	Statistical model used	Additivity assumption tested	Non-linearity tested	Methods to assess model performance	Compared with other algorithms	Validation, internal
Framingham 1991	Complete case	Significance testing. First, covariates were chosen separately for each sex that appeared to model age well (e.g. quadratic age terms, interactions between age and covariates), then covariates representing additional risk factors were added.	No	Weibull, non-proportional hazard	Yes	Yes	C statistic	NR	NR
Framingham 1998	Complete case	Significance testing	Yes	Cox regression	Yes	Yes	C statistic	NR	NR
UKPDS 2001	Complete case	Significance testing, likelihood ratio tests	No	Discrete time Gompertz regression	Yes		Plotting modelled and observed survival rates	Framingham (as implemented by the Joint British Societies Cardiac Risk Assessor Computer Program)	NR
PROCAM 2002	NR	Significance testing	Yes	Cox regression	Yes	NR	ROC, calibration (Homer-Lemeshov X^2)	Framingham as presented in ATPIII	Cross-validation: data divided into 5 equal sets. 4 used to generate model, fifth to test performance. Repeated for every possible 4+1 combination.
ARIC 2003	Complete case	Significance testing. Age used in all models, but added after selecting all other variables. When selecting non-traditional risk factors, calculated increase in population attributable risk and % increase AUC after adding predictor.	Yes	Cox regression	NR	NR	ROC	Framingham, 1998	NR
SCORE 2003	Complete case	*A priori*	Yes	Weibull regression	NR	NR	ROC. Diagnostic performance assessed by positive clinical likelihood ratios for various thresholds of risk. Lin's concordance coefficient was used to measure concordance between risks estimated using cholesterol vs. cholesterol/HDL cholesterol ratio.	NR	NR
QRISK1 2007	Multiple imputation	*A priori.* Model testing BIC	Yes	Cox regression	Yes	Yes	ROC, calibration (deciles of predicted risk vs Kaplan-Meier). Also, by age band and quintile of Townsend score. D-statistic, $R^{2.}$ Proportions re-classified into higher and lower categories.	Framingham 1991, ASSIGN	Tested performance in validation dataset (1/3 of data).

QRISK2 2008	Multiple imputation	Model testing BIC	Yes	Cox regression	Yes	Yes	ROC, calibration (deciles of risk vs K-M estimates), Brier score, D-statistic, R^2 Proportion reclassified at 20% threshold and checked observed 10-yr risk in those reclassified.	QRISK 1, Framingham (NICE modification)	As above
Reynolds women 2007	NR	First step: Cox model with stepwise selection, multiple regression trees. Final selection: Model testing BIC	Yes	Cox regression	Yes	Yes	ROC, calibration Hosmer-Lemeshov X^2 (cats defined by 2% increments in predicted risk), Brier, entropy (a likelihood-based function for dichotomous outcomes for which smaller values indicate better fit); the Yates slope (the difference in predicted risk between cases and noncases for which larger values indicate better fit).	Framingham 1998, ATP-III	Tested performance prospectively in validation dataset (1/3 of data).
ASSIGN 2007	NR	Significance testing	Yes	Cox regression	Yes	NR	ROC, calibration (quintiles of risk), rank correlations, kappa statistics.	Framingham 1991	NR
Reynolds men 2008	Complete case	No variable selection: only known risk factors used. Compared two models: traditional; plus hsCRP and family history. Models compared via LRT and BIC	Yes	Not specified	NR	NR	ROC, calibration, reclassification	ATP-III	NR
Framingham 2008	Complete case	Significance testing	Yes	Cox regression	Yes Treated and untreated blood pressure in the model.	Yes	C statistic and calibration (Hosmer-Lemeshov X^2 with Kaplan-Meier estimates of observed incidence, risk deciles) Sensitivity and specificity of the top quintile of predicted risk. Net reclassification improvement.	Framingham, 1998	Bootstrap resampling.
NZDCS 2010	Replaced missing clinical data with those from previous visits within 2 years.	AIC, significance testing.	No	Cox regression	Yes	Yes	C statistic, calibration plots using pre-specified risk grouping.	Framingham, UKPDS.	Derivation cohort from north, validation from south of NZ (1/4rth the size of derivation cohort).
Pooled Cohorts, 2013	Complete case	Significance testing, NRI, IDI	Yes	Cox regression	Yes	Yes	C statistic, Hosmer-Lemeshov X^2 , calibration slope	NR	10x10 cross-validation technique
QRISK3 2017	Multiple imputation with chained equations	Part of predictors (those in QRISK2) pre-specified Additional variables retained if adjusted HRs <0.90 or >1.10 and p-value >=0.01	Yes	Cox regression	Yes	Yes	Harrell's C, D-statistic, R^2, calibration plots, % reclassified to high risk group (>=10%) using new models	QRISK2 2008	Randomly assigned ¾ of practices to derivation cohort, and ¼ of practices to validation cohort.

Missing values

Three out of fourteen studies did not report how missing data were handled (24, 29, 34). Six studies selected participants with no missing data on covariates (12, 23, 26, 27, 30, 33) thereby performing complete case analyses. One study excluded people who were missing information on SBP or DBP, smoking, total cholesterol, high-density lipoprotein or diabetes and assumed "no left ventricular hypertrophy" if missing (9). All three QRISK studies used multiple imputation procedures to replace missing values for SBP, blood lipids, smoking status, and BMI; in QRISK3, missing values for blood glucose and HbA1c were also imputed in some models (31, 32, 35). The ARIC study excluded participants with missing data on baseline history of CHD or on the risk factors considered in their models (25). The NZDCS supplemented missing clinical data with records up to two years before the baseline visit, and developed separate models that included blood pressure lowering treatment status using data from a sub-cohort of complete cases (28).

Variable selection and assessment of model performance

Since the multiple publications of Framingham study risk prediction models in the 1970s, 80s and 90s, there has been a general consensus that certain variables should be included in prediction models due to their established clinical significance. For this reason, in all of the included models some degree of *a priori* variable selection was made before initiating modelling. In some instances, the number of candidate predictors had to be restricted due to their unavailability in all participants. This was the case with the SCORE and the PCE studies which included multiple datasets from several countries/states and only included clinically significant variables that were uniformly defined and collected (12, 29). Only one study, the Reynolds Risk Score for women, reported using automated methods of selecting predictors (stepwise selection) (27).

The statistical approaches to variable selection included testing the significance of associations between predictors and outcome, model fit (e.g. maximum likelihood (ML), Akaike information criterion (AIC) or Bayesian information criterion (BIC)), discrimination (such as area under a receiver operating characteristic curve (AUROC) or concordance statistic), net reclassification improvement (NRI) and integrated discrimination improvement (IDI). Four papers reported variable selection by testing their statistical significance (23, 30, 33, 34). Two papers used discrimination methods: age-adjusted Cox model and its accompanying C-statistic to test for the relation between various independent variables and the outcome (24); and the

percentage increase in AUC to evaluate the effect of additional variables on prediction (25). Two publications selected variables by model fit (maximum likelihood method) (9, 33). Three publications combined several methods: QRISK1 and QRISK2 used Cox models to estimate coefficients and hazard ratios for each potential risk factor and compared models using the BIC; the NZDCS used statistical significance and AIC (28); the Reynolds risk score for women used Cox models with stepwise selection and additive regression trees as a first step and performed final selection by BIC (27); the PCEs study used a combination of statistical significance, NRI and IDI (12). The QRISK3 study included predictors from QRISK2 models *a priori*, and selected additional predictors by inspecting the magnitude of adjusted hazard ratios and their statistical significance (35).

Overall model performance, discrimination, calibration

Discrimination was the most commonly assessed aspect of models performance: all but one publication (33) reported using the C statistic to assess discrimination. The majority (11/14) of studies reported assessment of calibration, mostly by plotting observed incidence against predicted risk (12, 23, 26-28, 30-35). The Reynolds risk score for men and one of the Framingham study publications (23) calculated the net reclassification improvement. QRISK1 and QRISK2 publications compared the proportions of patients who would be classified into low and high risk categories in a validation cohort when applying their models or Framingham or ASSIGN equations; QRISK3 was compared with QRISK2 in the same way.

Only half of the studies reported their methods of internal validation (12, 27, 28, 31, 32, 34, 35). Five studies reported using cross-validation methods (i.e. dividing the study sample into a training set to train the model and a test set to evaluate it) (12, 31, 32, 34, 35). One study reported using temporal validation, when the development data are split into two parts: one part containing patients who joined the study early to develop models, and another part containing the patients who joined the study more recently to assess models' performance (26). A further study reported geographic validation, where the dataset was split into development and validation cohorts based on the geographic region of a patient's residence (28). Bootstrap resampling was used in another study (23), which is a method for estimating the sampling distribution of an estimator by resampling with replacement from the original sample (38).

Most of the models included in this review have been validated multiple times in independent datasets. However, a summary of the external validation approaches was out of scope for this review.

Table 2.5 summarises the model-building methods and approaches to internal validation and assessment of model performance reported in the publications.

Discussion

This chapter has summarised the range of approaches to developing CVD risk prediction equations used by research groups whose CVD risk prediction tools have been extensively published and have been included or recommended in clinical guidelines. It focused on identifying the range of analytical aspects of CVD prediction models and built on an existing systematic review to identify candidate studies. The included models were largely consistent on the timeframe of risk prediction, choice of the type of regression model, inclusion of a standard set of variables and stratifying data by sex. From the perspective of developing new risk prediction equations, there appears to be a recognised set of risk factors for inclusion in CVD prediction models, a preference for sex-specific models, and a general agreement on the use of proportional hazards models (mostly Cox or Weibull regression) and data obtained from cohort studies. Also, many of the more recent studies consider predictors related to ethnicity and socioeconomic status a necessary addition to models. The approaches to defining outcomes, accommodating effect modifications, non-linearity and methods for variable selection were more diverse.

Naturally, each of the studies had some limitations which may have influenced both their analytical approaches and the models' performance in new patients. Sample size may have precluded Framingham 1991 and UKPDS studies from building separate models for men and women. Studies like SCORE and PCE combined multiple cohorts, in order to improve the representativeness of the sample and generalisability of models. This created certain limitations; for example, the PCE models excluded some CVD subtypes from the endpoint definition due to substantial study-by-study variation in ascertainment, and geographic variation in use of cardiovascular procedures. Also, the PCEs had to dismiss a range of novel biomarkers because of their unavailability in some of the cohorts included in the study. Similarly, unavailability of some of the important predictors prevented inclusion of these in

other models, e.g. diabetes status was absent in SCORE models. This may have weakened the models' discriminative ability. Substantial missing data on important predictors (such as blood lipids in QRISK studies) meant that these had to be statistically imputed.

Several limitations may influence the generalisability of models: insufficient sociodemographic diversity as in the Framingham study which mostly included white and working/middle class Americans, or the PCEs inability to include some common US ethnicities (Hispanic-American, Asian-American, American-Indian) due to insufficient numbers. Similarly, using occupational cohort data, as in case of PROCAM and both Reynolds studies, meant that their findings may be influenced by the "healthy worker" bias and therefore compromise generalisability of these models to broader patient populations. Furthermore, using population from randomised clinical trials as in case of Reynolds study and the UKPDS, may impact their models' generalisability since RCTs include volunteers and often have highly selective inclusion criteria.

Substantial variation of CVD outcome definitions was observed in this review. This means that the scores from models with very different definitions of outcomes are not directly comparable. Such variation leads to large variation of absolute CVD risk estimates, which may in turn lead to variation in selection of high risk individuals, and therefore variation in treatment recommendations in clinical practice - depending on the model used and its "high risk" threshold (39). Models with more severe, and therefore less frequent outcomes will produce lower absolute risk scores, which may have an adverse effect on communication of risk to patients (39). A more comprehensive (with wider coverage of more manifestations of the underlying cardiovascular disease process) and a more standardised approach to defining outcomes would increase studies' power and would enable transparent assessment of statistical performance of different models and standardised evaluations of the health impact of risk-based preventive interventions.

Increased availability of large and recent population-based datasets with extensive sets of variables present new opportunities to explore prediction modelling. Large datasets, such as in the QRISK programme, which are derived from primary care electronic records linked to national datasets, enable the inclusion of more predictors, the exploration of novel biomarkers and generally place fewer restrictions on the modelling process. For example, instead of needing to address possible effect modifications in a global model, populations can be stratified

into clinically meaningful sub-populations at risk and separate models developed for each of the sub-populations. This can potentially improve the accuracy of prediction.

However, large datasets will also pose challenges, as the effect of a large sample size may preclude meaningful interpretation of statistical significance. Also, the increasing computerisation of health care means that data collected for purposes other than research are becoming available to researchers. These data may not be suitable for the research task in question or be of sufficient quality.

The strengths of this review are that it provides a detailed summary of the range of methods employed to develop cardiovascular risk prediction scores by highly competent research groups, and extracted the information on methods used to build prognostic models which should be reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) recommendations (3). This review has some limitations. It included only English-language publications that reported models developed in relatively similar contexts. Also, due to incomplete reporting, some relevant information on model development was not available.

Although the analytical approach to modelling was seldom completely transparent, the publications included in this review provided a valuable insight into the current context of CVD prognostic modelling at the beginning of this doctoral research. The TRIPOD recommendations (3, 21) also provided comprehensive guidance on reporting of prognostic models, along with the good practice advice regarding what should be done in prognostic modelling, and what should be avoided.

Summary

This review of analytical approaches to CVD prediction modelling has shown that approaches to building established CVD risk prediction tools varied, but also shared many common features. The limitations that earlier studies faced are not necessarily present in more recent larger cohort studies. This provides opportunities to investigate new approaches to cardiovascular risk prediction modelling.

Chapter 3. Methods of Prognostic Model-Building Using Survival Data

Introduction

Chapter 2 presented an overview of methodological approaches applied by established research groups to develop CVD risk prediction models. All the CVD prediction models considered in Chapter 2 were developed using survival data, which arise when a time is measured for each subject from a defined time point until the occurrence of a particular event. Chapter 3 will focus on statistical methods specific to survival data which were used to derive and evaluate risk prediction models in this thesis.

Methods used in prediction research vary depending on the research question. This thesis focuses on clinical prediction models estimating absolute risk of CVD for individual patients. Clinical prediction models are usually developed by selecting the relevant predictors and combining them statistically using regression analyses: logistic regression for binary outcomes and survival methods where the outcome is the time to event rather than an event itself. These models are inherently multivariable, since, as a rule, health professionals integrate several patient characteristics to estimate future risk (40).

Building a prediction model from a set of candidate predictors is a complex process and there is no one commonly agreed approach to model-building and evaluation. New ideas and methods keep emerging and approaches evolve in response to the changing context of prediction research. Because methodological issues in development, validation, and updating prediction models evolve, the methods used in prediction modelling need to be regularly revised to incorporate the new evidence (21).

The candidate's aim was to adopt a transparent and systematic approach to development and validation of CVD risk prediction models, based on the best current methodological recommendations and informed by the modelling strategies applied in CVD risk models published recently by established research groups. Based on the literature review (Chapter 2) and other articles on methodological approaches to clinical prediction modelling, a decision was made to follow the general approach outlined in a recent work by E. Steyerberg - "Clinical Prediction Models: a practical approach to development, validation, and updating" (38), which offered comprehensive hands-on guidance on analytical aspects of modelling, as well as to adhere to the TRIPOD reporting recommendations (3, 21). These publications provided an excellent coverage of the key issues that need to be dealt with during the development and validation of prediction models. However, while these sources were comprehensive, they were

incomplete in the detail required. The literature review (Chapter 2) helped fill in these gaps and other missing details were addressed through further reading of the modelling literature and taking advice from two co-supervisors of this project who are also biostatisticians.

Aims and objectives

The main aim of this chapter is to describe the statistical methods applied to develop the CVD risk prediction models and to provide rationale for the choices made. A second aim is to present the concept of validation of risk prediction models and the approaches to internal and external validation implemented in this thesis.

To achieve these aims, the objectives of this chapter are to:

- Describe the type of statistical model used, its main properties and the methods of testing models' validity.
- Describe the model-specification approach adopted in this thesis.
- Present the concepts of calibration, discrimination, explained variation and clinical utility, and the measures used the assess these.
- Describe the differences between internal and external validation.
- Describe methods of external validation of existing models used for comparative assessment of new model's performance.

Methods for building prognostic models

Cox regression

Cox regression models (41) (also known as proportional hazards models) were selected to develop the prediction models presented in this thesis. This type of model is the most common approach to analysis of right censored survival outcomes. Right censoring happens when an event has not occurred during the observed duration of follow up, and status beyond this time is unknown. Other model structures, including Weibull, exponential, and log-normal, were considered but none appeared to have a significant advantage over the Cox approach.

The Cox model specifies the hazard, defined as the risk that an event will occur in an interval after time *t* given that the patient has survived to time *t.* In general, the Cox model is written:

$$\lambda(t;i) = \lambda(t;0) \times \theta_i$$

where the parameter θ_i represents the rate ratio which compares the individual's rate $\lambda(t;i)$ with $\lambda(t;0)$ - the baseline hazard describing how the hazard changes over time at baseline levels of predictors. The assumption is that all time variation in the individual rate $\lambda(t;i)$ is captured by the baseline rate $\lambda(t;0)$.

An important characteristic of the Cox model is that both the individual and the baseline rates vary with time while their ratio is assumed to remain constant. Hence the rates for subject *i* are assumed to be proportional to the baseline rates at all times *t*, an assumption which is described as proportional hazards.

The predictive version of the Cox model can be expressed in terms of the survival function as

$$S(t \mid \chi_i, \beta) = S_0(t)^{\exp(\beta\chi)}$$

where $S(t \mid \chi_i)$ is the probability of surviving beyond time t given predictors χ_i, S_0 is the baseline survival function at time t, and $\beta\chi$ is the prognostic index, a linear combination of predictor values weighted by regression coefficients.

Model diagnostics

The validity of models may be assessed in a number of different ways. In this thesis, the focus was on testing the underlying statistical assumptions and investigating observations which were poorly represented by the model or had a relatively large influence.

The proportionality assumption was assessed using the global test based on scaled Shoenfeld residuals (42), and plotting log(-log(survival) versus log(time) for variables that appeared non-proportional. The assumption of linearity of continuous variables was assessed by plotting lowess (locally weighted scatterplot smoothers) plots of the Martingale residuals against the variable in question (43, 44). Plots of deviance residuals (a transformation of Martingale

residuals to make them symmetrical around zero) were examined to identify outliers (observations poorly predicted by a model).

Covariate profiles of influential observations were inspected and decisions made about deletion in each individual instance. Checking for observations with disproportionate influence was done using DFBETA (difference beta) (45) plots, which quantify the effect of removing an observation on the regression coefficient of a predictor. The DFBETA measures the influence on a coefficient-by-coefficient basis, and compares the estimated parameter, β, obtained from the full data, with estimated parameters β_i, obtained by fitting the model to the N-1 subjects remaining after the i^{th} subject is removed. If $\beta - \beta_i$ is close to 0, the i^{th} subject has little influence on the estimate.

Also, likelihood displacement values were used (46), which approximate what happens to the model log likelihood (to be precise, twice the log likelihood) when an observation is omitted. Unlike DFBETA, the likelihood displacement values allow the combined influence of all regressors in a model on each individual observation to be assessed.

Model specification

All analyses were stratified by sex as it is now the standard approach in cardiovascular risk research, particularly when large datasets are available.

The inclusion of variables was decided *a priori*, based on clinical knowledge and reports from previous research. Stepwise selection was not applied as it may result in the inclusion of unimportant variables, and exclusion of important ones, and therefore it is not generally recommended to subject clinically important predictors to this type of selection process (38).

All continuous variables were treated as continuous unless there were reasons to categorise. It is reasonable to expect the underlying relationship with outcomes to be smooth but not necessarily linear, and usually but not necessarily monotonic (47). Keeping variables continuous makes full use of the information at hand, when non-linearity is assessed and the relationship with outcomes is correctly modelled. Categories are often unnatural and unrealistic due to the assumption that there is a discontinuity in response when interval boundaries are crossed (48). Categorisation also assumes that the relationship between the predictor and the response is flat within intervals; this assumption is less reasonable than a linearity assumption in most cases (48). Keeping variables continuous also prevents the loss of power and loss of

precision: dichotomising, for example, is equivalent to losing a third of data, which substantially diminishes the ability to detect real relationships (49), and it can also increase the probability of false positive results (50). Besides, there is often an uncertainty about defining the cut points, as these are arbitrary and opinions might differ as to what they should be. Finally, the arbitrariness of categories can make comparison between studies difficult.

Due to assumptions of the underlying model structure, it is important to detect any strong non-linear relationships between predictors and the outcome. As a more specific and objective investigation of possible non-linear relationships between predictors and outcome, the fractional polynomial procedure was used as proposed by Royston and Sauerbrei (51). Fractional polynomials are formulated as a power transformation of a predictor x: x^p, where p is chosen from the set -2, -1, -0.5, 0, 0.5, 1, 2, 3 as proposed by Royston and Altman (52). This defines 8 shapes of the continuous covariates, including inverse (x^{-1}), log (x^0), square root ($x^{0.5}$), linear (x^1), squared (x^2) and cubic (x^3). Besides these 8 fractional polynomials (FP1) functions, an additional 28 FP2 functions can be considered of the form $x^{p1} + x^{p2}$, and, when p1=p2, another 8 FP2 functions can be defined as $x^p + x^p \log x$, which results in a total of 36 FP2 functions. The algorithm applies a special type of backward stepwise selection procedure to determine reasonable functional forms for each continuous predictor. FP1 functions are always monotonic (always have a positive or negative slope), while FP2 functions may be monotonic or unimodal (51).

To reduce the chance of overfitting the data, and to impose monotonicity, powers were restricted to the first order FP transformations (FP1) in all analyses. The FP algorithm was also used to check for continuous by continuous, and continuous by categorical interactions, while applying pre-specified non-linear transformations. This is important as the ability to detect an interaction involving continuous variables depends on correct modelling of the relationship between continuous variables and the outcome. That is, if the relationship with the outcome is non-linear but the variable is fitted as a linear term, this may impede the detection of an interaction.

The interaction terms identified by the FP analyses were only included if they met a strict pre-determined threshold of statistical significance, if the effect modification was clinically plausible, and if plotted data supported effect modification. The threshold was set at $p<0.001$ because of multiple testing and the large sample size.

The outcome used for the new models was the time to first CVD-related hospitalisation or CVD death. The follow up time was calculated from the index assessment to the first CVD event (non-fatal or fatal), or death due to other causes or end of follow up. As mentioned earlier, the time scale in Cox models can be the time on study, age (attained age, age at risk, age at onset), or the calendar time (more relevant if investigating a one-off exposure, e.g. industrial accidents). Time on study was chosen primarily because it is the most common approach in CVD prediction models, but also because much more is understood about the effect of age as a predictor than it is about the effect of other predictors at different ages – which would be the case if age were used as the time scale (personal communication T. Lumley 2016).

When applying published models to the PREDICT dataset (external validation), outcomes were defined in the same way as used in the derivation of that model. Cardiovascular risk factors may have different impacts on subtypes of CVD, and using a different definition of an outcome during validation would impede accurate assessment of model performance. Besides, comparing observed and expected events when the former and the latter have different definitions would be illogical (53).

By default, the baseline function of a Cox model corresponds to all covariates equalling zero, which is difficult to interpret and may lead to computational issues. For the baseline survival value to have a meaningful interpretation, it was instead estimated on the mean values of continuous variables and the reference values of categorical variables. As a result, the baseline represents a realistic "average patient" rather than a patient with impossible values of risk factors not represented in the data.

Absolute risk

The Cox model specifies that the risk to time *t* (in years), *R(t)* is of the form

$$R(t) = 1 - S_0(t)^{\exp(\beta X)}$$

Where $S_0(t)$ is the "baseline" survival function at time t.

To calculate 5-year absolute risk of CVD, coefficients from the Cox model were used as weights in the linear predictor (βX) component of the equation, which was then combined with

the baseline survival probability at 5 years that had been estimated at the mean values of continuous variables and the reference group of categorical variables.

Assessment of model performance (validation)

It is important in prediction modelling to evaluate the validity of a model, and obtain evidence regarding the accuracy of predictions for new patients who were not used to develop the model. Good performance at validation is a necessary requirement for a model to be recommended for use as a clinical decision aid. The validation process should ideally include both internal and external validation. Internal validation is the first step of model assessment and determines how well a model performs in the population it has been derived from. In addition to the standard measures of internal validity (discussed in the next section), it is common to randomly split the dataset into "derivation" and "validation" sub-cohorts, and apply the model developed in the derivation cohort to the validation cohort. However this approach has important limitations: the derivation and validation datasets are likely to be very similar, particularly in large datasets like the PREDICT study, and hence will produce an overly optimistic impression of a model's performance. In smaller datasets, estimates of model performance are also likely to be less precise as less data is used to derive the model. Methods such as bootstrapping have been proposed as better alternatives, particularly for the smaller-sized datasets. In contrast, with large samples such as in the PREDICT study, the standard measures of internal validity are already good indicators of model performance and it has been demonstrated that splitting a large cohort into derivation and validation sub-cohorts is unnecessary (38).

External validation, on the other hand, tests generalisability of a model: how well it performs in independent yet comparable populations. The optimal degree of similarity between the derivation and validation dataset is a matter of opinion, although certain dissimilarities, such as very different distributions of age, make it unreasonable to expect good performance if a model is tested in a new population. Therefore, distributions of known predictors need to be compared in the two datasets. However, if an unknown but important predictor is not included in a model, but the distribution of this predictor is very different in the derivation and validation dataset, this may also prevent the model from performing well during validation. Hence, even if a model is adequately developed, it may not perform well in a new population due to differences in characteristics of the patients in the two datasets (differences in case-mix).

Given the size of the PREDICT cohort, split sampling, to create derivation and validation sub-cohorts, was not performed as the main analyses, however it was undertaken as a sensitivity analysis. Rather than a random split, which would have created two very similar sub-cohorts and made the validation exercise less useful, a non-random geographic split based on the District Health Board (DHB) area of residence was applied, which resulted in two similarly sized cohorts with important non-random differences. The distributions of risk factors, outcomes and duration of follow up in both datasets were inspected, to ensure they were comparable in terms of patients and settings. The process of model-building and internal validation was then repeated by developing models in the derivation cohort and assessing their performance in the validation cohort, and compared with the results for the models developed in the complete cohort.

The differences in regression coefficients were small, as expected, and more stable using the complete cohort (38). Thus the final models, were derived from the complete cohort. Another reason for using the complete cohort to develop models was because it was largely representative of the New Zealand primary care patient population. Approximately 90% of all eligible patients in the study region had their CVD risk assessed and the study population included approximately one-third of all adult New Zealanders considered eligible for a CVD risk assessment by national guidelines. Unfortunately, a meaningful external validation of the new equations was not possible, as there are no other relevant New Zealand datasets of sufficient scale or representativeness.

Measures used to assess internal and external validity

A number of measures can be used to evaluate the internal and external validity of prediction models. These include calibration - the agreement between the predicted and observed outcome of interest; discrimination - the ability of a model to differentiate between high risk and low risk patients; and the overall predictive ability of a model – the explained variation. Another important feature of interest is the clinical utility of a model.

Calibration

The calibration aspect of model performance assesses agreement between the predicted outcome of interest and the observed outcome, usually for groups of patients.

The Hosmer-Lemeshow goodness-of-fit test is often reported as a measure of calibration yet it is no longer recommended (54) as it only provides a p-value for differences between observed

and predicted events in arbitrarily selected groups of patients and does not indicate the direction of any miscalibration. Moreover, this p-value would depend both on extent of any miscalibration and the sample size; it will almost certainly be significant if the sample is large, and non-significant if the sample is small (54). Thus, large sample sizes are likely to lead to null hypotheses of good calibration being rejected, even when calibration is very good.

In this project, the primary assessment of calibration was by grouping the estimated 5-year risks into deciles and plotting mean predictions from the model against the mean Kaplan-Meier survival probabilities (55) at 5 years for each decile (38). The Kaplan-Meier method is a non-parametric approach (i.e. doesn't make assumptions about the distribution of the survival or hazard curves) and is the most widely used method in medical research. Information from every individual is used, irrespective of whether they had the event of interest or not, and of the duration of follow up they contributed. The duration of follow up is split into intervals small enough to distinguish each individual event, and survival probabilities are calculated based on the intervals when an event occurs. The number of people at risk are adjusted at each interval for censored observations. This gives an accurate empirical description of the cumulative survival of the cohort and therefore is a suitable method to summarise the survival experience when assessing a model's calibration. The interpretation of the calibration plots is intuitive: if point estimates in the plot are scattered along the 45 degree line, the model is well-calibrated; if they are above or below the line, the model over- or underestimates risk, respectively. This type of calibration plot can be described as a visual representation of the Hosmer-Lemeshow test, because it categorises patients into groups according to predicted risk, however provides additional information about the direction of miscalibration (38).

Regression on the prognostic index

Another validation measure for survival models is the "calibration slope" (56), originally introduced for binary outcomes by Cox (57). It was further developed by Miller and Hui (58) and extended to survival outcomes by van Houwelingen (56). This measure is related to the average strength of the predictor effects in the model and assesses the degree of agreement between the observed and predicted values using a regression model. Cox, who proposed this measure for a logistic regression model, described it as a way of testing the agreement between an observed binary sequence and a corresponding sequence of probabilities (57). If the slope $\beta>1$ the probabilities from a model show the right general pattern of variation, but do not vary enough. If the slope $\beta<1$, the suggested probabilities vary too much (57).

The calibration slope is derived by fitting a regression model with the linear combination of predictors from the model (prognostic index) as the only predictor. In a logistic regression model, this will produce an intercept α and slope β. For an ideal plot showing a 45-degree line, α equals 0 and β equals 1. If the estimated slope β is smaller than 1, it indicates overfitting which means that predictions are too extreme: too low for low risk patients and too high for high risk patients. If the slope β is greater than1, it suggests the opposite: predictions are too high for low risk patients and too low for high risk patients (38). The intercept α assesses "calibration-in-the-large", or the overall agreement between the sum of all predicted probabilities and total number of observed outcomes. Provided the calibration slope is 1, if the intercept, α, is less than 0 this may indicate that the predicted probabilities are systematically too high, and if it is above 0, the predicted probabilities are systematically too low (38).

Interpretation of the calibration slope in a time-to-event (survival) model is essentially the same as for a logistic model, however as an intercept term cannot be directly obtained from a Cox model, "calibration-in-the-large" cannot be assessed in the same way (38) .

While calibration slope is often described as a measure of calibration (38), some authors have characterised it as a measure of discrimination and model fit (59). The rationale is that if the slope of the relationship between the prognostic index and outcome equals 1, it shows that the predicted probabilities vary, or spread, to the same degree as the observed event rates. If the slope is <1, probabilities vary too much, and if >1, probabilities don't vary enough. The degree to which predicted probabilities are spread out in relation to events is the fundamental definition of model discrimination, so it is not unreasonable to interpret the calibration slope as a measure of discrimination. From this perspective, if the slope of PI in the validation data is approximately 1, then discrimination of the tested model in the new dataset is about the same as in the derivation data. Such conclusions would be true under the condition that the case-mix in the two datasets is similar. However, if the prognosis of patients is less heterogeneous than in the derivation data, the spread of the PI will be less and this will impact the discrimination measures. If the slope is <1, the discrimination is poorer, and if it is >1, discrimination is better in the validation data than in the derivation data (59).

Regardless of how it is classified, the 'calibration' slope is a useful measure based on individual-by-individual comparisons of predicted probabilities with observed outcomes, and therefore this performance indicator was used in this thesis along with the other measures of calibration and discrimination.

Discrimination

For binary outcomes, the discrimination aspect of a model refers to its ability to separate patients having good outcomes from those having bad outcomes. A wider spread of the distribution of estimated risk is therefore favourable. This definition can be extended to survival data as the model's ability to separate those with longer event-free survival from those with a shorter event-free survival. The distinction between the binary and time-to-event context is important as the pairs of observations to be compared when calculating this index may vary, even within the same data.

Measures of discrimination based on concordance probability

The C-index (concordance index) quantifies the agreement between the ranking of the predicted and observed outcomes. For binary outcomes, the C-index is identical to the area under the receiver operating characteristic curve (AUROC) (60), which is a plot of the true-positive proportion (sensitivity) against the true-negative proportion (1 - specificity), evaluated at consecutive thresholds of the predicted probability. The AUROC represents the concordance, or agreement, of ranking between the predicted probabilities of having the event for a random pair of patients, one who experiences an event and one who doesn't. The value is the probability that the patient who has an event will have higher predicted probability than the patient who doesn't have an event. For a model with perfect discriminatory ability, this corresponds to sensitivity = 100% and specificity = 100% for each threshold value and AUC = 1. A value of AUROC = 0.5 indicates that model has no discriminatory ability.

For survival outcomes, concordance is measured between the observed and predicted orders of failure. Applying AUROC methodology to time-to-event data is difficult as it does not take into account that some observations are censored before the time point for which risk is to be calculated. To overcome this problem, extensions of the C-index or AUC have been developed by Harrell et al. (37, 61) and by Gönen and Heller (61, 62) as well as by other authors.

The C index proposed by Harrell et al. (37, 61) (known as Harrell's C), extends the AUROC to accommodate right censored outcomes and assesses the degree of agreement between predictions and outcomes comparing survival time not only in pairs of event and non-event, but also pairs of patients where both had events which happened at different times. The overall C index is calculated as the proportion of all usable pairs in which the predictions and outcomes agree, i.e. patients who develop an event first have a worse predicted prognosis. Only comparable pairs are used in the calculation of this index. That is, in each randomly selected

pair the survival times must not be identical; for censored observations, a pair is comparable if the shorter time corresponds to an event.

Gönen and Heller [48] suggested there could be bias in Harrell's C-index due to censoring, and developed a new measure of concordance probability, for use with the Cox model. They reversed the definition proposed by Harrell et al. (37, 61) and generated a measure based only on the regression coefficients and observed risk factor distributions. Therefore, Gönen and Heller's K index can be calculated using only Cox regression coefficients and the risk factor values. The actual survival experience is not directly used in this metric, instead it is used through the regression coefficients, which are obtained from the full survival data. In this way, the information contributed by the censored observations is also utilised.

Pencina et al. (63) pointed out that there were important differences in the definitions and properties of these two discrimination metrics. Namely, while both the binary C-statistic and Harrell's C compare how well the predictions agree with the observed survival experience, Gönen and Heller's K index calculates the probability that of any two patients, the one with a more adverse model-based risk profile will have a shorter survival time, i.e. how well the observed matches the predicted. All pairs of patients are used in the calculation of Gönen and Heller's K, including where both patients did not experience an event. Because of this, for Gönen and Heller's K index to reach 1, not only must the predicted probabilities be ordered according to decreasing survival times but also the duration of follow up needs to be sufficiently long for every patient to develop an event. That is, unlike for Harrell's C, a perfect ordering of events and perfect discrimination between events and non-events are not sufficient.

Pencina et al. (63) suggested that different concordance indices may be preferable in different situations. In prospective studies with lengthy follow-up, such as cardiovascular cohort studies focused on long-term risk prediction, not only avoiding the event but extending survival is important, and not all subjects experience the event of interest. In this context, discrimination is better captured by Harrell's C.

Measure of discrimination based on prognostic separation

Another discrimination measure quantifies the spread of the observed risks across the range of predicted risks and is based on the idea of separation of survival curves as a measure of prognostic information. Royston and Sauerbrei (64) proposed a measure of prognostic separation, called the separation statistic, or D statistic. The D statistic can be calculated by

first transforming the prognostic index derived from the model using Blom's approximation (65) to produce a standard normal order rank statistic Z; D is then the coefficient of Z in a model fitted with Z as the only predictor. For survival outcomes, the D statistic can be interpreted as the log hazard ratio between 'low' and 'high' risk groups obtained by dichotomising the predicted prognostic index at their median value. Higher values suggest improved discrimination and an increase of at least 0.1 indicates an important difference in prognostic separation between different risk scores (64).

Difference between calibration and discrimination

To be useful in clinical practice, it is essential for models to have both good discrimination and good calibration. However, good calibration of a model does not necessarily also mean there is good discrimination. For example, a model that predicts a similar risk for all patients at a level close to the incidence in the study population will be perfectly calibrated, yet it would discriminate poorly between events and non-events (54). The range of predictions will be very narrow in this hypothetical example, while for a good discrimination the range should be as broad as possible. In fact, ideally the predictions should be close to 100% for those who develop an event and close to 0% in those who do not.

It can be argued that inadequate discrimination is a more important problem, since poor calibration can be improved by recalibration (if the calibration plot is a straight line with different slope or intercept (56)), whereas nothing can be done about poor discrimination (59). Therefore, when developing prediction models, it is important to include performance measures that can identify poor discrimination (59).

Measure of explained variation: R^2

Measures of explained variation quantify the predictive ability of the variables in a model. The explained variation index is interpreted as the proportion of variation in the outcome that can be explained by predictors in a model. The more variability in the outcome that is explained, the better the predictive ability of a model. The value of R^2 ranges between 0% and 100%; a maximum value of 100% indicates that the variation in the outcome is completely explained by the predictors in a model, while the minimum value 0% indicates that none of the outcome is explained by the predictors.

This measure is known as a correlation coefficient in linear regression modelling where it measures how much of the variation in the ***y*** variable is accounted for by the linear relationship

with the ***x*** variable. The total variation in ***y*** can be thought of as the sum of the squared distances from the mean of ***y*** : $\sum(y-\bar{y})^2$.

The variation about the line is given by the sum of squared distances from each ***y*** to the corresponding, predicted value ***y′*** on the regression line: $\sum(y-y')^2$.

The difference between them is that part of the variation 'explained' by the linear regression of y on x. It can be shown that the proportion of the variation explained by the regression is equal to r^2 ,

$$r^2 = \sum(y-\bar{y})^2 - \sum(y-y')^2 / \sum(y-\bar{y})^2$$

The measure has been extended for logistic and time-to-event contexts. A study by Choodari-Oskooei et al (66) has identified 17 versions of R^2 for survival models (64, 67-80). It is not straightforward to immediately apply the classical definition of R^2 to survival models, particularly to the Cox model. The distinguishing feature of survival data is the presence of censored observations; also, the Cox model has no distributional assumptions and no error term. Various proposed measures apply different approaches to deal with these issues, which explains why so many measures have been suggested. Opinions differ on the most appropriate measure of explained variation in the censored survival data.

From the large spectrum of R^2-type measures in survival models, Choodari-Oskooei et al (66) identified the following categories: measures of explained variation; measures of explained randomness; measures of predictive accuracy. Among the three categories, explained variation measures can be described as the most intuitive because they are immediate extensions of the common interpretation of the R^2 in linear regression models: decomposition of the total variation into the variation explained by the model and the remaining unexplained variation, which gives an idea how much is known, and how much is unknown about a particular health phenomenon.

In the analyses presented in this thesis, the R^2 proposed by Royston and Sauerbrei (64) was used, which is a measure of explained variation on the log relative hazard scale based on the authors' D-statistic. It measures the relative gain in prognostic separation quantified by the D statistic and can be interpreted as the proportion of prognostic separation explained by the model. The conversion of D statistic to R^2_D is given by

$$R_D^2 = \frac{D^2 / k^2}{\sigma^2 + D^2 / k^2}$$

where σ^2 is defined as $\sigma^2 = \frac{\pi^2}{6} \cong 1.645$ and factor k is defined as $k = \sqrt{8/\pi} \cong 1.596$.

This measure has been shown to perform well in simulation studies as it was robust to influential observations and less affected by censoring compared with the alternative measures (66). R^2_D assumes the prognostic index is normally distributed; an assumption that can be checked by plotting the histogram of the prognostic index.

Assessment of clinical utility: decision curve analysis

Although important for assessment of statistical properties of a model, measures of calibration and discrimination do not provide a clear answer as to whether the model is worth using in clinical practice or which of two or more models is preferable (81). The 'net benefit' approach attempts to tackle this problem by incorporating clinical consequences of using a model into the model evaluation process (81). The net benefit measure was originally proposed by Pierce (82) and is widely applied in health economics. The key concept in net benefit analysis is the trade off, for example, between patients correctly treated (because they will otherwise become true cases within the predicted timeframe) and treated unnecessarily (because they will not). Net benefit is defined as $net\ benefit = \frac{true\ positives}{n} - \frac{false\ positives}{n}\left(\frac{p_t}{1-p_t}\right)$ where n is the total number of participants in the study and p_t is the threshold probability (82). The threshold probability of a disease or event is the risk threshold at which a patient would opt for treatment (or a clinician would advise treatment) and is informative of how the patient (or clinician) weighs the relative harms of a false-positive and a false-negative prediction (83). In any specific population, the net benefit can range between negative infinity and the incidence of a disease. A model being tested can be compared with two theoretical alternatives: 1) assume that no one has (or will develop) a disease and treat none; 2) assume that everyone has the disease (or will develop it) and treat everyone. In the former case, there will be zero true positives and zero false positives and therefore the net benefit will also be zero. In the latter case, the true positives will equal the number of events (incidence), and the false positives will equal to the number of non-events.

The decision curve analysis (DCA) proposed by Vickers and Elkin (83) built on the idea of net benefit and summarised the net benefit analysis into plots. DCA was originally developed for

binary data, and then extended to time to event analyses. This method assumes that the threshold probability of an event at which an intervention (treatment) is required, reflects how a clinician or a patient weighs the relative harm of false positive or false negative predictions. This theoretical relationship is then used to calculate and plot the net benefit of the model across different threshold probabilities in the form of "decision curves" (83). Higher net benefit over a clinically relevant range of thresholds implies that the model leads to better decision making.

To compare models presented in this thesis, decision curves were constructed by plotting the net benefit of the new models and the comparison models, against the 'treat none' and 'treat all' scenarios, at various threshold probabilities. These plots were then visually inspected to identify the model that provided maximal net benefit for a given threshold. In all analyses, the threshold probability represented the probability of experiencing a fatal or non-fatal CVD event in five years.

Justifying the need for new equations: validation and recalibration of existing models; evaluation of additional predictors

When presenting new prediction models, it is recommended that relevant existing models are also evaluated to determine how well they perform in the population of interest. If they perform poorly, and recalibration does not improve their performance sufficiently, then this provides additional support for the development of new models.

Recalibration is generally required when a model is evaluated in an external population with a different underlying event rate. When an outcome is binary and a model systematically predicts risk higher or lower than observed risk, adjustment of intercept without changing the weights of the risk factors is used to improve performance of the model in the new population (84, 85). When using survival models, updating the baseline survival function is equivalent to updating the intercept in the binary outcome situation. A new baseline survival function is estimated by fitting a Cox model with the prognostic index (linear predictor) from the tested model as a covariate, in the validation dataset. The preferred approach to recalibration is to fit the prognostic index as an offset term since it forces the weights for the risk factors in the score to remain the same (53).

The performance of two existing CVD risk prediction models (i.e. the 2013 American College of Cardiology/American Heart Association Pooled Cohort Equations (PCEs) (12) and The New Zealand Diabetes Cohort Study (NZDCS) equation (28)) were evaluated in the PREDICT

cohort, using the standard performance metrics described above. Each model was then recalibrated, as described above, by updating the baseline hazard estimated by fitting a Cox model with the prognostic index from the corresponding model as an offset term in PREDICT dataset (53) and then re-assessing calibration.

The additional predictive power of variables available in the PREDICT dataset that were not included in the PCEs or NZDCS models were also evaluated. This was done by fitting Cox regression models that include the prognostic index of the corresponding model (offset term) and the additional predictors; the effect estimates (hazard ratios and 95% confidence intervals) from these models were then inspected.

Discussion

In this thesis, the aim was to adopt an explicit and transparent model-building procedure based on current best-practice recommendations for prognostic/prediction model development. To ensure comparability of the new models with published influential studies, the methodological approaches adopted by established research groups developing CVD risk estimation models were particularly influential in the approach taken to model-building.

Appropriate design of a prediction study and an appropriate approach to developing and validating a model are the preconditions of the model's good performance. Therefore, when reporting a new model, it is important to provide a clear description of how the model was developed and validated. In view of the growing number of prediction studies and the lack of systematic approach to reporting them, the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guideline was developed to improve the reporting of prognostic/prediction models and enable informed critical appraisal of the model development procedure (21). Accordingly, as stated in the introduction to this chapter, the TRIPOD guidelines were followed in the reporting of the general population models and T2D specific models in Chapters 6 and 7.

In comparison to the methodology for predictive model development, the methods for assessment of their predictive performance are even less standardised. Yet because validation is such an important part of predictive modelling, full attention needs to be given to the measures employed in assessing a model's performance. Many performance measures used to evaluate the time-to-event predictive models were originally developed for binary outcomes and subsequently extended to survival analysis, dealing with the issue of censoring (i.e. missing

information on the outcomes due to discontinuing the follow-up before an event occurs) in different ways. Therefore, these is no consensus as to which of the adaptations is the most appropriate. When multiple options of these adaptations were available, the aim was to provide a rationale for why a particular option was chosen, based on recommendations in the literature.

Interpretability is one of the most important properties of a model's performance measures. Of the statistical measures discussed in this chapter, R^2 (the proportion of explained variation) and the D-statistic (the hazard ratio for high-risk vs low-risk groups) are the easiest to explain the meaning of to a general clinical audience. The C-statistic is a commonly used measure of discriminative ability and therefore is familiar to many health researchers. However, with the many different versions of C-statistic and R^2 measures available, one needs to be aware of the variation in their definitions and properties, since their suitability will depend on the context of a particular study.

Calibration plots are the most straightforward and clearly interpretable measures of model performance and are, in effect, a graphical representation of the Hosmer-Lemeshow test. Regression on the prognostic index is perhaps the least intuitive performance measure for a non-statistical audience. There is also a lack of agreement in the literature as to whether regression on the prognostic index (to derive the 'calibration slope') measures calibration or discrimination. Nevertheless, this measure provides information on how well a model predicts an outcome of interest in a new dataset and therefore should not be discarded.

Although both good discrimination and good calibration are required for a model to be clinically useful, metrics of discrimination and calibration are statistical abstractions that have no straightforward clinical interpretation. Yet the ultimate purpose of a prediction model is to help a health care professional to make better clinical decisions for an individual patient. Therefore measures of clinical utility are an additional way to assess a model's performance. The decision curve analysis approach described in this chapter is one such method which provides plots that compare models with each other as well as with the standard theoretical alternatives (i.e. 'treat none' or 'treat all').

Summary

This chapter has described the statistical methods used in this thesis for building prediction models with time-to-event outcomes. The chapter has also discussed approaches to validation,

the key aspects of models that should be evaluated when conducting a validation study, and the rationale for using specific measures of predictive performance for survival outcomes.

Chapter 4. Using Routinely-Collected Data to Build CVD Prediction Models: the PREDICT Cohort Study. Description of the PREDICT CVD-free and Diabetes Sub-Cohorts

Introduction

Chapter 2 summarised methodological approaches to building risk prediction models adopted by recognised research groups, based on published reports of their CVD risk prediction models. Chapter 3 discussed methods of prediction model-building using survival data. This chapter describes the origin and main characteristics of the PREDICT cohort study, and the data used for building the prediction models presented in this thesis.

Aims and objectives

The aim of this chapter is to describe the PREDICT cohort study and the main characteristics of its data.

To achieve this aim, the objectives of this chapter were to:

- Describe the background and design of the PREDICT cohort study.
- Describe how and what data were collected by the study.
- Describe how the data from primary care and from national/regional datasets were linked into one dataset.
- Describe the two sub-cohorts used to generate the models presented in this thesis.
- Discuss the strengths and limitations of the PREDICT cohort data.

This chapter follows the recommendations of the Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement (16). Additionally, because PREDICT is an observational study by design, this chapter also complies with the STROBE guidelines for reporting observational studies (86).

PREDICT study setting

The PREDICT-CVD project was established to provide guideline-based clinical decision support for assessment and management of CVD risk in routine general practice (10), and to simultaneously collect CVD-relevant data for prognostic research (87). Recruitment started in August 2002, when a web-based CVD risk assessment software (PREDICT) was introduced into participating Primary Health Organisations (PHOs), initially in the Auckland region and then in the Northland region of New Zealand. These PHOs provide primary health care services to approximately 80% of the two regions' populations, which constitute more than one-third of the total New Zealand population. The PREDICT software is integrated into patients' electronic health records (EHR) in general practice and includes a CVD risk prediction calculator that uses a modified version of one of the published Framingham risk prediction equations (9) to estimate a patient's 5-year risk of a CVD event. The NZ-Framingham modification adjusts for several additional factors recommended by New Zealand Guidelines, including ethnicity (Māori, Pacific and Indian), diabetes with microalbuminuria, and family history of premature CVD (10, 88). If one or more of these additional factors was present, a 5% absolute increment was added to the 5-year CVD risk calculated using the Framingham equation. The 5% was only added once, even if a patient has more than one of the additional risk factors.

Eligibility criteria and recruitment process for the PREDICT primary care cohort

Eligibility was defined by the NZ Guidelines for assessment and management of CVD risk (10, 88, 89). These guidelines recommended regular formal risk assessment at least every 5 years for all New Zealand men aged 45-74 years, all women aged 55-74 years, and starting 10 years earlier for patients of Māori, Pacific and Indian ethnicity and for those with known CVD risk factors (such as diabetes, current smokers, and those with high levels of blood pressure, blood lipids or who were obese). The frequency of recommended re-assessments was based on the patient's estimated risk and ranged from annually to five yearly. Patients who were pregnant or under 18 years old were excluded.

Patients were automatically recruited when a primary healthcare provider completed a standardised CVD risk assessment template using PREDICT-CVD software (templates in Appendix 4.1). Eligible patients were not required to be individually consented, although notices describing the PREDICT study were provided in the waiting rooms of all general

practice surgeries (approved by regional and national Ethics Committees). This meant that 100% of patients who completed an electronic CVD risk assessment using PREDICT software were recruited. With permission of primary care providers, the CVD risk factor profiles generated for each patient by the PREDICT-CVD software were stored in both the patient's EHR and on a secure third-party server (held by a private IT company, Enigma Solutions Ltd).

This thesis is based on patient data from risk assessments done between August 2002 and October 2015, however PREDICT is an open prospective cohort which is constantly growing. Figure 4.1 shows number of patients recruited each year into the total PREDICT cohort.

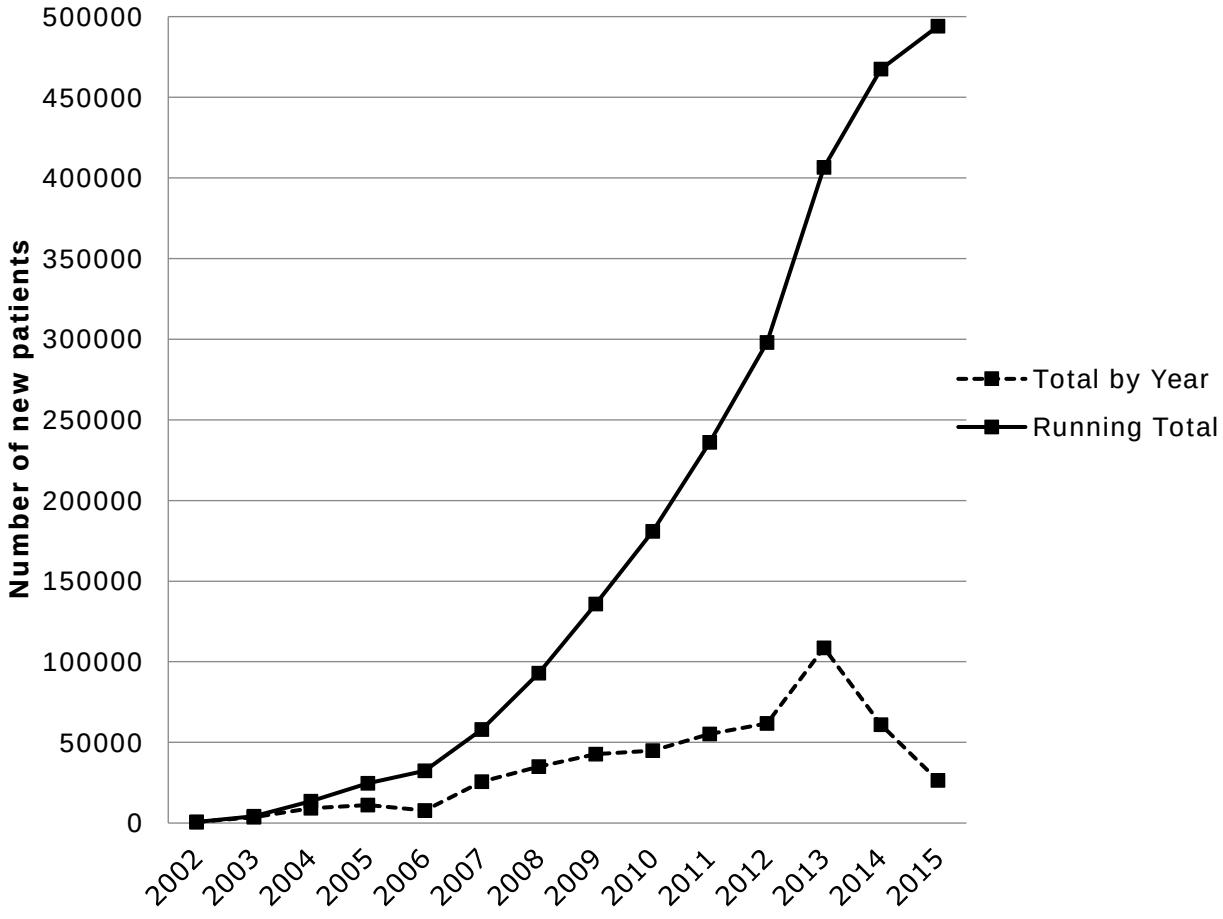


Figure 4.1. PREDICT cohort recruitment: running total and number of new patients aged 18-105, recruited by year

Figure 4.2 shows number of patients in the CVD-free and in the diabetes sub-cohorts recruited each year into the PREDICT cohort.

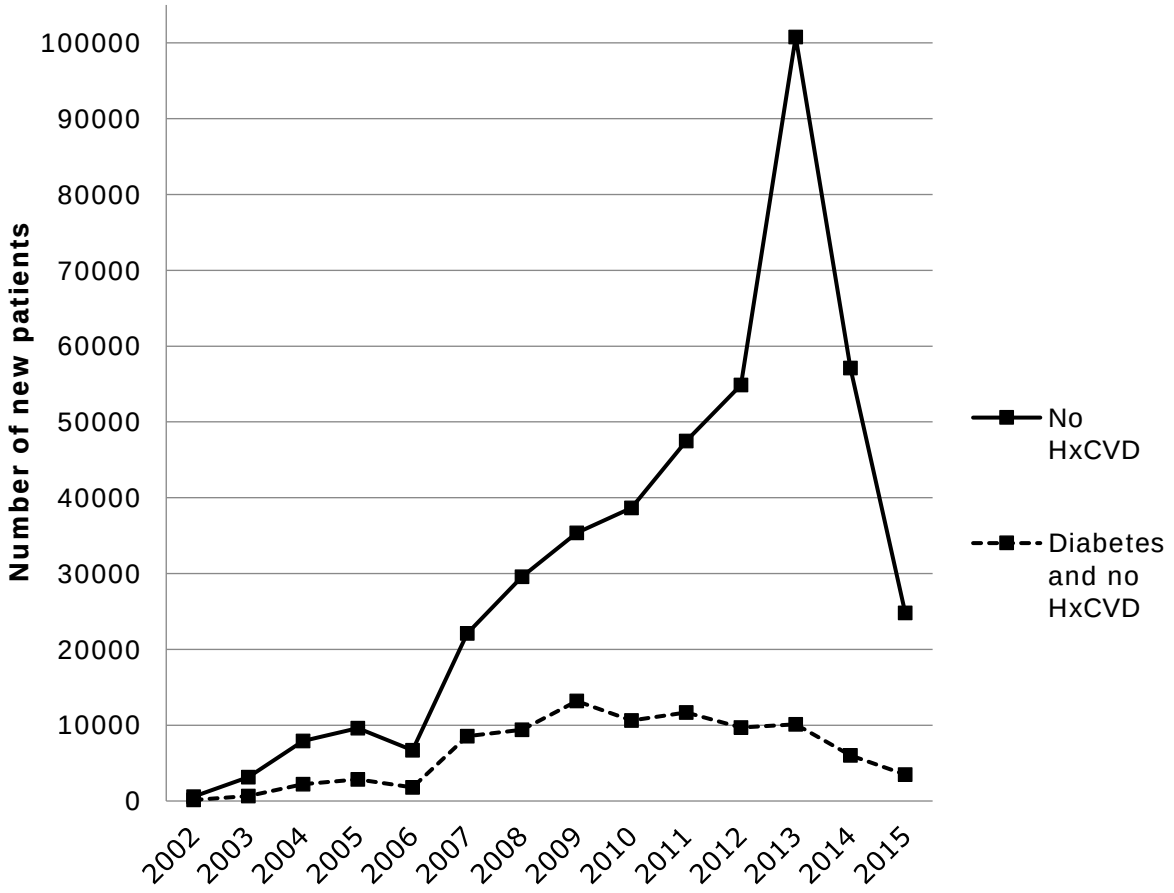


Figure 4.2. PREDICT cohort recruitment in the CVD-free and diabetes sub-cohorts: number of new patients recruited by year*

*The spike around the year 2013 on Figure 3.2 is due to increased CVD risk assessments in response to the national target set by the government. In 2011, CVD risk assessments and screening for diabetes were made a national health priority, with the goal of 60% of eligible adults to be screened by PHOs by 2012, 75% by July 2013 and 90% by July 2014 (90). According to CVD risk assessment performance, PHOs would receive modest incentive payments and would be benchmarked quarterly against others (91).

Prognostic factors and medical history: data collected in primary care

The CVD risk calculation in the PREDICT software could only be completed if all mandatory fields in the PREDICT risk assessment template were filled in. When the template was opened, the software attempted to autopopulate the required fields from a patient's EHR, which were then checked and corrected by the health care provider. When completed, the PREDICT template was submitted (example of a completed template page in Appendix 4.2), and the patient's estimated absolute risk was calculated and presented back within a few seconds

(example in Appendix 4.3). At the same time, the patient's risk profile was automatically stored both in the patient record and on the secure PREDICT server managed by Enigma Solutions Ltd.

The information collected on the PREDICT risk assessment template consisted of the risk factors included in the New Zealand version of the Framingham equation and some additional medical history and demographic information and was required for all participants. These were as follows:

- Age in years, derived from the patient's date of birth
- Sex
- Self-identified ethnicity
- History of CVD (including angina, myocardial infarction, percutaneous coronary intervention or coronary artery bypass grafting, ischaemic stroke or transient ischaemic attack, peripheral vascular disease or procedures)
- Electrocardiogram-confirmed atrial fibrillation
- Diabetes (type 1, type 2, type unknown)
- Family history of premature CVD (self-report, defined as ischaemic CVD occurring before 55 years of age in a first-degree male relative, or before 65 years of age in a first-degree female relative)
- Diagnosed genetic lipid disorder (familial hypercholesterolaemia, familial defective apolipoprotein B-100, familial combined dyslipidaemia, other genetic lipid disorder)
- Smoking status (no-never; no – quit over 12 months ago; no – quit less than 12 months ago; yes – up to 10/day; yes – 11-19/day; yes – 20+/day)
- Systolic and diastolic blood pressure in mmHg (mean of the two most recently recorded sitting blood pressure levels at the time of the index assessment, measured by either a general practitioner or practice nurse)

- Fasting or non-fasting TC:HDL, single measurement (the most recently recorded TC and HDL levels at the time of the index assessment, measured in community laboratories, and automatically downloaded into patient records).

Also, if a patient was diabetic, the year of diabetes diagnosis, Hb_{A1C}, and renal disease status (confirmed microalbuminuria, overt diabetic nephropathy, non-diabetic nephropathy) were recorded on the main template as well.

A second template, on CVD risk management, was only filled if the clinicians wanted the PREDICT software to generate individualised guideline-based recommendations for patient management. The additional variables on this template were therefore more likely to have missing data. Examples of these variables are height, weight and body mass index (BMI), waist circumference, triglyceride (fasting), treatment with CVD medications (aspirin, clopidogrel, warfarin, ACE inhibitor, angiotensin II receptor blocker, beta blocker, thiazide, calcium antagonist, statin, fibrate and other lipid lowering drugs). Some of these variables were more likely to be complete in specific subgroups of patients, e.g. BMI was nearly complete in those with diabetes.

A separate diabetes management template was automatically generated by the PREDICT software if a patient was recorded as having diabetes on the CVD risk assessment template. The diabetes management template had fields to record additional biomarkers and medical history only in those with diabetes, including serum creatinine, albumin to creatinine ratio, estimated glomerular filtration rate, Hb_{A1C} or fasting glucose, and diabetes medications (oral hypoglycaemic agents and insulin).

The PREDICT software was launched in 2002 without the diabetes management template which was added in 2004, after new CVD and diabetes guidelines were introduced (10). The earlier version of PREDICT CVD risk assessment template is provided in Appendix 4.4.

Outcomes and additional information on prognostic factors and clinical history: data linkage to the national and regional datasets

The CVD risk factor profiles stored on the PREDICT webserver were regularly linked, using an encrypted national identifier (the National Health Index [NHI]), to national health databases containing all public hospitalisations, mortality records, all medications dispensing by

community pharmacies, and all laboratory test results done in the Auckland and Northland regions, which are the main areas where PREDICT software is used. The NHI is used in all interactions with the publicly funded or subsidised health services in New Zealand and over 99% of New Zealand residents have a unique NHI number. All data used for these analyses were anonymised using encrypted NHIs (eNHIs) before being made available to the VIEW research team.

The first step in linking PREDICT data with national and regional health databases, was to discard any non-valid eNHIs from the PREDICT dataset. These constituted less than 0.1% of the study population. Then, the PREDICT data were linked with the National Health Index (NHI) dataset (92) in order to compare the PREDICT demographics (age and sex as recorded in primary care records) with those associated with the NHI recorded by the Ministry of Health. Any apparent mismatches were discarded from the dataset (0.03% of the cohort).

Linkage was then done with the National Minimum Dataset (NMDS), a national collection of public and private hospital discharge information, including coded clinical data for inpatients and day patients (93). Because it is not compulsory for private hospitals to submit their discharge information, it may be incomplete or delayed and therefore only public hospitalisation data were used for identifying prior ischaemic CVD and outcome ascertainment in this thesis. Only about 1% of cardiovascular events reported to the NMDS are from private hospitalisations.

The original NMDS was implemented in 1993, when data started being submitted electronically in an agreed format by public hospitals. The private hospital discharge information for publicly funded events, e.g., birth events and geriatric care, has been collected since 1997. Other data is being added as it becomes available electronically, with the current NMDS introduced in 1999. The NMDS has undergone many changes over the years, with some datasets removed and transformed into separate collections (e.g. the Mortality Collection and New Zealand Cancer Registry) and additional fields included to report the events in more detail (93).

In the PREDICT study, the NMDS data were used to identify CVD events occurring during follow up as well as to define a patient's history and co-morbidities at the time of CVD risk assessment (such as history of CVD, diabetes, atrial fibrillation, renal disease etc.).

Fatal CVD events were identified from the National Mortality Collection which classifies the underlying cause of death for all deaths registered in the country (94).

In New Zealand, every hospital codes their own discharges using the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM), the Australian Classification of Health Interventions (ACHI) and Australian Coding Standards (ACS). In contrast, mortality coding is completed centrally by the Ministry of Health, using the ICD-10-AM classification and the World Health Organization International Classification of Diseases Rules and Guidelines for Mortality Coding.

The PREDICT study uses an inclusive list of ICD-coded CVD outcomes. The primary outcome was a participant's first major CVD event defined by ICD-10-AM codes as a hospitalisation with or death from: ischaemic heart disease, including angina; ischaemic or haemorrhagic cerebrovascular events, transient ischaemic attacks; peripheral vascular disease, congestive heart failure, or other atherosclerotic CVD deaths. The hospitalisation outcomes included both primary and secondary diagnosis fields. In addition to deaths with CVD as an underlying cause, deaths were classified as due to CVD if they occurred within 28 days of a first CVD-related admission. Outcomes were not individually adjudicated.

Patients' risk profiles are additionally linked to the Pharmaceutical Claims Data Mart (PHARMS) (95), a national data warehouse that contains information on all publicly funded drugs dispensed from community pharmacies starting from 1 July 1992. This dataset was used to construct treatment variables and to supplement the information on co-morbidities (such as use of insulin as a marker for history of diabetes or loop diuretics for history of heart failure).

Further linkage is extended to data from TestSafe (96), a clinical information sharing service provided by the four Northern Region District Health Boards (DHBs), which contain data on all hospital and community diagnostic (laboratory & radiology) results and reports in the region. This source was used to supplement the information on relevant laboratory tests (e.g. blood lipids, glycaemic control measured by Hb_{A1C} and renal function tests such as serum creatinine or albumin to creatinine ratio etc.), when these were not available in some patients' PREDICT template records, or to crosscheck and validate the results recorded in primary care.

The PREDICT CVD-free sub-cohort, including patients with and without diabetes

The PREDICT participants included in the main analyses in this thesis were CVD-free, which was defined as "without a history of atherosclerotic CVD."

Participants were considered to have a history of atherosclerotic CVD if primary care clinicians recorded that they had prior angina, hospitalisation for ischaemic heart disease, transient ischaemic attack, cerebrovascular disease, or peripheral vascular disease on the PREDICT templates, or if they had prior ICD-10 AM coded atherosclerotic CVD-related hospitalisations (MI, angina, stroke, TIA, cardiac, cerebral or peripheral vascular procedures) recorded in the NMDS. They were also considered to have a history of atherosclerotic CVD if they had been dispensed at least one anti-anginal medication (glyceryl trinitrite, isosorbite dinitrate, isosorbide mononitrate, nicorandil, or perhexiline) on at least three occasions in the five years before index assessment (recorded in the national PHARMS database). Figure 4.3 presents components of the CVD events included in the history of atherosclerotic CVD definition.

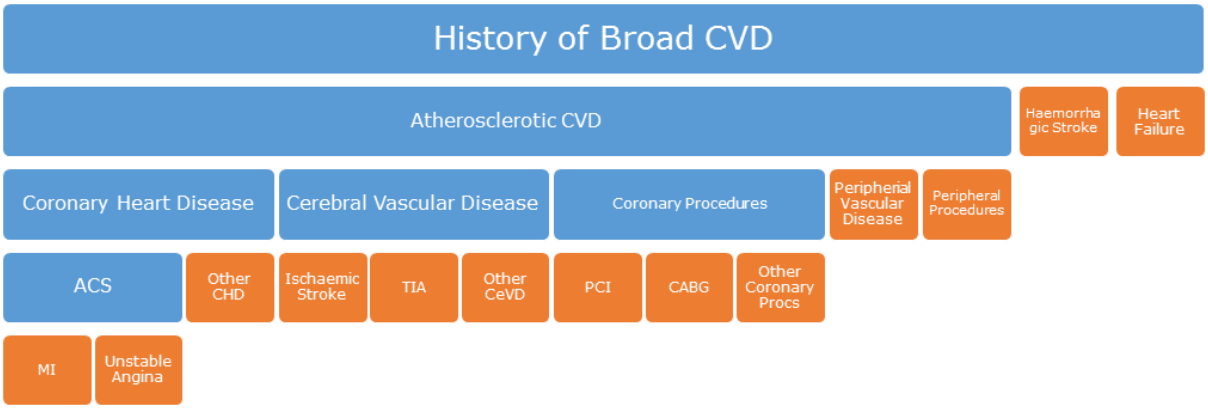


Figure 4.3 Definition of the history of atherosclerotic CVD in PREDICT study

Patients with atrial fibrillation (AF) were not excluded from study population as AF is not necessarily a consequence of atherosclerotic CVD.

New Zealand guidelines recommend that people with diabetes plus overt nephropathy are considered to be clinically at high risk and are therefore not eligible for calculated CVD risk assessment. These patients were also excluded from all analyses in this thesis if recorded on the PREDICT template as having an overt diabetic nephropathy or non-diabetic nephropathy, or had an eGFR<30 mL/min/1.73m^2 according to the laboratory records.

Table 4.1 summarizes the main demographic and clinical characteristics of the sub-cohort free of atherosclerotic CVD at the time of the first risk assessment between August 2002 and October 2015 (N=438,734).

Table 4.1. Medical history and CVD risk factors in the CVD-free PREDICT cohort, aged 18-105 years, August 2002 to October 2015

Variable	Total (n=438734) n (%)	Age group (%) <25 (n=1,281)	25-34 (n=7,214)	35-44 (n=60,614)	45-54 (n=149,845)	55-64 (n=135,789)	65-74 (n=66,068)	75-84 (n=15,226)	85+ (n=2,697)
Medical history:									
Family history CVD (yes)	50,811 (11.6)	11.1	16.8	13.7	12.1	11.4	9.4	7.2	4.6
Atrial fibrillation (yes)	8,864 (2.0)	0.2	0.3	0.5	1.0	1.8	4.2	9.0	16.5
Heart Failure	7,722 (1.8)	0.2	0.6	0.6	0.8	1.4	3.4	9.2	21.1
Diabetes	71,100 (16.2)	65.2	42.1	17.6	13.5	14.3	18.0	28.4	28.2
Risk factor:									
[1]Smoking									
Yes (current)	62,437 (14.2)	22.4	23.4	22.1	17.5	11.3	7.3	4.0	1.8
Past (former)	71,536 (16.3)	7.3	11.9	12.7	14.9	17.3	20.3	20.3	17.9
No (never)	304,759 (69.5)	70.3	64.7	65.2	67.5	71.4	72.4	75.7	80.3
[2]BMI									
Underweight/Normal (<25)	88,139 (20.1)	26.7	13.2	13.9	18.1	22.5	23.4	28.4	37.8
Overweight (25-29.9)	124,985(28.5)	22.0	23.0	27.1	28.7	28.2	30.0	31.4	27.9
Obese (30+)	131,048(29.9)	40.8	49.7	42.5	31.9	25.3	23.6	20.3	12.9
	mean (SD)								
[3]Systolic BP (mm Hg)	129 (16)	121 (14)	125 (15)	125 (15)	127 (15)	130 (16)	135 (16)	137 (16)	138 (17)
[4]Diastolic BP (mm Hg)	79 (9)	76 (10)	79 (10)	80 (10)	80 (10)	80 (9)	79 (9)	77 (8)	75 (9)
[5]Total cholesterol : HDL ratio	4.1 (1.2)	4.2 (1.5)	4.8 (1.5)	4.6 (1.4)	4.2 (1.2)	3.9 (1.2)	3.8 (1.1)	3.6 (1.0)	3.5 (1.0)
[6]Medications at baseline:									
BP-lowering drugs	108,489(24.7)	9.4	15.0	12.2	16.8	27.0	41.4	58.7	65.6
Lipid-lowering drugs	72,453 (16.5)	4.1	11.2	9.2	11.4	18.4	27.5	34.1	24.4
Antiplatelet/anticoagulants	51,400 (11.7)	1.3	4.4	4.6	6.7	12.2	22.5	36.5	42.6
[7]NZDEP quintile									
1 (least deprived)	95,679 (21.8)	14.8	11.5	12.6	21.1	25.6	25.2	22.6	21.3
2	85,596 (19.5)	14.6	14.1	14.3	19.2	21.5	21.4	20.6	18.0
3	79,105 (18.0)	16.3	16.5	16.7	17.5	18.6	19.4	18.9	20.5
4	81,741 (18.6)	21.9	23.7	22.0	18.4	17.4	17.8	19.6	21.7
5 (most deprived)	96,573 (22.0)	32.3	34.3	34.4	23.9	16.9	16.2	18.2	18.6
Ethnicity									
European	243,173 (55.4)	39.3	32.2	24.3	49.4	66.2	72.3	76.0	85.1
NZ Māori	54,937 (12.5)	23.8	20.1	23.1	15.4	8.2	6.2	4.7	2.5
Pacific	54,702 (12.5)	22.8	21.5	26.4	14.2	7.4	6.7	6.7	4.4
Indian	35,984 (8.2)	7.7	15.9	18.4	8.8	5.2	4.3	3.2	1.9
Chinese	24,899 (5.7)	2.1	3.5	1.6	4.8	7.8	7.2	6.3	3.8
Other Asian	17,841 (4.07)	2.7	4.1	4.6	5.2	3.7	2.3	2.1	1.9
[8]Other	7,198 (1.6)	1.6	2.8	1.7	2.1	1.4	1.1	1.0	0.4

[1]Missing values for 2 patients; [2]Missing values for 94,562 (21.6%) patients; [3]Missing values for 1 patient; [4]Missing values for 5 patients; [5]Missing values for 57 (0.01%) patients; [6]Medications categories are not mutually exclusive; [7] Missing values for 40 patients; [8]Includes Middle Eastern, Latin American, African and 'not specified".

Table 4.1 shows that few CVD risk assessments are done before the age of 34 years old. Also, the youngest patients have the highest proportions of diabetes and obesity, because adults with diabetes usually have a CVD risk assessment done in conjunction with the annual diabetes management review (97) and are included in the PREDICT cohort regardless of their age (89, 98). The group older than 75 years have the highest proportion of patients "on treatment" across all medications categories. They also have the largest share of those with the history of heart failure and atrial fibrillation and the lowest levels of obesity compared with all other age groups.

The proportions of Māori, Pacific and Indian patients are higher in the three youngest age groups which reflects adherence to national guidelines recommendation for earlier risk assessment as well as the younger age structure of these populations. Also, more than half the patients in the three younger age groups (<45 years old), reside in the most deprived areas (NZDep quintile 4 and 5) while those 75 years old and more were evenly distributed across NZDep quintiles.

The diabetes sub-cohort

Diabetes status was recorded on the PREDICT template as "Type 1", "Type 2" or "Type unknown". When defining the T2D sub-cohort, patients recorded as Type 1 were excluded but those recorded as "Type unknown" (<5% of all diabetics) were included based on clinical advice from diabetologists that most patients entered as "type unknown" were likely to have T2D (personal communication P.Drury, 2017).

The cohort of patients with Type 2 diabetes included those recorded as such by a primary care clinician, or recorded in the national hospitalisations database as discharged with one or more of the following ICD-10 codes: 25000, 25002, 25010, 25012, 25020, 25022, 25030, 25032, 25040, 25042, 25050, 25052, 25060, 25062, 25070, 25072, 25080, 25082, 25090, 25092, or ICD 10 codes E1153, E1165, E1102, E1112 - E1116, E1122, E1123, E1129, E1132 - E1136, E1139, E1142, E1143, E1149, E1152, E1159, E1162- E1164 E1169, E1172, E1173, E118, E119; or who were dispensed at least one diabetes medication (all subsidised forms of insulin and oral hypoglycaemic agents including acarbose, chlorpropramide, glibenclamide, gliclazide, glipizide, metformin, pioglitazone, rosiglitazone, tolazamide, tolbutamide) within the 6 months before the baseline risk assessment.

The demographic and clinical characteristics of all CVD-free primary care patients with diabetes (N=71,100) are presented in Table 4.2.

Table 4.2. Medical history and CVD risk factors in the CVD-free PREDICT cohort with diabetes, aged 18-95 years, Aug 2002 to Oct 2015

Variable	Total (n=71100) n (%)	Age group (%) <25 (n=1,281)	25-34 (n=7,214)	35-44 (n=60,614)	45-54 (n=149,845)	55-64 (n=135,789)	65-74 (n=66,068)	75-84 (n=15,226)	85+ (n=2,697)
Medical history:									
Family history CVD (yes)	7,603 (10.7)	6.7	9.2	12.0	12.2	10.9	8.9	7.1	4.9
Atrial fibrillation (yes)	2,035 (2.9)	0.4	0.3	0.9	1.4	2.6	5.1	9.9	15.0
Heart Failure	3,322 (4.7)	0.2	0.9	1.7	2.7	4.4	7.6	14.3	26.3
Diabetes type 1	3,009 (4.2)	50.5	17.2	5.9	3.2	2.5	1.9	1.5	1.3
Diabetes type 2	65052 (91.5)	45.6	78.6	90.1	92.4	93.2	93.7	94.5	94.0
Diabetes type unknown	3,039 (4.27)	3.8	4.2	4.0	4.4	4.3	4.3	4.0	4.7
Risk factor:									
[1]Smoking									
Yes (current)	10,830 (15.2)	20.6	25.7	22.8	19.0	13.1	7.5	4.0	2.2
Past (former)	12,408 (17.5)	8.3	12.6	14.2	15.5	18.5	21.8	23.1	19.6
No (never)	47,861 (67.3)	71.1	61.6	63.1	65.6	68.5	70.7	72.9	78.2
[2]BMI									
Underweight/Normal (<25)	9,894 (13.9)	30.2	12.3	9.3	10.5	14.0	17.8	24.5	35.6
Overweight (25-29.9)	19,363 (27.2)	24.6	20.0	21.2	24.7	28.7	32.3	37.5	36.4
Obese (30+)	39,175 (55.1)	44.0	66.0	66.8	61.3	53.1	45.3	33.5	23.1
[3]eGFR									
90+	32,658 (45.9)	94.3	87.2	74.6	58.9	38.4	15.4	1.6	0.8
60-89	28,550 (40.2)	4.9	9.8	19.6	33.5	48.7	60.4	56.7	38.8
45-59	4,691 (6.6)	0.2	0.6	0.8	2.0	5.9	14.6	25.0	30.6
30-44	1,642 (2.3)	0.1	0.2	0.3	0.7	1.7	4.3	10.8	20.5
<30	799 (1.1)	0	0.4	0.4	0.6	1.2	1.7	3.2	7.4
[4]ACR									
<3	34,870 (49.0)	62.3	53.4	46.0	45.9	49.6	53.6	51.0	46.9
3-30	14,093 (19.8)	24.1	26.4	20.9	18.3	18.1	19.2	25.9	33.3
>30	4,644 (6.5)	8.1	9.4	7.7	6.5	5.9	5.4	7.3	7.1
	mean (SD)								
Systolic BP (mm Hg)	133 (16)	120 (13)	125 (14)	128 (15)	132 (16)	134 (16)	136 (16)	137 (16)	137 (17)
Diastolic BP (mm Hg)	80 (10)	75 (10)	80 (10)	82 (10)	82 (10)	80 (9)	78 (9)	75 (9)	74 (9)
[5]Total cholesterol : HDL ratio	4.3 (1.3)	4.1 (1.5)	4.7 (1.5)	4.7 (1.5)	4.4 (1.3)	4.1 (1.3)	3.8 (1.2)	3.6 (1.1)	3.6 (1.1)
[6] Hb_{A1C}	61 (21)	78 (27)	71 (24)	66 (23)	63 (22)	59 (19)	55(15)	53 (12)	51 (11)
[7]Medications at baseline:									
BP-lowering drugs	37,150 (52.3)	12.9	22.6	32.0	40.5	50.8	55.5	56.4	42.8
Lipid-lowering drugs	31,531 (44.4)	6.0	11.2	9.2	11.4	18.4	27.5	34.1	24.4
Antiplatelet/anticoagulants	23,881 (33.6)	1.7	8.8	18.4	28.0	39.1	47.1	54.4	57.8
Insulin	7,350 (10.3)	58.8	25.7	11.7	8.4	9.2	8.2	8.2	5.9
Oral hypoglycaemic agents	36,041 (50.7)	30.3	52.7	49.3	47.9	52.4	53.8	54.1	46.3

As shown in Table 4.2, half of the patients below the age of 25 years old were diagnosed with type 1 diabetes, and the proportion of these sharply decreased with increasing age. Also, the proportions on insulin was highest in younger patients (<35 years old). The proportions of overweight or obese patients were high across all age groups, with the highest proportions in 25 to 74 year olds (ranging from 82% to 88%), and the lowest of 59% in those aged 85 years or more. Glomerular filtration rate physiologically decreases with age and this was evident with the eGFR levels decreasing with increasing age. Those over 74 years old had the highest proportions of eGFR under 30 mL/min/1.73m^2. Levels of proteinuria were slightly higher in patients younger than 34 years (possibly reflecting longstanding type 1 diabetes) and older than 74 years old.

Definition of the study populations and the naming conventions

The study populations used in this thesis were derived from the PREDICT CVD-free sub-cohort described above. National guidelines only provide specific recommendations for people aged 35 to 74 years and as a result, there were limited numbers and a possible lack of representativeness in younger and older age groups. Therefore, the initial plan was to restrict the study population to this age group (i.e. 35 to 74 years), however after requests from a number of general practitioners and cardiologists, the age group was changed to include people from age 30 years (i.e. 30 to 74 years). The small number of patients whose ethnicity was recorded as Middle Eastern, Latin American, African, or was unknown, were excluded.

The PREDICT-1^o study population (or PREDICT-1^o cohort) included individuals from the original PREDICT study, restricted to those aged 30-70 years with no history of CVD, heart failure, or significant renal disease, and no missing data on any of the pre-specified predictors (detailed in Chapter 5). The new CVD risk prediction models derived from PREDICT-1^o cohort will be referred to as PREDICT-1^o models.

The PREDICT-1^o T2D study population (or PREDICT-1^o T2D sub-cohort) was a subset of the PREDICT-1^o cohort restricted only to individuals with type 2 diabetes aged 30-70 years, with no history of CVD, heart failure, significant renal disease, history of renal dialysis or renal transplant, and no missing data on pre-specified predictors (detailed in Chapter 5). The new CVD risk prediction models derived from PREDICT-1^o T2D cohort will be referred to as PREDICT-1^o T2D models.

The PREDICT-1^{o} and PREDICT-1^{o} T2D study populations inclusion and exclusion criteria are presented in Figure 4.4.

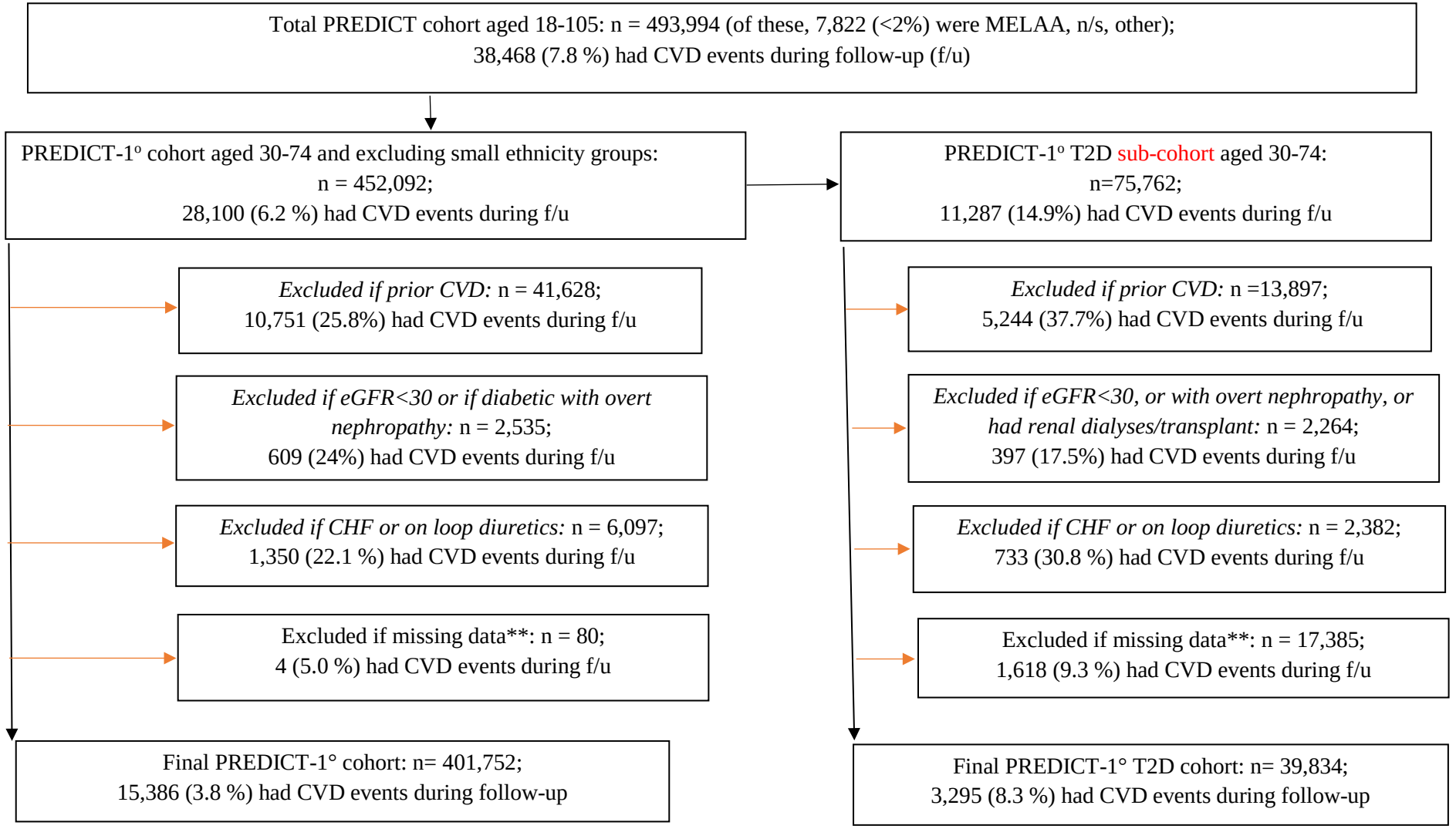


Figure 4.4 Flowchart of exclusions, PREDICT-1º and PREDICT-1º T2D cohorts

Ethical approval

Information about the PREDICT study is available in the waiting rooms at all participating general practices, and patients are advised they can request removal of their anonymised data from the study. In New Zealand, health information is available to researchers without individual patient consent provided the data are of a statistical and non-identifiable nature (87, 99). The PREDICT study was approved by the Northern Region Ethics Committee in 2003 (AKY/03/12/314) with subsequent annual approval by the National Multi Region Ethics Committee since 2007 (MEC07/19/EXP).

Implications of using routinely-collected data in PREDICT study

Routinely-collected data are defined as data collected without specific *a priori* research questions posed before the utilisation of the data for research. These data are typically generated during health care delivery, and through monitoring of health outcomes. The sources of routinely-collected data could be from clinical management (for example, primary care databases), health system planning (e.g. administrative health data), documentation of clinical care (e.g. electronic health record data repositories), or epidemiological surveillance (e.g. cancer registries and public health reporting data) (16). These data, generated in various health care settings and geographic locations, present opportunities for large-scale, innovative, efficient and cost-effective research to inform decisions in clinical medicine, health services planning and public health (100).

The broad spectrum of routinely-collected health data sources and the increasing use of these in research makes it challenging to identify the strengths, limitations and associated biases of individual data sources. Multiple errors and biases may interfere with routine data collection and processing (e.g., data linkage problems, misclassification bias and underreporting), additional to the usual sources of bias in observational studies, and can compromise the results of a study using routinely-collected data (101, 102).

Most of the data in the PREDICT study are either from routinely-collected sources (e.g. CVD events from hospitalisation and mortality records) or are collected by clinicians as part of a visit to their general practice in order to conduct a CVD risk assessment and use this as a basis for clinical management.

One of the main advantages of using routinely available data and/or data collected in routine practice is the potential to generate information on large and regionally or nationally representative populations. The composition of the PREDICT cohort was largely determined by national CVD risk assessment and management guidelines screening criteria which are described in the section "Eligibility Criteria and Recruitment Process for the PREDICT Primary Care Cohort". Whether a patient visiting a GP clinic is risk assessed or not, and therefore whether they enter the cohort, is at the discretion of the patient and their doctor or nurse. However, over the past decade, the Ministry of Health has provided additional funding to primary care to screen high-risk groups such as Māori, Pacific, Indian, those living in areas classified as NZDep quintile 5, and patients with diabetes (87). Also, national guidelines highlight the increased risk of CVD associated with Māori, Pacific, and Indian ethnicity, and recommend assessing patients from these ethnic groups 10 years earlier than other ethnic groups. The majority of people in these ethnic groups live in the upper half of the North Island of New Zealand, where PREDICT software is used. As a result, these ethnic groups are well-represented in the PREDICT cohort as illustrated by Poppe and colleagues (103) (Table 4.3 – included with permission of K.Poppe).

Table 4.3. Over-representation of high-risk ethnic groups in PREDICT cohort compared to their distribution in the general population*

	Aged 35-74 years		
	2013 NZ population* n (%) of total	**PREDICT cohort n (%) of total**	**PREDICT as % of NZ population**
All	2,090,970	446,961	21
Māori	235,860 (11%)	59,185 (13%)	25
Pacific	99,645 (5%)	56,147 (13%)	56
Indian	66,152 (3%)	37,170 (8%)	56
Asian	143,303 (7%)	42,934 (10%)	30
European/Caucasian	1,546,010 (74%)	251,525 (56%)	16

* NZ Ministry of Health population using a single prioritised ethnicity per person

Moreover, the Ministry of Health set a national target in 2011 to screen over 90% of all eligible adults by 2014, which meant that, while the assessments done before the target was set were more likely to include prioritised high risk patients, after the target was set, the cohort became increasingly representative of the total eligible New Zealand population. Further, if

participating patients changed health care providers or moved within New Zealand, they were not lost to follow-up since comprehensive national datasets were used to establish outcomes.

The outcomes in the PREDICT study were established through linkage to national databases of hospitalisations and deaths. The national morbidity and mortality datasets use the WHO-endorsed ICD-coding system and are subject to rigorous quality control. Still, some potential sources of misclassification of outcomes have been identified from research in other contexts. For example, when hospital reimbursement is based on complexity of case mix, hospitals may liberally apply more complex case mix to maximise reimbursement (104). An introduction of a provider billing incentive codes may also change the likelihood of a code being used over time (105, 106). Neither of these reimbursement-related sources of misclassification are believed to be important issues in New Zealand due to the way hospitals are funded. Changes in code classification systems, such as transition from ICD-9 to ICD-10 may alter validity of ascertainment using coded data (107, 108), however the PREDICT study was established after ICD-10 coding was introduced into New Zealand. While the accuracy of ICD coding for specific diagnoses can be unreliable, the broad definition of 'global' CVD as the outcome in the PREDICT study is likely to counterbalance these possible biases. High sensitivity and positive predictive values for ICD-coded CVD events in national datasets has been reported elsewhere (109). Also, a recent New Zealand study confirmed high level of capture of coronary intervention and associated Acute Coronary Syndromes in the All New Zealand Acute Coronary Syndrome-QI cardiac register and excellent agreement with national administrative datasets (110).

While the TestSafe dataset provided access to all laboratory tests results carried out in the Auckland and Northland regions, there are some issues associated with using this information. TestSafe data are aggregated from numerous laboratory facilities with variations in standards of recording the results, with some fields recorded as free text rather than numeric format which means these data require considerable data cleaning during linkage. Also, because the dataset includes all the laboratory tests performed in the region from a number of laboratory systems, there were multiple versions of tests for a particular biomarker. Therefore, decisions had to be made as to which of the tests would be appropriate for the particular research question. For example, there were several types of proteinuria test available in the TestSafe dataset, some of which were less appropriate for diabetes monitoring as they could be biased by other factors in a patient's clinical history: protein to creatinine ratio may be ordered instead of albumin to

creatinine ratio because of clinician's concern about possible non-albumin (tubular) proteinuria (personal communication Cam Kyle, 2017). The most relevant tests were selected in consultations with clinicians and chemical pathologists.

Another issue encountered with the laboratory data was the recording of some results using different units of measurement. For example, during the PREDICT study period, there was a change in the reporting units for Hb_{A1C} from percent to mmol/mol. This change was linked to the standardisation of routine assays for Hb_{A1C} to a new reference method (111). From 2009, both percent and mmol/mol were used in New Zealand when recording Hb_{A1C}, and since October 2011, reporting has completely switched to mmol/mol. This meant that both percent and mmol/mol records were present in TestSafe data and these had to be unified according to the current standard units of reporting.

The PHARMS data warehouse is jointly administered by the New Zealand Ministry of Health and the Pharmaceutical Management Agency of New Zealand, and collects data on government-subsidised medications dispensed by community pharmacies nation-wide (95). While dispensing data is a more accurate measure of use than prescribing data, no information was available on the actual use. The PHARMS data reliability and coverage has increased over time with reliable identification of dispensing episodes by NHI increasing from 64% in 2004 to 92% in 2006 to over 96% from 2010 onwards (87). Therefore, the patients assessed early in the evolution of PREDICT may have more missing information on their treatment status, although the proportion of patients assessed between 2002 and 2004 was relatively small.

The issue of unmeasured variables affects all observational studies, but can be particularly prominent when using routinely-collected data. The opportunities to collect additional data while using PREDICT software needed to be balanced with the existing workload of GPs and therefore some potential predictors, e.g. diet and physical activity, although present in the templates, were not mandatory fields and usually not recorded. Another type of unmeasured variable are 'unmeasurable' or 'difficult to measure variables.' For example, in chapter 6 of this thesis, the candidate found that patients in the PREDICT cohort on blood pressure lowering drugs had an increased risk of CVD, after accounting for blood pressure and other risk factors. This suggests confounding by indication, i.e. increased risk of an event associated with being treated due to more adverse risk profiles in those prescribed treatment. Certain unmeasured variables are related to a patient's characteristics that can influence the clinician's decision to prescribe medications but are seldom able to be adequately captured when specifically focusing

on one aspect of clinical care (CVD risk assessment variables vs the whole medical and social history of each patient).

Likewise, missing data can be problematic in all observation studies yet are particularly so in routinely-collected data as researchers do not control the collection of the data they use. Missing data can also result in selection bias if there are missing identifiers that prevent records from being linked, although this was not a significant issue for the PREDICT dataset. The PREDICT investigators provided clinicians with templates embedded in their EHRs, with automatic populating of fields where possible. Moreover, the software only calculated risk and recruited participants when all compulsory fields were completed, so there was no missing data on compulsory variables. However, some of the fields were not compulsory, and were therefore incomplete, and systematic differences between patients with complete and incomplete data are likely. An additional reason for missing values in the PREDICT primary care dataset is related to a change to the templates that took place in 2004. The original version of the template had fewer fields, so patients enrolled using this version of template were missing some information that was available on the updated templates (Appendix 4.4).

Discussion

Summary of the strengths and limitations of the PREDICT cohort dataset

In prediction modelling, the average effects of predictors identified in groups of patients from a study are used to make inferences about the risk for future patients in general populations of patients. Therefore, both representativeness and the quality of data used to generate predictive models are among the main prerequisites of their good performance, and are necessary preconditions for improving risk estimation accuracy. All datasets have certain weaknesses: random and systematic error, missing values, missing variables, to name a few. However, there is a big difference between using data carefully collected to answer specific research questions and using data collected for other purposes. As discussed above, administrative data and other data collected in routine clinical practice (e.g. the PREDICT CVD risk template data) are increasingly available for health research due to increasing computerisation of health care and this creates both opportunities (such as augmenting data and triangulation of information) and challenges (e.g. unknown sources of bias, distortions of information).

One of the main strengths of the PREDICT cohort is its size and representativeness. As discussed, over 90% of New Zealanders eligible for CVD risk assessments according to national guidelines were assessed between about 2010 and 2015 as a result of a funded national health target. With more than one-third of all New Zealand primary care practices using PREDICT software, the cohort is both large and representative of the relevant clinical population. The study population includes New Zealand's largest urban population (Auckland) but also significant rural populations in both the Auckland and Northland regions. All of New Zealand's diverse socioeconomic and ethnic groups are well represented in the study.

Another strength of the study is the contemporaneous nature of the information. The data linkage is updated every year, which means that new participants and recent event data are promptly included in the study.

A major advantage of PREDICT over other studies using data generated from routine practice was that the data collection procedure was prospectively designed specifically for deriving new CVD risk prediction models. While the main risk profile information came from routine clinical practice, this was the very data used by providers to calculate patients' CVD risk, and so is directly relevant to risk estimation context. The quality of data was likely to be reasonably high because the risk assessment software was integrated into the electronic health record, and data were directly read from the patient record into the risk assessment form. Also, automated validity checks and built-in range checks reduced input errors (87). Near 100% completeness of the main set of risk factors due to compulsory fields is quite unique for such a large study. In contrast, another large primary care-based cohort in the UK, QRISK, had a substantial proportion of missing data on some important risk factors, e.g. almost 80% missing a full lipid profile (112).

The national and regional datasets used to supplement and validate the CVD risk profiles also come from the same health care setting where the risk prediction models will be used. TestSafe data include all diagnostic laboratory results for primary care in the Auckland and Northland regions where most of the cohort lives. While PHARMS is an administrative dataset collected to support the management of pharmaceutical subsidies, all medications included in the PREDICT study analyses were publicly-subsidised. The ability to identify dispensing episodes recorded in PHARMS Data Mart using the NHI has increased from 64% in 2004 to nearly 100% from 2010 onwards (87).

The main limitation of the study is that some potentially useful risk factor data were either not compulsory fields, and therefore incomplete, or were not included. Examples of non-compulsory fields were BMI and waist circumference, markers of renal function and renal damage, and other lipid fractions (such as triglycerides, LDL). Fortunately, some of the incomplete data could be supplemented by linkage to the TestSafe laboratories dataset and using the information from patients' follow-up visits. Also, data on additional biomarkers that might be potentially useful for risk stratification of CVD (inflammatory and metabolic biomarkers, neurohormones etc.) could be obtained this way, although patients with these are unlikely to be representative of all primary care patients.

Of course, information in primary care records will inevitably have some errors as illustrated by Wells et al. with incomplete recording of prior cardiovascular disease at the time of CVD risk assessment (113). The reasons can be numerous, such as lack of time due to unrealistic workload in primary care and the deficiencies in design of health information sharing (113).

Another limitation is the loss to follow up for some participants. Although hospital admission and mortality records are comprehensive in New Zealand, if a participant leaves New Zealand, their outcome information is unlikely to be recorded in the PREDICT dataset, unless events are subsequently documented in primary care records or national datasets.

The quality of ethnicity coding is of relevance to risk CVD prediction, given the well-known differences in CVD risk by ethnicity. The national ethnicity coding system (114) currently only collects what is known as Level 2 ethnicity information. This allows for the specific identification of Indian and Chinese people, but aggregates other Indian subcontinent ethnic groups (Pakistani, Bangladeshi, Sri Lankan) and other Asian ethnicities, (e.g. Vietnamese, Malaysian, Thai, Korean, Japanese, Filipino etc.) in an 'Other Asian' category. Combining these heterogeneous groups into one category may disguise important information about ethnic differences in CVD risk. This issue might improve with the introduction of the new ethnicity coding protocols by the Ministry of Health (115) which reflect the requirement for information systems to capture the greater population diversity and improved granularity of information. The updated protocols support a transition from the previous minimum requirements of recording up to three ethnicities at level 2 classification to recording up to six ethnicities at level 4 classification.

Other instances of incomplete or otherwise suboptimal recording of information in primary care or national datasets, such as recording of family history of premature CVD, will be addressed by modifying the standard PREDICT risk assessment templates in the near future.

Summary

Despite some data limitations due to the PREDICT study's reliance on routine health data or data collected as part of routine clinical care, many of these potential limitations have been, or can be addressed through the design of the web-based PREDICT data templates, or through extending linkage to multiple national and regional datasets. As a result, the data used in the PREDICT study are highly relevant for cardiovascular prognostic research, in terms of sample size, representativeness, and the quality of data.

Chapter 5. Methodological Procedures Applied in Chapters 6-8

Introduction

In this thesis, uniform methodological procedures were applied when developing and accessing both the general population and the T2D-specific models, and when assessing performance of the existing alternative models in the PREDICT-1^{o} and PREDICT-1^{o} T2D study populations. To avoid repetition in Chapters 6-8, these procedures are described here, in Chapter 5.

Aims and objectives

The main aims of this chapter were, firstly, to describe the methodological procedures applied in development of new general population and T2D-specific CVD risk prediction models from the CVD-free PREDICT study cohort; and, secondly, to describe the external validation and recalibration of the existing alternative models in the PREDICT study population. These alternative models were the internationally recommended Pooled Cohort Equations (PCEs), which served as the comparison for the new PREDICT-1^{o} models, and the New Zealand Diabetes Cohort Study equations, the comparison for the PREDICT-1^{o} T2D models.

To achieve these aims, the objectives of this chapter were to:

- Describe the methodological procedure for developing new sex-specific models using data from patients in the PREDICT-1^{o} and PREDICT-1^{o} T2D cohorts.
- Describe the methods of internal validation of the new PREDICT-1^{o} and PREDICT-1^{o} T2D models.
- Describe the evaluation of the performance of PCEs and NZDCS models in the PREDICT-1^{o} and PREDICT-1^{o} T2D study populations.
- Describe the recalibration of the PCEs and NZDCS using the PREDICT-1^{o} and PREDICT-1^{o} T2D study populations.
- Describe the assessment of additional predictors which were included in PREDICT-1^{o} or PREDICT-1^{o} T2D models, but not included in PCEs or NZDCS models.

This chapter follows the recommendations of the Transparent Reporting of multivariable prediction models for Individual Prognosis Or Diagnosis (TRIPOD) guidelines (21) for reporting prediction model development and validation.

Sources of data

The PREDICT study is based on a large cohort of primary care patients opportunistically screened for CVD risk between August 2002 and October 2015. The full PREDICT cohort recruitment, data sources and linkage are described in detail in Chapter 4.

Participants, exclusion criteria

The participants included in the general primary prevention study (PREDICT-1^{o}) were primary care patients of European, Māori, Pacific, Indian, Chinese and other Asian ethnicities who had not experienced a CVD event, congestive heart failure (CHF) or significant renal disease prior to the index CVD risk assessment. The definitions of these exclusion criteria are described in Chapter 4.

Participants were also excluded if younger than 30 or older than 74 years and/or had missing data on any of the pre-specified predictors (age, ethnicity, NZDep, smoking, family history of premature CVD, systolic blood pressure, TC:HDL, history of diabetes, history of atrial fibrillation, blood pressure lowering, lipid lowering, or antithrombotic treatments at baseline).

The PREDICT-1^{o} T2D cohort was a subset of the PREDICT-1^{o} cohort, and consisted of patients with T2D. These were identified from at least one of the following: T2D or 'type unspecified' recorded in the PREDICT template (based on the primary care record); an ICD-10 code for diabetes type 2 recorded at any prior hospitalisation; dispensing of at least one diabetes medication within 6 months before baseline risk assessment. In addition to the PREDICT-1^{o} exclusions, T2D participants were also excluded if they had a history of renal dialysis or renal transplant, or had missing data on any of the pre-specified T2D-specific predictors (years since diagnosis of T2D, eGFR, albumin to creatinine ratio, HbA_{1c}, and body mass index).

Similarly to most modern cohorts used to derive CVD risk prediction equations, a substantial proportion of the patients included were receiving medical therapy at baseline and during follow up, for CVD risk and diabetes management.

Outcomes

The outcome of interest was the time from baseline risk assessment to first fatal or non-fatal CVD event. A CVD event was defined as a public hospitalisation or death associated with coronary heart disease (CHD), ischaemic or haemorrhagic stroke, transient ischaemic attack, peripheral vascular disease, or congestive heart failure, that occurred before 31 December 2015 (ICD-10-AM codes in Table 5.1). The follow up time was defined from index assessment till the first of the following: first CVD event, death due to other cause or extract date (December 31, 2015).

For the validation of the North American PCEs, a 5-year Atherosclerotic CVD (ASCVD) outcome was constructed in accordance with the ACC/AHA guideline committee definition and included: coronary death, fatal stroke, or the first nonfatal myocardial infarction or stroke. The relevant ICD-10-AM codes in Table 5.1, used to build the ASCVD outcome in PREDICT dataset, are highlighted in bold.

Table 5.1. International Classification of Disease-10-Australian Modification (ICD-10-AM) codes for total CVD events during follow-up, from hospital discharge and mortality records

Outcome type	**ICD-10-AM codes**
Myocardial infarction	**I210, I211 - I214, I219 - I221, I228, I229**
Unstable angina	I200
Other coronary heart disease	**I201, I208, I209, I230, I231, I232, I233, I234, I235, I236, I238, I240, I248, I249, I252, I253, I254, I255, I256, I460, I469,**
Ischaemic stroke	**I630 - I636, I638, I639, I64**
Haemorrhagic stroke	**I600 - I616, I618, I619**
Transient ischaemic attack	G450 - G453, G458 - G468
Peripheral vascular disease	E1050 - E1052, E1150 - E1152, E1451, E1452, I7021 - I7024, I7100 - I7103, I711, I713, I715, I718, I739 - I745, I748, I749
Congestive heart failure	I110, I130, I132, I50, I500, I501, I509
Other ischaemic CVD-related deaths	E1059, E1159, E1459, I250, I2510 - I2513, I252, I258, I259, I461, I650 - I653, I658 - I664, I668 - I670, I672, I690, I691, I693, I694, I698, I700, I701, I7020, I708, I709, I714, Z951, Z955, Z958, Z959.

* The codes in bold represent the Pooled Cohort Equations (PCEs) Atherosclerotic CVD and were used in the external validation of the PCEs. Both fatal and nonfatal events with myocardial infarction and stroke codes were included, but only fatal events with 'Other coronary heart disease' codes.

Predictors in PREDICT-1º and PREDICT-1º T2D models

Seven pre-specified risk factors in PREDICT-1º models were documented on the PREDICT template: sex, current age, ethnicity, family history of premature CVD, smoking status, systolic blood pressure (SBP), and total cholesterol to high-density lipoprotein ratio (TC:HDL). These were measured as follows:

- Age in years (continuous). This variable was derived from the participant's index assessment date and their date of birth. Date of birth is a component of the National Health Index (NHI) dataset on all New Zealanders and was automatically linked to the PREDICT dataset.

- Ethnicity was self-reported according to national categories (114). The ethnicity variable is a component of the NHI dataset on all New Zealanders and was linked to the PREDICT dataset. Because a patient is able to select several ethnicities, the following prioritisation order was applied: Māori, Pacific, Indian, Chinese, Other Asian, European, Middle Eastern, Latin American, African (MELAA), other/unknown. The categories used in the analyses were European, Māori, Pacific, Indian, Chinese/Other Asian. Due to small numbers and heterogeneity of the category, those classified as MEELA and other/unknown, were excluded.

- Family history of premature CVD in a first-degree male relative before the age of 55 years or a first-degree female relative before the age of 65 years (no, yes). This variable was recorded by the health professional completing the electronic form, based on the above definition which was stated on the form.

- Smoking status (three categories: never smoker, ex-smoker, current smoker). This variable was recorded by the health professional completing the electronic form. The categories provided were: no-never; no – quit over 12 months ago; no – quit less than 12 months ago; yes – up to 10/day; yes – 11-19/day; yes – 20+/day. These were combined into three categories based on preliminary risk models.

- Systolic blood pressure (SBP) in mmHg (mean of two measures, continuous). These were the two most recently recorded sitting blood pressure levels at the time of the index assessment, measured by either a general practitioner or practice nurse.

- Total cholesterol to High Density Lipoprotein cholesterol (TC:HDL) ratio (one fasting or non-fasting measure, continuous). These are the most recently recorded TC and HDL levels at the time of the index assessment. They are measured in community laboratories and are automatically downloaded into patient records.

Other risk predictors, such as diabetes, history of atrial fibrillation, treatment with blood pressure lowering and lipid lowering medications, antithrombotics, and the diabetes-related treatments for the T2D models, were generated using both primary care and national records.

- Diabetes was recorded by the health professional completing the PREDICT templates. The categories provided were: none; Type 1; Type 2 (including Type 2 on insulin); Type unknown; current gestational diabetes. Participants with current gestational diabetes were excluded from all analyses. In addition, if participants had been previously hospitalised with diabetes (ICD-10 codes specified in Chapter 4) or were dispensed diabetic medications within 6 months prior to the index assessment (specified in Chapter 4), they were also classified as having diabetes.

- History of atrial fibrillation was recorded by the health professional completing the PREDICT templates (ECG confirmed AF) and/or in the linked hospital records (clinical code I48).

- Preventive drug treatments were defined as binary indicators (yes/no) using records from primary care at index assessment and from national data, where treatment was defined as at least one dispensing in the 6 months before baseline of blood pressure lowering drugs (Table 5.2a), lipid lowering drugs (Table 5.2b), or antithrombotics (Table 5.2c). Diabetes treatments were all subsidised forms of insulin and oral hypoglycaemic agents including acarbose, chlorpropramide, glibenclamide, gliclazide, glipizide, metformin, pioglitazone, rosiglitazone, tolazamide, tolbutamide.

Table 5.2a. Medications included in blood pressure lowering treatment variable

	Class of blood pressure-lowering drugs*					
	ACE inhibitor	**ARB**	**Beta blocker**	**CCB**	**Other**	**Thiazide**
Medication name	Benazepril Captopril Cilazapril Enalapril Lisinopril Perindopril Quinapril Trandolapril	Candesartan Losartan	Acebutolol Alprenolol Atenolol Bisoprolol Carvedilol Celiprolol Labetalol Metoprolol Nadolol Oxprenolol Pindolol Propranolol Sotalol Timolol	Amlodipine Diltiazem Felodipine Isradipine Nifedipine Verapamil	Amiloride Clonidine Clopamide Hydralazine Methyldopa Triamterene	Bendrofluazide Chlorthalidone Chlorothiazide Cyclopenthiazide Hydrochlorothiazide Indapamide Methyclothiazide

* Alpha blockers, loop diuretics (bumetanide, frusemide), metolazone and spironolactone excluded as primary indication not usually to reduce blood pressure

Table 5.2b. Medications included in lipid lowering treatment variable

	Class of lipid lowering treatment	
	Statin	**Other lipid lowering treatment**
Medication name	Statin	Atorvastatin Fluvastatin Pravastatin Simvastatin Acipimox Bezafibrate Cholestyramine Clofibrate Colestipol Ezetimibe Ezetimibe with simvastatin Gemfibrozil Nicotinic acid

Table 5.2c. Medications included in antithrombotic treatment variable

	Class of antithrombotic drugs	
	Antiplatelets	**Anticoagulants**
Medication name	Aspirin Clopidogrel Dipyridamole Prasugrel Ticagrelor Ticlopidine	Dabigatran Phenindione Rivaroxaban Warfarin

Level of socioeconomic deprivation was derived from the National Health Index dataset. It was represented by an area-based deprivation score constructed from the data collected during the 5-yearly national census (116). A new NZDep Index is developed following each New Zealand census, and PREDICT participants were allocated a NZDep score derived from the NZDep Index based on the census closest to their index assessment (i.e. NZDep 2001, 2006 or 2013). NZDep2013 combines nine variables from the 2013 census reflecting eight dimensions of deprivation and provides a deprivation score for each meshblock in New Zealand. Meshblocks are geographic units containing a median of 81 people in 2013. The dimensions of deprivation included are listed below in order of decreasing weight in the index (117):

- Communication: People aged < 65 years with no access to the Internet at home
- Income: People aged 18 – 64 years receiving a means tested benefit
- Income: People living in equalised (i.e. controlled for household composition) households with income below an income threshold
- Employment: People aged 18 – 64 years unemployed
- Qualifications: People aged 18 – 64 years without any qualifications
- Owned home: People not living in own home
- Support: People aged < 65 years living in a single parent family
- Living space: People living in equalised households below a bedroom occupancy threshold
- Transport: People with no access to a car

The index is a continuous measure that is typically expressed as an ordinal score. The index was derived via first principal component analysis which has been scaled to have a mean of 1000 index points and a standard deviation 100 index points. This score is also divided into deciles, where decile 1 represents residences in the 10% of the least deprived areas in NZ and decile 10 represents residences in the 10 % of the most deprived areas. In clinical and population health, these deciles are typically aggregated into quintiles, as represented in the models described in this thesis.

In the PREDICT-1º T2D models, the risk predictors described above were included, along with additional, diabetes-specific risk predictors generated using both primary care and national records. These were: time since diabetes diagnosis, treatment with insulin and oral hypoglycaemic agents (as specified above), estimated glomerular filtration rate (eGFR), albumin to creatinine ratio (ACR), and haemoglobin A_{1C} (HbA_{1c}).

The eGFR and ACR was based on the most recent creatinine measurement recorded before baseline risk assessment, or, if none were available, up to 18 months after baseline assessment. The eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation (118) as recommended in the 2013 guidelines from the Kidney Disease: Improving Global Outcomes Foundation (KDIGO) (119).

The HbA_{1c} result was the most recent recorded before baseline risk assessment, or, if none were available, up to 14 days after baseline. If a patient didn't have at least one result in this period, but was not on blood glucose-lowering drugs (insulin or oral hypoglycaemic), an HbA_{1c} was included up to 18 months post baseline.

Missing data

In the PREDICT-1º population, there were virtually no patients with missing data on the eight mandatory variables required to predict CVD risk using PREDICT software. A very small proportion of patients were not represented in the National Health Index dataset, and therefore had their socioeconomic status information missing. Observations with missing data were excluded (N=80).

In the PREDICT-1º T2D population, missing data were also present in measurements of HbA_{1c}, ACR, eGFR, and years since diagnosis of diabetes, and observations where these were missing were excluded as well. To evaluate whether patients with missing data on ACR, eGFR and years since diagnosis of diabetes are different from patients with complete data, sensitivity analyses were performed by fitting models with these predictors categorised into clinically meaningful groups. The categories were as follows: ACR <3, 3-30, >30, unknown; eGFR >=90, 60-89, 45-59, 30-44, unknown; years since diagnosis <=1, 2-5, >5, unknown. These models were assessed using standard performance indicators applied to all other models in this study, and compared with the models from the complete case analysis.

Statistical analyses

Model derivation

Cox regression models were used to estimate beta coefficients and the baseline survival function at 5 years (41). Model diagnostics included testing the proportionality assumption using the global test based on scaled Shoenfeld residuals (42), and plotting log(-log(survival) versus log(time) for variables that appeared non-proportional; inspecting non-linearity of covariates by plotting LOWESS smoothing of Martingale residuals versus continuous covariates (43, 44); and checking for patients with disproportionate influence using delta beta (DFBETA) plots (45) and likelihood displacement values.

Predictors were selected *a priori*, based on clinical relevance and previous research. Continuous variables were not categorised if possible, as there are several issues associated with categorisation, as described in Chapter 3.

Nonlinearity of continuous variables and interactions between continuous and categorical variables were investigated using the fractional polynomials (FP) procedure proposed by Royston (51). To reduce the chance of overfitting the data, and to impose monotonicity, powers were restricted to the first order FP transformations (FP1). Also, the FP algorithm was used to check for continuous by continuous, and continuous by categorical interactions, while applying the non-linear transformations.

All analyses were stratified by sex.

Absolute risk calculations

The Cox model specifies that the risk to time *t (in years)*, *R(t)*, be of the form

$$R(t) = 1 - S_0(t)^{\exp(\beta X)}$$

Where $S_0(t)$ is the "baseline" survival function at time t.

To calculate absolute 5-year risk of CVD, the coefficients from the Cox models were used as weights which were combined with baseline survival values at 5 years estimated at the mean values of continuous variables and the reference groups of categorical variables.

Assessment of model performance

As discussed in detail in Chapter 3, the key aspects of a risk prediction model that need to be assessed are discrimination (ability to distinguish between high risk and low risk patients), calibration (agreement between the observed and predicted outcome) and explained variation (the proportion of variation in the outcome that can be explained by the predictors in the model).

The discrimination metrics used were Harrell's C-index of concordance, which can be interpreted as the proportion of all pairs of patients where a patient with higher predicted survival is the one who survived longer (61) and the D statistic, a measure of prognostic separation developed specifically for censored survival data, where higher values suggest improved discrimination and an increase of at least 0.1 indicates an important difference in prognostic separation between different risk scores (64).

Calibration was assessed in two ways – graphically and in terms of the "calibration slope" (57, 120). For the graphs, the estimated 5-year risk was grouped into deciles and the mean value in each decile plotted against the mean Kaplan-Meier survival probabilities in that decile (55). To derive the "calibration slope", the prognostic index (linear combination of the predictors weighted by estimated regression coefficients) derived from the model was entered as a single predictor in a Cox model predicting the combined CVD outcomes of interest (53). The Hosmer-Lemeshow goodness-of-fit test was not used since, as discussed in Chapter 3, this metric is not suitable for a large study.

Explained variation was assessed using Royston's (72) modification of O'Quigley, Xu & Stare's (74) modification of Nagelkerke's (73) R-squared statistic for censored survival data, as discussed in Chapter 3.

To assess clinical utility across various risk thresholds for intervention, decision curve analysis was performed as proposed by Vickers et al. (83).

Internal validation using split cohorts

Internal validation using split cohorts was undertaken as a sensitivity analysis only. The study population was divided into two geographically defined sub-cohorts based on the District Health Board (DHB) area in which participants lived. Auckland and Counties Manukau DHBs

formed the derivation sub-cohort and Waitemata and Northland DHB formed the validation sub-cohort. Prediction models were developed using methods described above, in the derivation cohort, and tested models' performance in the validation cohort using methods described above for assessing the performance of the new models.

External validation

An external validation was not undertaken because the PREDICT-CVD cohort is largely representative of the New Zealand population eligible for CVD risk assessment according to national guidelines, and there are no other New Zealand datasets of sufficient size containing all required information.

Validation and recalibration of PCEs and NZDCS models in the PREDICT cohort

Before developing new prediction models, it is generally recommended that researchers investigate whether there are existing models which perform well in the population of interest (3, 21). This is of particular relevance when the available study population is smaller than the studies on which existing models are based because new models may be more prone to random error than existing models. However, the PREDICT study population was much larger than the Framingham study populations, and it had already been determined, in an earlier extract from the PREDICT study, that the NZ-modified Framingham Heart Study equation overestimated risk in the general PREDICT population, but appeared to underestimate risk among high-risk ethnic groups (121). Therefore it was decided there was already sufficient justification to develop new equations in the PREDICT population. Nevertheless, following the publication of new American CVD risk prediction equations in 2013 (i.e the Pooled Cohort Equations or PCEs), it was decided that it would also be appropriate to assess their performance in the PREDICT population, albeit after the development of the new equations.

The first step was to apply the Pooled Cohort Equations (12) to the PREDICT-1^{o} population used to derive the new PREDICT-1^{o} equations described above, and then assess the performance of the risk scores using the standard performance metrics described above and in Chapter 3. The outcome used in these analyses was the atherosclerotic cardiovascular disease (ASCVD) outcome as defined by the PCEs developers.

The next step was to recalibrate the PCE models by updating the baseline survival functions estimated through fitting Cox models with the prognostic index from the PCE models, as an offset term, in the PREDICT-1^{o} dataset (53). Then, calibration was reassessed.

As a final step, the additional variables available in the PREDICT cohort but not included in the PCE models (i.e. ethnicity, deprivation, atrial fibrillation, family history of premature CVD, and treatment with lipid lowering and antithrombotic medications) were also assessed to determine if they were independent predictors over and above the predictors included in the PCEs. This was done by fitting Cox regression models that included the PCEs prognostic index (offset term) and the additional predictors, and then inspecting the effect estimates (Hazard Ratios and 95% confidence intervals) from these models.

In a similar manner, to externally validate the NZDCS models in the PREDICT-1^{o} T2D population, absolute risks were calculated using the NZDCS equation (28) and the obtained risk scores were assessed using standard performance metrics. The NZDCS models were then recalibrated by updating the baseline hazard estimated through fitting a Cox model with the prognostic index from the NZDCS model as an offset term in PREDICT dataset (53) and calibration was re-assessed. The additional predictors available in the PREDICT dataset but not included in the NZDCS model (i.e. BMI, eGFR, SES, AF, family history of CVD, and treatments with oral hypoglycaemic agents, insulin, lipid lowering and antithrombotic drugs) were investigated by appending them to the NZDCS model in the same way as described above.

In contrast to the validation of the PCEs after the development of the new PREDICT-1^{o} equations, the validation of NZDCS models was done as a first step, before developing the PREDICT-1^{o} T2D models. This was because the NZDCS models were relatively recent and developed in the NZ patient population, hence it was important to investigate whether the new PREDICT-1^{o} T2D equations were likely to be needed.

Statistical packages used

All analyses were performed using Stata13 software (122).

Chapter 6. Development and Internal Validation of the General Population PREDICT-1º CVD Risk Prediction Models in People without Prior CVD or Equivalent Risk

Introduction

In this chapter, the development and internal validation of the general population CVD risk prediction models in people without prior CVD or equivalent risk, known as the PREDICT-1^{o} CVD risk prediction models, are presented. The CVD prediction models specific to patients with T2D are described in Chapter 7.

Many models for CVD risk prediction are available, with the Framingham Heart Study equations being the most well-known and utilised risk estimation tools. A number of CVD risk prediction models have been derived from the Framingham Study cohort (as described in Chapter 2), and one of these, published in 1991, was recommended for use in New Zealand guidelines on the management of raised blood pressure and of dyslipidaemia in the early 1990s (10, 123). In 2002-2003 the New Zealand Guideline Group combined all clinical guidelines on the management of various CVD risk factors into one national CVD risk assessment and management guideline (10). For these new guidelines, the development team modified the previously recommended 1991 Framingham equation and established the 'NZ-modified Framingham Heart Study absolute CVD risk prediction equation' (9) as the recommended method for predicting risk. Concerns remained that a risk model developed in a white US population in the 1970s and 1980s might not be accurate enough in a 21st century multi-ethnic New Zealand population or in other high risk groups such as people with diabetes. Therefore, in 2002, the PREDICT study (described in detail in Chapter 3) was initiated to assess the performance of the Framingham equations and, if required, to develop new CVD risk prediction models for the New Zealand primary care population

During the data collection phase of the PREDICT study, the American College of Cardiology and American Heart Association (ACC/AHA) (12) published new risk equations, the so-called 'pooled cohort equations' (PCEs) for atherosclerotic cardiovascular disease (ASCVD). The PCEs were introduced as a replacement for the older Framingham equations recommended in previous US guidelines (123). These 'new' equations were race- and sex- specific models developed using combined data from a set of cohorts funded by the National Heart, Lung and Blood Institute, which included the original Framingham (124) and Framingham offspring (125) cohorts, as well as the ARIC (Atherosclerosis Risk in Communities) study (126), Cardiovascular Health Study (127), and CARDIA (Coronary Artery Risk Development in Young Adults) study (128). Combining the datasets was done to increase the diversity and therefore generalisability of the models.

This chapter presents new CVD risk prediction models for men and women without prior CVD or equivalent risk, derived from PREDICT cohort, and compares their performance with the more contemporary US pooled cohort equations, rather than the Framingham equations, as originally intended.

Aims and objectives

The main aim of this chapter is to describe the development of new general population CVD risk prediction models derived from the PREDICT study participants without prior CVD or equivalent risk and, secondarily, to compare their performance with the US pooled cohort equations (PCEs).

To achieve this aim, the objectives of this chapter were to:

- Develop new sex-specific models (PREDICT-1^{o}) in the PREDICT cohort study.
- Evaluate the performance of the new PREDICT-1^{o} equations.
- Evaluate the performance of the PCEs in the PREDICT-1^{o} cohort.
- If necessary, recalibrate the PCEs to the PREDICT-1^{o} population, and evaluate the re-calibrated equations.
- Compare the performance of new PREDICT-1^{o} equations with the performance of PCEs.

This chapter follows the Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) recommendations (21) as a guide. The sources of data, inclusion criteria, outcomes and predictors' definitions, treatment of missing data, and statistical analyses methods are specified in detail in Chapter 5.

Results

Participant numbers, follow-up time and events

The full PREDICT primary care cohort included 493,993 patients. After removing patients of Middle Eastern, African and Latin American ethnicities, as well as those with an unspecified

ethnicity, and applying the main exclusions (Figure 6.1), there were 401,752 participants (56.3% men) aged 30-74 who experienced 15,386 first CVD events during 1,685,521 person-years of follow-up (mean 4.2 years). The risk profile data were collected from August 2002 to October 2015. Outcomes data were collected up to 31 December 2015.

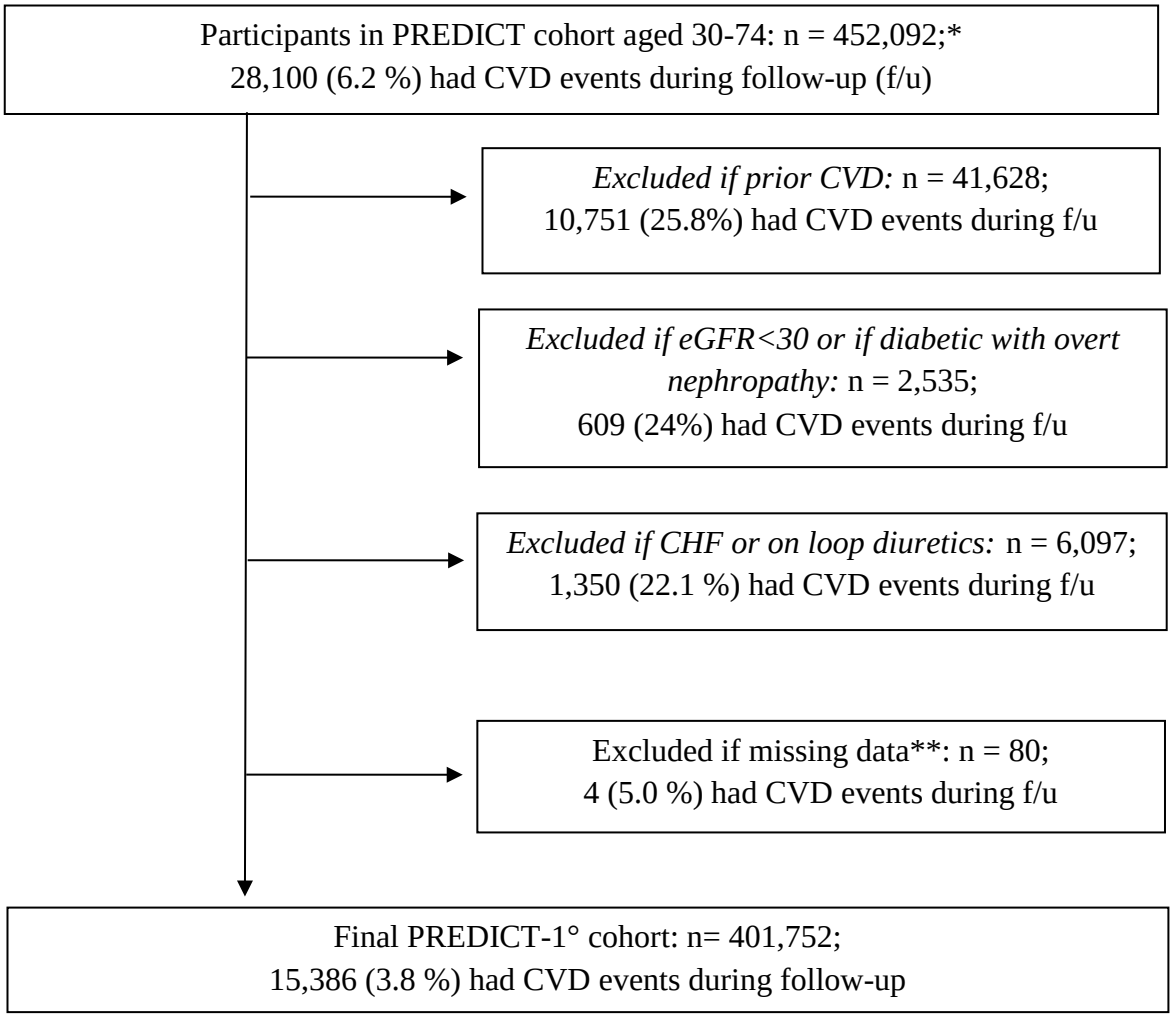


Figure 6.1. Flowchart of exclusions, PREDICT-1º cohort aged 30-74 years

*Excludes 448 people with inconsistent demographic variables across data sources and 7,822 people in ethnic groups with fewer than 1000 participants.

**The number of patients with missing data on each predictor: TC:HDL - 45 ; NZDep – 35.

Participant characteristics

The baseline description of the study cohort is presented in Table 6.1.

Table 6.1. Baseline characteristics of the PREDICT-1° cohort, aged 30-74 years

	Women	**Men**
Participants; n (% of total cohort)	175,699 (43.7%)	226,053 (56.3%)
Incident total CVD events; n (% sex-specific cohort)[a]	5,650 (3.2%)	9,736 (4.3%)
Total person-years observed	743,640	941,881
Crude incidence of total CVD (per 1000 per year)	7.6 (7.4, 7.8)	10.3 (10.1, 10.5)
Follow-up time in years; mean (SD)	4.2 (2.7)[b]	4.2 (2.7)[b]
People with follow up ≥5 years	58,493 (33.3)	72,417 (32.0)
Age in years; mean (SD)	56 (8.9)	51.8 (9.9)
Self-identified ethnicity:		
European	96,032 (54.7%)	128,503 (56.9%)
Māori	23,853 (13.6%)	27,573 (12.2%)
Pacific	22,537 (12.8%)	28,073 (12.4%)
South Asian	14,188 (8.1%)	20,232 (9.0%)
Chinese/other Asian	19,089 (10.9%)	21,672 (9.6%)
NZ Deprivation quintile:		
1 (least deprived)	38,523 (21.9%)	50,379 (22.3%)
2	34,230 (19.5%)	44,609 (19.7%)
3	31,808 (18.1%)	40,684 (18.0%)
4	32,626 (18.6%)	41,553 (18.4%)
5 (most deprived)	38,512 (21.9%)	48,828 (21.6%)
Smoking:		
Never smoker	129,158 (73.5%)	149,139 (66.0%)
Ex-smoker	24,838 (14.1%)	39,856 (17.6%)
Current smoker	21,703 (12.4%)	37,058 (16.4%)
Family history of premature CVD	22,996 (13.1%)	24,495 (10.8%)
Atrial fibrillation	1,777 (1.0%)	3,680 (1.6%)
Diabetes	27,377 (15.6%)	30,942 (13.7%)
Systolic blood pressure (SBP) mmHg; mean (SD)	129 (17.7)	129 (16.2)
Total Cholesterol to HDL Cholesterol ratio (TC:HDL), mean (SD)	3.7 (1.1)	4.4 (1.3)
Medications at index assessment:[c]		
On blood pressure lowering medications	45,973 (26.2%)	43,253 (19.1%)
On lipid lowering medications	27,540 (15.7%)	33,372 (14.8%)
On antithrombotic medications	17,831 (10.2%)	21,723 (9.6%)

[a] Values are n (% of sex-specific cohort) unless otherwise stated.

[b] Follow-up time ranged from one day to 13.3 years, in both men and women.

[c] 33.1% of women and 26.8% of men were treated with one or more class of drugs at index assessment.

The crude annual rate of fatal or non-fatal CVD events was 7.6 per 1000 person-years in women and 10.3 per 1000 person-years in men. Just over 30% were followed for 5 years or more. Women were, on average, older than men which reflects the national guidelines recommendation to risk assess women 10 years later than men. By ethnic group, 12.8% of participants self-identified as New Zealand Māori, 12.6% as Pacific, 10.1% as Chinese or other Asian, 8.6% as Indian, and 55.9% as European. In terms of clinical history, 15.6% of women and 13.7% men had diabetes; just over 10% reported a family history of premature CVD; and about 1.5% had a history of atrial fibrillation prior to index assessment.

About 30% of participants were on at least one of the lipid, blood pressure lowering or antithrombotic drugs. Blood pressure lowering treatment was the most common cardiovascular preventive therapy in the cohort with 26.2% women and 19.1% men recorded as receiving these medications at baseline. Approximately 15% of patients were on lipid lowering and 10% on antithrombotic drugs.

Outcome events

Myocardial infarction was the most common outcome (34%) with coronary heart disease accounting for over half of all events (Table 6.2). Stroke and transient ischemic attacks accounted for 26% of events, congestive heart failure for 12%, and peripheral vascular disease for 6%. Only 9.8% were fatal events. Just over half (55.5%) of the PREDICT-defined total CVD events met the PCEs definition of hard atherosclerotic CVD.

Table 6.2. Number and type of first CVD events in the PREDICT-1° cohort, aged 30-74 years

Outcome type	Non-Fatal events, n	Fatal events, n[a]	Proportion of all CVD events, %
Myocardial infarction	4,984	188	33.6
Unstable angina	2,275	11	14.9
Other coronary heart disease	343	436	5.1
Ischaemic stroke	2,124	156	14.8
Haemorrhagic stroke	445	205	4.2
Transient ischemic attack	1,123	0	7.3
Peripheral vascular disease	790	62	5.5
Congestive heart failure	1,795	113	12.4
Other Ischaemic CVD-related deaths	n/a	336	2.2
Total CVD events (N = 15,386)[b]	13,879	1,507	100

[a] If a patient died within 28 days of a non-fatal CVD event, the event was counted as fatal.

[b] If a patient had more than one type of CVD event, only the first was counted.

PREDICT-1° models development and specification

The final sex-specific models included 12 predictors: age at CVD risk assessment (continuous), ethnicity (European, Indian, NZ Māori, Pacific, Chinese/other Asian), smoking status (current, former, never), diabetes (yes/no), history of atrial fibrillation (yes/no), family history of premature CVD (yes/no), SBP (continuous, average of two measures), TC:HDL (continuous), socioeconomic deprivation quintile (ordinal, treated as continuous), and binary treatment indicators: blood pressure lowering, lipid lowering treatments, and treatment with antithrombotic medications. The hazard ratios for the final models are presented in Table 6.3.

Table 6.3. Adjusted[a] hazard ratios for first cardiovascular event, by sex

	Adjusted hazards ratios (95% CI)	
	Women	Men
Age (per year)	1.08 (1.07, 1.08)	1.07 (1.07, 1.07)
Ethnicity:		
European	1	1
Māori	1.48 (1.37, 1.60)	1.34 (1.26, 1.42)
Pacific	1.22 (1.12, 1.33)	1.19 (1.12, 1.27)
South Asian	1.13 (1.00, 1.27)	1.34 (1.24, 1.45)
Chinese/other Asian	0.75 (0.66, 0.85)	0.67 (0.61, 0.74)
NZ Deprivation quintile (per 1 quintile)	1.11 (1.09, 1.14)	1.08 (1.07, 1.10)
Smoking:		
Non-smoker	1	1
Ex-smoker	1.09 (1.01, 1.18)	1.08 (1.02, 1.14)
Smoker	1.86 (1.73, 2.00)	1.66 (1.57, 1.75)
Family history of premature CVD	1.05 (0.97, 1.12)	1.14 (1.08, 1.21)
Atrial fibrillation	2.44 (2.12, 2.81)	1.80 (1.62, 2.00)
Diabetes	1.72 (1.61, 1.85)	1.75 (1.66, 1.85)
SBP (per 10 mmHg)[b]	1.15 (1.12, 1.17)	1.18 (1.16, 1.20)
TC:HDL (per 1 unit)	1.13 (1.11, 1.15)	1.14 (1.12, 1.15)
Medications at index assessment:		
Receiving blood pressure lowering medications	1.40 (1.31, 1.50)	1.34 (1.27, 1.42)
Receiving lipid lowering medications	0.94 (0.88, 1.01)	0.95 (0.90, 1.00)
Receiving antithrombotic medications	1.12 (1.04, 1.21)	1.10 (1.03, 1.17)
Interactions:		
Age x Diabetes	0.978 (0.972, 0.984)	0.980 (0.977, 0.984)
Age x SBP (per 10 mmHg)	0.996 (0.994, 0.997)	0.996 (0.995, 0.997)
Receiving blood pressure lowering medications x SBP (per 10 mmHg)	0.958 (0.931, 0.985)	0.948 (0.926, 0.971)

[a] Hazard ratios are adjusted for all other variables included in the model.

[b] The hazard ratios for SBP are per 10 mmHg but were modelled per 1 mmHg for absolute risk calculations.

In these models, all continuous variables were fitted as linear terms after assessment using the fractional polynomials procedure and Martingale residuals plots provided no compelling support for fitting non-linear terms. Every additional year of age was associated with an increased estimated 5-year CVD risk of 7-8% in relative terms. Māori, Pacific and Indian peoples were all at increased (13 - 48% higher) estimated risk compared to Europeans, whereas Chinese and other Asian peoples were at between 25% - 33% lower risk than Europeans. Estimated risk increased by 11% in women and 8% in men per quintile of the social deprivation score and family history of premature CVD was a statistically significant predictor in men only. Smoking, diabetes, atrial fibrillation, increasing SBP and increasing TC:HDL were all statistically significant predictors, as were blood pressure-lowering and antithrombotic medications at the index assessment, but not lipid-lowering medications. Interactions between diabetes and age, between SBP and age, and between blood pressure lowering medication and SBP, were statistically significant in both sexes.

Variable coefficients, means of centred variables and baseline survival functions for the sex-specific 5-year CVD risk PREDICT-1° equations, along with an example calculation, are presented in Table 6.4.

Table 6.4. Beta Coefficients in the PREDICT-1° Equations for women & men, with example 5-year total CVD Risk calculation

Variable	Coefficient (Women)	Coefficient (Men)	EXAMPLE CALCULATION[a]		
			Woman	Coefficient x Variable[#]	Product
Age	0.0756412	0.0675532	55 years	0.0756412 x c.Age[#]	-0.08597757
			European		
Māori	0.3910183	0.2899054		0	
Pacific	0.2010224	0.1774195		0	
Indian	0.1183427	0.2902049		0	
Chinese/Asian	-0.28551	-0.3975687		0	
NZDep quintile	0.1080795	0.0794903	Quintle 3	0.1080795 x c.NZDep[#]	0.00099152
Ex-smoker	0.087476	0.0753246	Ex-smoker	0.087476 x 1 (Ex-smoker)	0.087476
Current smoker	0.6226384	0.5058041		0	
Family History CVD	0.0445534	0.1326587	No	0	
Atrial fibrillation	0.8927126	0.5880131	No	0	
Diabetes	0.5447632	0.5597023	Diabetes	0.5447632 x 1 (Diabetes)	0.5447632
SBP	0.0136606	0.0163778	135mmHg	0.0136606 x c.SBP[#]	0.08172727
TC:HDL	0.1226753	0.1283758	5 units	0.1226753 x c.TC:HDL[#]	0.15625546
OBPLM[b]	0.339925	0.2947634	Yes	0.339925 x 1 (OBPLM)	0.339925
OLLM	-0.0593798	-0.0537314	No	0	
OATM	0.1172496	0.0934141	No	0	
Age x Diab	-0.0222549	-0.020235		-0.0222549 x c.Age x 1(diab)	0.02529603
Age x SBP	-0.0004425	-0.0004184		-0.0004425 x c.Age x c.SBP	0.0030091
OBPLM x SBP	-0.004313	-0.0053077		-0.004313 x 1(OBPLM) x c.SBP	-0.02580339
			Sum Coefficients x Variables		**1.1276626**

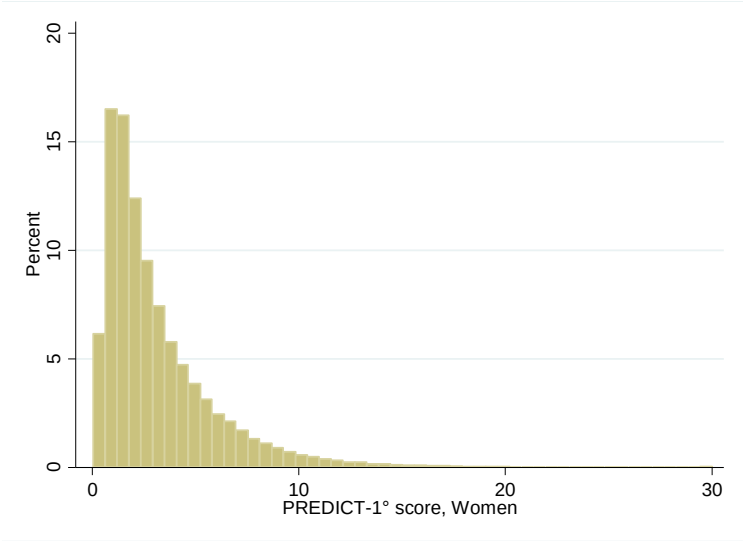
Means for centering			[#]Calculated centred variable for example	
Age	56.13665	51.79953	c.Age = 55 – 56.13665	- 1.13665
NZDep quintile	2.990826	2.972793	c.NZDep = 3 – 2.990826	0.009174
SBP	129.0173	129.1095	c.SBP = 135 – 129.0173	5.9827
TC:HDL	3.726268	4.38906	c.TC:HDL = 5 – 3.726268	1.273732
Baseline survival function (at 5 years)			Calculating 5-year CVD risk (1-Baseline surv $^{\exp(\text{sum of coefficients x variables})}$) x 100	
For women = 0.983169213058			**= (1- 0.983169213058 $^{\exp(1.1276626)}$) x 100 = 5.11%**	
For men = 0.974755526232				

[a] Example 5-year CVD risk calculation for a 55-year European women, in the middle social deprivation quintile (i.e. quintile 3), no family history of premature CVD, an ex-smoker, has diabetes, no atrial fibrillation, systolic blood pressure = 135mmHg, TC:HDL cholesterol = 5, on blood pressure lowering medications.

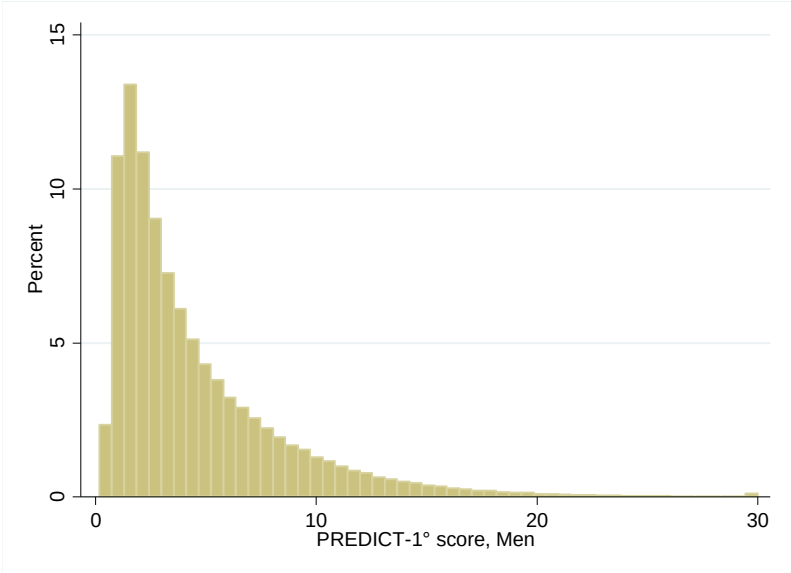
[b] Abbreviations: TC:HDL = Total Cholesterol to HDL Cholesterol ratio; CVD = cardiovascular disease; SBP=systolic blood pressure; NZDep=New Zealand deprivation score; On blood pressure lowering medications (OBPLM), On lipid lowering medications (OLLM), On antithrombotic medications (OATM).

[#] Denotes centred variables

The distributions of the PREDICT-1° cohort risk scores are shown in Figure 6.2, stratified by sex. Mean estimated 5-year CVD risk was 3.2% in women and 4.6% in men, and median risk was 2.3% (IQR: 1.3%, 4.2%) in women and 3.2% (IQR: 1.8%, 6.0%) in men.



6.2a. Women

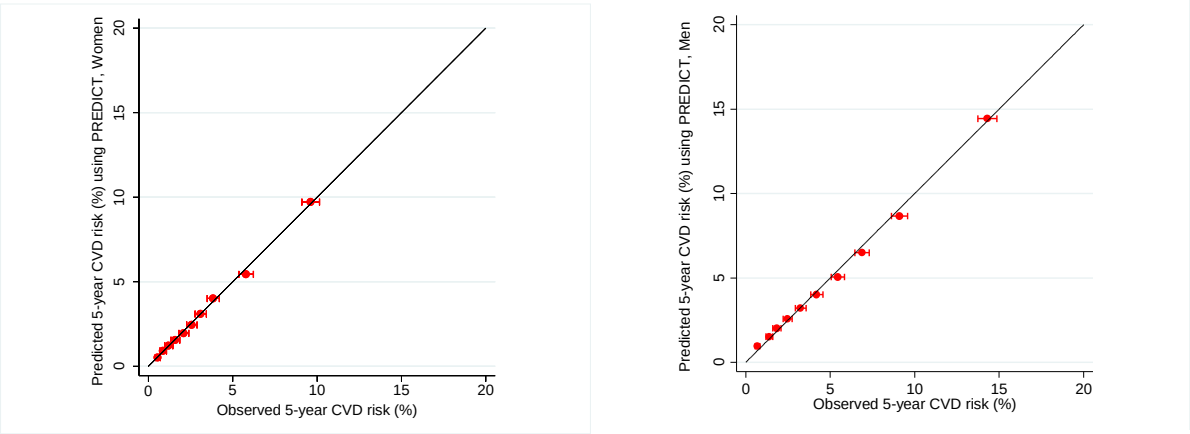


6.2b. Men

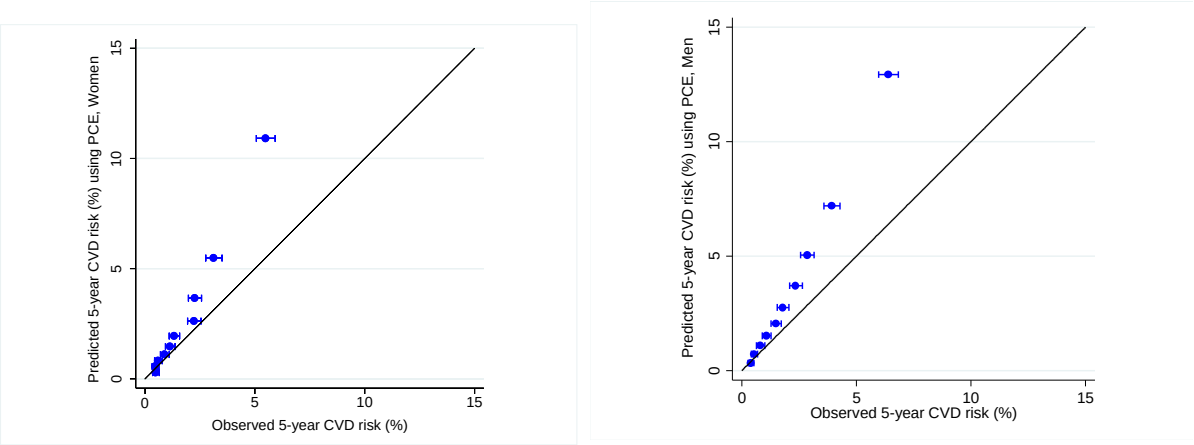
Figure 6.2. Distribution of PREDICT-1° risk prediction scores, by sex

Performance of PREDICT-1º sex-specific models

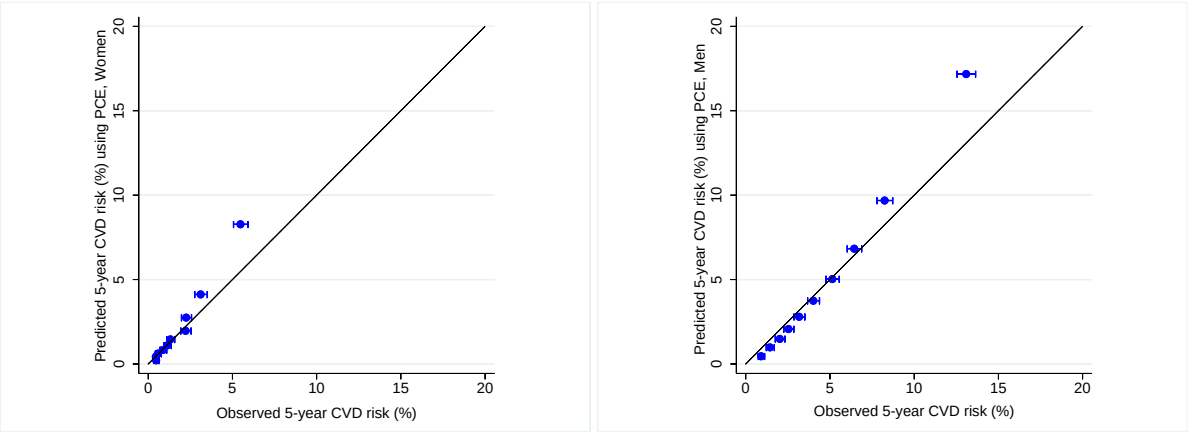
Predicted versus observed 5-year CVD risk plots for the PREDICT-1º equations (Figure 6.3a) demonstrated excellent calibration across all risk deciles in both sexes. Under- or over-prediction did not exceed 0.5% in any predicted risk decile.



6.3a. PREDICT-1° Equations (total CVD); women (left) & men (right)



6.3b. Original Pooled Cohort Equations (atherosclerotic CVD); women (left) & men (right).



6.3c. Recalibrated Pooled Cohort Equations (atherosclerotic CVD); women (left) & men (right).

Figure 6.3. Calibration plots: predicted versus observed 5-year CVD risk (%)*

*The diagonal line represents perfect calibration.

The models also showed good discrimination (Table 6.5). These findings are discussed further below, in comparison with the PCEs.

Table 6.5. Standard performance metrics for PREDICT-1° equations (estimating 5-year total CVD risk) & the PCEs (estimating 5-year atherosclerotic CVD risk), applied to the PREDICT-1° cohort, in women & men aged 30-74 years

	PREDICT-1° Equations	**Pooled Cohort Equations**
Women		
R^2	30 (29, 31)	26 (24, 28)
Harrell's C statistic	0.73 (0.72, 0.73)	0.71 (0.70, 0.72)
Royston's D statistic	1.334 (1.291, 1.377)	1.225 (1.162, 1.288)
Calibration slope	1.000 (0.968, 1.032)	0.753(0.714, 0.792)
Men		
R^2	29 (28, 30)	24 (23, 26)
Harrell's C statistic	0.73 (0.72, 0.73)	0.71 (0.70, 0.72)
Royston's D statistic	1.318 (1.285, 1.351)	1.157 (1.112, 1.202)
Calibration slope	1.000 (0.975, 1.025)	0.741 (0.710, 0.772)

95% confidence intervals were calculated for R^2 and Royston's D statistic using 1000 bootstrap replicates.

Validation and recalibration of the PCEs models in the PREDICT-1º cohort

The interquartile ranges and medians of the PCEs score calculated in the PREDICT-1º dataset are not presented, as the PCE ASCVD outcome only includes approximately 60% of the outcomes included in the new PREDICT-1° equations and so the comparison is not meaningful. However the calibration graphs, which for the PCEs were based on the ASCVD outcome, showed significant over-prediction of ASCVD risk in both sexes (Figure 6.3b). Recalibration of PCEs somewhat improved the fit at lower risk deciles, but the risk was still overestimated at the highest two risk deciles in men and women (Figure 6.3c).

Assessment of additional predictors which were not included in PCE models showed that ethnicity, high levels of socioeconomic deprivation, history of atrial fibrillation, and treatment with lipid lowering medications were statistically significant in both sexes, whereas family history of premature CVD and treatment with antithrombotic drugs was only statistically significant in men (Table 6.6).

Table 6.6. Hazard ratios for additional predictors combined with prognostic index from the Pooled Cohorts Equations[a]

	Adjusted hazard ratios (95% CI)	
	Women	**Men**
Ethnicity:		
European	1	1
Māori	1.64 (1.47, 1.83)	1.39 (1.26, 1.51)
Pacific	1.44 (1.28, 1.62)	1.37 (1.25, 1.50)
South Asian	1.30 (1.11, 1.53)	1.65 (1.49, 1.83)
Chinese/other Asian	0.82 (0.68, 0.98)	0.76 (0.67, 0.86)
NZ Deprivation quintile		
1	1	1
2	1.11 (0.96, 1.28)	1.05 (0.95, 1.16)
3	1.12 (0.97, 1.29)	1.12 (1.02, 1.24)
4	1.12 ((0.97, 1.29)	1.19 (1.08, 1.32)
5	1.43 (1.25, 1.64)	1.27 (1.15, 1.40)
Family history of premature CVD	1.08 (0.97, 1.21)	1.24 (1.14, 1.35)
Atrial fibrillation	2.20 (1.78, 2.71)	1.63 (1.40, 1.90)
Medications at index assessment:		
Receiving lipid lowering medications	0.86 (0.78, 0.95)	0.82 (0.76, 0.88)
Receiving antithrombotic medications	0.59 (0.86, 1.06)	0.89 (0.82, 0.96)

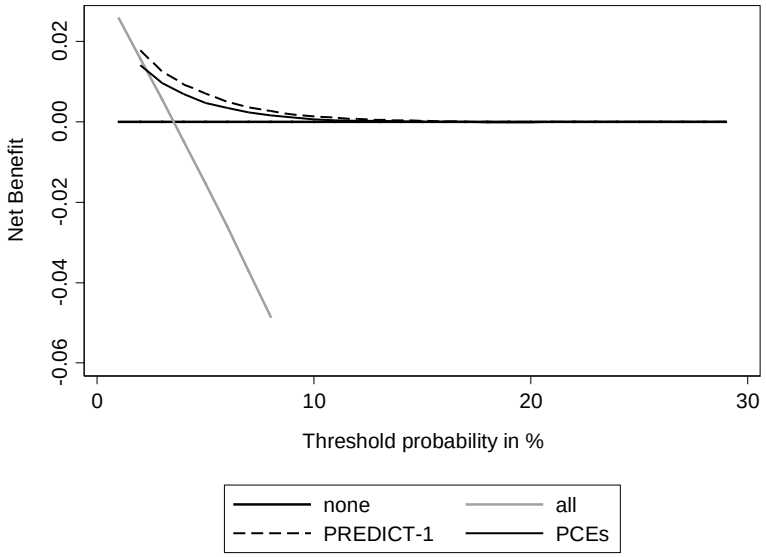
[a] PCE prognostic index included in Cox regression models as an offset term.

Comparison of performance of PREDICT-1° models and PCEs

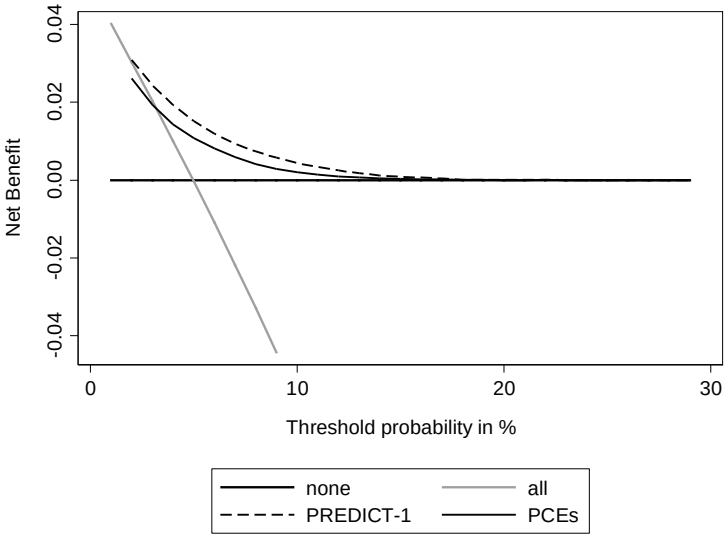
The performance statistics for PREDICT-1° models and PCEs are presented in Table 6.5. The discrimination of PREDICT-1° models was superior in comparison with the PCEs. In women, the Harrell's C-statistic was 0.73 and 0.71, and the D statistic was 1.334 and 1.225, respectively. In men, the C-statistic was 0.73 and 0.71; the D-statistic was 1.318 and 1.157, respectively. The proportion of explained variation was 4% higher in women and 5% higher in men for PREDICT-1° models compared with PCEs. The calibration slopes of the PCEs were 0.75 and 0.74 for women and men, respectively, while for PREDICT-1° models, the calibration slope was 1. The calibration plots for PREDICT-1° models showed that estimated event rates were close to observed, in both men and women, while PCEs overestimated risk in both sexes, even after recalibration (Figure 6.3).

The net benefit (the trade off between patients correctly treated and those treated unnecessarily at various definitions of high risk) of using PREDICT-1° was greater compared with the PCEs across clinically meaningful ranges of 5-year risk where preventive therapy might be considered (Figure 6.4). In women, below a 2% and above a 12% treatment threshold, the

PREDICT-1° model provided no additional benefit beyond a treat-none or treat-all approach, respectively. In men, the PREDICT-1° model was superior in the range between 2% and 17%.



6.4a. Women



6.4b. Men

Figure 6.4. Net benefit curves for PREDICT-1º and PCEs models

Note: the 'treat none' line is the horizontal line through zero net benefit. PREDICT-1º and PCE models are the curved lines.

PREDICT-1º Derivation and Validation cohort: sensitivity analyses

There were 235,141 people in the Derivation cohort (57% men) and 166,611 in the Validation cohort (56% men). Approximately 50% of the Derivation cohort were European compared to about 65% of the Validation cohort. The Derivation cohort also had fewer Māori participants, but the proportions of Pacific, Indian, and Chinese participants were higher than in the Validation cohort (Table 6.7).

Table 6.7. Description of the Derivation and Validation cohorts, aged 30-74 years

	Women (Derivation)	**Women (Validation)**	**Men (Derivation)**	**Men (Validation)**
Participants; n (% of total cohort)	102,242	73,457	132,899	93,154
Incident CVD events; n (% of sex-specific cohort)[a]	3,108	2,542	5,422	4,314
Total person-years observed	424,288	319,352	549,774	392,107
Crude incidence of CVD (per 1000 per year)	7.3 (7.07, 7.6)	8.0 (7.7, 8.3)	9.9 (9.6, 10.1)	11.0 (10.7, 11.3)
Follow-up time in years, mean (SD)[b]	4.1 (2.7)	4.3 (2.7)	4.1 (2.7)	4.2 (2.7)
People with follow up >=5 years	32,094	26,399	40,946	31,471
Age in years; mean (SD)	56 (9.0)	57 (8.7)	51 (10.0)	53 (9.7)
Ethnicity:				
European	48,857 (47.8)	47,175 (64.2)	66,338 (49.9)	62,165 (66.7)
Māori	11,445 (11.2)	12,408 (16.9)	13,592 (10.2)	13,981 (15.0)
Pacific	17,767 (17.4)	4,770 (6.5)	22,296 (16.8)	5,777 (6.2)
South Asian	11,428 (11.2)	2,760 (3.8)	16,197 (12.2)	4,035 (4.3)
Chinese/other Asian	12,745 (12.5)	6,344 (8.6)	14,476 (10.9)	7,196 (7.7)
NZ Deprivation quintile:				
1 (least deprived)	23,871 (23.4)	14,652 (20.0)	30,547 (23.0)	19,832 (21.3)
2	19,132 (18.7)	15,098 (20.6)	24,995 (18.8)	19,614 (21.1)
3	15,998 (15.7)	15,810 (21.5)	20,982 (15.8)	19,702 (21.2)
4	17,435 (17.1)	15,191 (20.7)	22,949 (17.3)	18,604 (20.0)
5 (most deprived)	25,806 (25.2)	12,706 (17.3)	33,426 (25.2)	15,402 (16.5)
Smoking:				
Never smoker	78,489 (76.8)	50,669 (69.0)	90,758 (68.3)	58,381 (62.7)
Ex-smoker	11,957 (11.7)	12,881 (17.5)	20,383 (15.3)	19,473 (20.9)
Current smoker	11,796 (11.5)	9,907 (13.5)	21,758 (16.4)	15,300 (16.4)
Family history of premature CVD	11,455 (11.2)	11,541 (15.7)	12,826 (9.7)	11,669 (12.5)
Atrial fibrillation	927 (0.9)	850 (1.16)	1,916 (1.44)	1,764 (1.9)
Diabetes	17,970 (17.6)	9,407 (12.8)	19,929 (15.0)	11,013 (11.8)
SBP mmHg; mean (SD)	128 (17.6)	130 (17.8)	128 (16.2)	130 (16.2)
Mean TC:HDL; mean (SD)	3.7 (1.1)	3.7 (1.1)	4.4 (1.2)	4.4 (1.3)
Medications at index assessment:				
On blood pressure lowering medications	26,372 (25.8)	19,601 (26.7)	24,926 (18.8)	18,327 (19.7)
On lipid lowering medications	16,928 (16.6)	10,612 (14.5)	20,544 (15.5)	12,828 (13.8)
On antithrombotic medications	10,699 (10.5)	7,132 (9.7)	12,987 (9.8)	8,736 (9.4)

[a] Values are n (% of sex-specific cohort) unless otherwise stated.

[b] The follow-up time ranged from one day to 13.3 years, in both men and women.

About a quarter of the Derivation cohort were in the highest deprivation quintile, compared with only 17% of the Validation cohort. Proportions of participants with diabetes were slightly higher in the Derivation than in the Validation cohort; 18% and 13% respectively in women, 15% and 12% respectively in men. Mean SBP was 2mmHg lower in the Derivation cohort in both sexes, but mean TC:HDL values were identical and drug treatment was similar in the two cohorts.

The hazard ratios were similar in Cox regression models fitted in the whole and Derivation cohorts (Table 6.8).

Table 6.8. Adjusted hazard ratios in risk models derived from the Derivation cohort and the full PREDICT-1° cohort, by sex

	Women (Derivation cohort)	Women (full PREDICT-1° cohort)	Men (Derivation cohort)	Men (full PREDICT-1° cohort)
Numbers of participants / first CVD events	102242/3108	175699 / 5650	132899/5422	226053 / 9736
Predictors				
Age (per year)	1.08(1.07, 1.09)	1.08 (1.07, 1.08)	1.07 (1.07, 1.07)	1.07 (1.07, 1.07)
Ethnicity:				
European	1	1	1	1
NZ Māori	1.47 (1.31, 1.64)	1.48 (1.37, 1.60)	1.33 (1.22, 1.46)	1.34 (1.26, 1.42)
Pacific	1.15 (1.03, 1.28)	1.22 (1.12, 1.33)	1.15 (1.06, 1.25)	1.19 (1.12, 1.27)
South Asian	1.10 (0.96, 1.27)	1.13 (1.00, 1.27)	1.33 (1.22, 1.46)	1.34 (1.24, 1.45)
Chinese/other Asian	0.71 (0.60, 0.83)	0.75 (0.66, 0.85)	0.71 (0.63, 0.80)	0.67 (0.61, 0.74)
NZ Deprivation quintile (per 1 quintile)	1.15 (1.11, 1.18)	1.11 (1.09, 1.14)	1.10 (1.07, 1.12)	1.08 (1.07, 1.10)
Smoking:				
Non-smoker	1	1	1	1
Ex-smoker	1.13 (1.00, 1.27)	1.09 (1.01, 1.18)	1.07 (0.99, 1.16)	1.08 (1.02, 1.14)
Smoker	1.93 (1.75, 2.13)	1.86 (1.73, 2.00)	1.64 (1.53, 1.75)	1.66 (1.57, 1.75)
Family history premature CVD	1.03 (0.92, 1.14)	1.05 (0.97, 1.12)	1.11(1.02, 1.20)	1.14 (1.08, 1.21)
Atrial fibrillation	2.28(1.86, 2.79)	2.44 (2.12, 2.81)	2.02 (1.75, 2.32)	1.80 (1.62, 2.00)
Diabetes	1.64 (1.50, 1.79)	1.72 (1.61, 1.85)	1.79 (1.67, 1.91)	1.75 (1.66, 1.85)
Systolic blood pressure (per 10 mmHg)	1.14 (1.11, 1.17)	1.15 (1.12, 1.17)	1.19 (1.16, 1.21)	1.18 (1.16, 1.20)
TC:HDL (per 1 unit)	1.15 (1.12, 1.19)	1.13 (1.11, 1.15)	1.14 (1.12, 1.16)	1.14 (1.12, 1.15)
Medications at index assessment				
Not on blood pressure lowering, anti-clotting or lipid lowering medications	1	1	1	1
On blood pressure lowering medications	1.47 (1.35, 1.61)	1.40 (1.31, 1.50)	1.37 (1.27, 1.47)	1.34 (1.27, 1.42)
On antithrombotic medications	1.18(1.07, 1.30)	1.12 (1.04, 1.21)	1.07 (1.00, 1.16)	1.10 (1.03, 1.17)
On lipid lowering medications	0.91 (0.83, 1.00)	0.94 (0.88, 1.01)	0.96 (0.89, 1.03)	0.95 (0.90, 1.00)
Interactions				
Age x diabetes	0.977 (0.968, 0.985)	0.978 (0.972, 0.984)	0.980 (0.974, 0.985)	0.980 (0.976, 0.984)
Age x SBP (per 10 mmHg)	0.996 (0.993, 0.998)	0.996 (0.994, 0.997)	0.996 (0.995, 0.998)	0.996 (0.995, 0.997)
OBPLM x SBP (per 10 mmHg)	0.963 (0.927, 0.999)	0.958(0.931, 0.985)	0.935 (0.906, 0.964)	0.948 (0.926, 0.971)

The calibration performance of the risk score developed in the Derivation cohort and applied to the Validation cohort was also similar to the performance observed in the whole cohort analyses, except for the highest decile of predicted risk where the observed risk was lower than predicted by 2-4% (Figure 6.5).

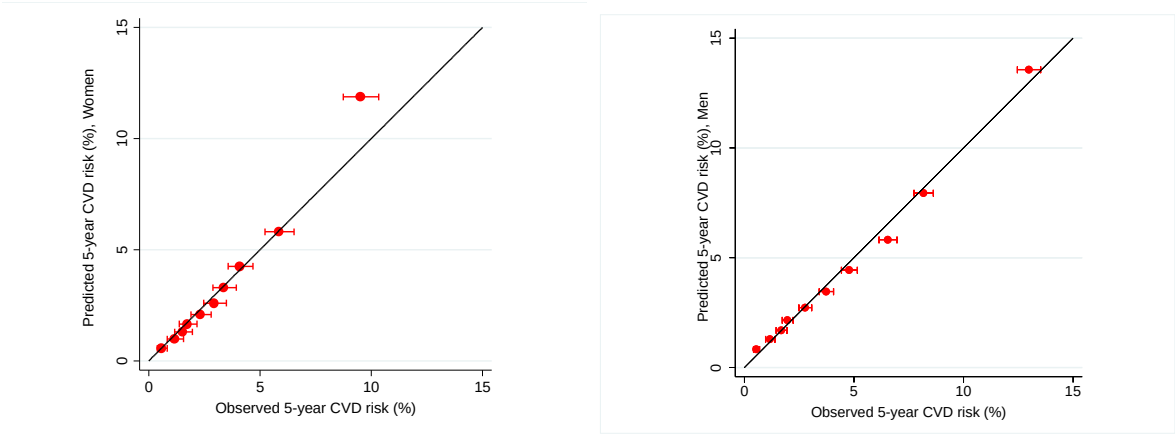


Figure 6.5. Calibration plots: predicted 5-year CVD risk (%) compared to observed 5-year CVD risk (%) using the Derivation cohort models and applied to the Validation cohort, in women (left) and men (right) aged 30-74 years

Performance indicators were comparable and consistent with those in the main analyses (Table 6.9).

Table 6.9. Performance statistics for the Derivation cohort models applied in the Validation cohort, and for the full PREDICT-1° models applied in the full cohort, by sex

	Derivation cohort model	Full PREDICT-1° cohort
Women		
R^2	27 (25, 29)	30 (29, 31)
Harrell's C statistic	0.72 (0.71, 0.73)	0.73 (0.72, 0.73)
Royston's D statistic	1.26 (1.19, 1.32)	1.334 (1.291, 1.377)
Men		
R^2	28 (26, 29)	29 (28, 30)
Harrell's C statistic	0.72 (0.71, 0.72)	0.73 (0.72, 0.73)
Royston's D statistic	1.26 (1.21, 1.31)	1.318 (1.285, 1.351)

95% confidence intervals for R^2 and Royston's D statistic were calculated using 5000 bootstrap replicates.

Discussion

Summary of findings

The PREDICT cohort provided a large, contemporary, representative study population of primary care patients with almost no missing data, and with about one third of the cohort followed up for five years or longer. Using PREDICT data, new CVD risk prediction models for primary care patients without prior CVD or equivalent high CVD risk were developed. The new sex-specific models include both the conventional predictors (age, smoking, diabetes, systolic blood pressure, TC:HDL) and new predictors (ethnicity, socioeconomic status, family history of premature CVD, CVD preventive medication use and atrial fibrillation). PREDICT-1^{o} models' performance was superior, across all pre-determined performance indicators, to the current Northern American ACC/AHA Pooled Cohorts Equations. The 2013 Pooled Cohorts Equations, in their original form, overestimated ASCVD risk in almost all deciles of risk. After recalibration, the sex-specific models still overestimated risk in the top two deciles, in both men and women. The calibration slope in this external evaluation of the PCEs in the PREDICT-1^{o} cohort, was less than 1.0, suggesting either overfitting of the development data or true differences in effects of predictors between the Pooled Cohorts and PREDICT-1^{o} study population; with a combination of these two issues being most likely.

The decision curve analysis graphs, which display the net benefit of using a model at different thresholds of risk, showed that PREDICT-1° models identified more cases without increasing the number treated unnecessarily when compared with the PCEs, across clinically relevant thresholds of risk.

Also, this study demonstrated that new predictors, such as SES, self-reported ethnicity, CVD-preventive treatment status and history of atrial fibrillation have additional predictive value over and above the predictors included in PCEs.

Interpretation of findings

This study showed that both self-reported ethnicity and an area-based indicator of socioeconomic status were independent predictors of CVD risk. Ethnic differences in patterns of disease may be attributed to several mechanisms beyond any differences in genetic profiles. Ethnicity is a complex social construct with unique societal and individual meanings and

therefore difficult to measure accurately. Yet some variations in health outcomes, some differences in health behaviours, and some responses to health interventions, cannot be quantified except by using the concepts of ethnicity or race. The social experiences of identifying with a particular ethnic group can impact on health through cultural variations in presentation of symptoms and perceptions of health and disease; through language competence, education and availability of information; through accessibility of services; through attitude, awareness and skill of clinical staff; or through ethnicity-related differences in clinician-patient relationships. Socioeconomic status (SES) profiles also differ across ethnic groups, and ethnicity is partly a representation of various socioeconomic inequalities, such as social class structures or economic disparities. SES is also difficult to assess accurately and there are many ways of measuring socioeconomic inequalities. For example, the measure used in this study is area-based and therefore may not correctly represent the SES of some individuals in the study population. Nevertheless, even though ethnicity and SES may overlap in undefined ways, they are still capturing non-overlapping aspects of prognostic information and therefore complement each other.

As in any modern cohort, many of the patients in this cohort were on blood pressure lowering medications at index assessment, and were observed to have an increased relative risk of CVD. This is contrary to the expected direction of the effect given results from clinical trials. However, this result is consistent with other reports from observational studies and has been termed “confounding by indication”: a distortion that arises from imbalance of prognostic factors between treated and untreated patients (129). In non-randomised studies, confounding by indication can complicate investigations, since in order to fully control for confounding by indication, one requires full information on prognostic factors, including a range of signs a clinician reads but does not fully record that also influences treatment choice. It is also possible that the increased risk among participants on blood pressure lowering medications in part reflects poorly controlled blood pressure that was not adequately accounted for by the blood pressure measurements included in the models. Blood pressure is notoriously difficult to measure accurately, and PREDICT researchers have previously demonstrated significant digit preference with blood pressure measurements in the PREDICT cohort (130). An increased risk was also associated with anthrombotic medications which could also be due to confounding by indication. In contrast, there was no increase in risk associated with lipid lowering medications. Possible explanation for why this differs from the findings for blood pressure lowering medications are firstly that lipid measurements are subject to much less measurement error than

blood pressure, so the prognostic effect of lipid levels is more accurately included in the models than the effect of blood pressure. Secondly, reaching target levels for lipids with statin therapy is relatively easier than reaching target blood pressure levels (often requires titration with multiple drugs).

The PREDICT-1^{o} models were derived and assessed using the full PREDICT cohort rather than the more common approach of randomly splitting the cohort and developing models in a derivation subcohort and assessing their performance in a validation subcohort (32). As discussed in Chapter 4, split-sample methods are problematic and generally an inefficient approach to validation, particularly in large studies (131-133). Simulation studies have shown that while large samples are required to make split-sample validation meaningful (134), at the same time, with a large sample, validation assessment based on the full cohort is already a good indicator of model performance. Also, it can be a waste of valuable information if the final model are only based on a sample of a large cohort. Differences in regression coefficients would generally be small, if the split is random, but the estimates of the full sample would be more stable (38).

Nevertheless, for completeness, a split-sample validation was done as a sensitivity analysis, but using a geographic rather than a random split. A spatial (i.e. non-random) split is more likely to result in a validation cohort that differs systematically from the derivation cohort thus reducing the chance of a 'home advantage' by overestimating the performance of prognostic models. Therefore, the demographic and clinical differences observed between the derivation and validation cohorts was expected and made the validation more meaningful. That the derivation cohort had a higher proportion of patients with diabetes and higher levels of deprivation, as well as higher proportions of patients of Pacific and Indian origin, might explain the small overestimation of risk in the highest decile observed in the calibration graphs derived from the validation cohort.

As discussed, the 2013 ACC/AHA PCEs were externally validated in the PREDICT-1^{o} cohort and compared with the PREDICT-1^{o} equations. The original plan was to externally validate the NZ-modified Framingham equation (9) used in PREDICT software (and in other equivalent software in New Zealand), however US guidelines now recommend replacing Framingham equations with the PCEs (12) and so they were considered more relevant comparison. The definition of the CVD outcome was more restrictive in the PCEs than PREDICT-1^{o}, which would have influenced estimated predictor effects, and consequently lead to different estimated

probabilities and model performance (135). Therefore, during external validation of the PCEs in the PREDICT cohort, the PCEs' ASCVD outcome was used, instead of the more global PREDICT CVD outcome. Of note, the more restrictive PCE definition of outcome would have limited the usefulness of the PCEs in the New Zealand context, regardless of their performance. This is because New Zealand guidelines recommend predicting the risk of a more comprehensive range of CVD events that are relevant in practice.

The overestimation of risk observed in the calibration of the PCEs was not unexpected, as many of the cohorts included in the dataset used to derive the PCEs are now dated. CVD mortality and incidence has declined by over 75% in New Zealand and in the US over the past five decades, and so the PCEs are largely based on a study population at considerably higher risk of CVD than the more contemporary New Zealand population. An external validation and comparison with the UK QRISK equations was also considered because it is based on a more contemporary population than the PCEs. However, it was difficult to obtain access to the full details of the equations and they included many variables that were not available in the PREDICT dataset.

Findings in the context of existing research

The hazard ratios derived from the new PREDICT-1^{o} equations are consistent with findings elsewhere. In regard to ethnicity, Indians and other South Asian populations have been shown to be at increased risk over and above the standard risk factors in the large QRISK primary care cohort in the UK (31), although the equations presented here are the first to identify an increased risk in Māori and Pacific peoples. The unique ethnic composition of the New Zealand population was an important consideration in the decision to derive equations from a New Zealand primary care cohort (136).

A family history of premature CVD was only a statistically significant predictor in men in the PREDICT-1^{o} equations, and the effect was modest. The Framingham investigators found a two-fold increase in CVD risk associated with premature parental CVD when their analyses were based on validated data through linkage of the Framingham Offspring cohort participants to their parent's data in the Original Framingham cohort (137). However, their analyses based only on self-reported data showed no increased risk. The observed lack of a strong association between self-reported premature family history of CVD in PREDICT and in Framingham

studies is most likely attributed to poor participant recall, as has been well documented previously (138).

As discussed above, the observed lack of association between lipid lowering treatment at index assessment, in contrast to increased risk as in case of systolic blood pressure, has been reported previously (12).

In line with the PCEs external validation reported in this chapter, previous external validation studies of the PCEs have found that on average they overestimate observed risk of ASCVD by 60% to 90% in more contemporary USA cohorts. They have also been shown to overestimate risk in several European and Asian cohorts (139).

In contrast to many other studies, data-driven approaches were not used to inform variable selection for the PREDICT-1^{o} models and predictors were selected *a priori*, based on previous studies and clinical judgement. This reduces the chance of accidental findings that are not generalisable to other populations. Stepwise statistical approaches to variable selection may result in the inclusion of unimportant variables and exclusion of important ones (38).

Strengths and limitations

The main strengths of this study are its large, representative, and contemporary study population, with substantial numbers of CVD events observed during follow up. There were a total of 15,386 CVD events which is equivalent to over 800 events per predictor variable - well above the minimum 10 events per predictor recommended when developing prediction models (140). The completeness of data on the included prognostic factors and on historical and follow up data, for a study of this size, is also a major strength. Most of the information on the predictor variables was recorded when general practitioners estimated their patients CVD risk and these calculations required complete data on the predictors. All event data came from direct linkage to national hospitalisation and mortality datasets which are considered to be complete due to the availability of a universal unique personal identifier throughout all the healthcare services. Hence, loss to follow up is likely to be negligible, with participants leaving the country being the only people whose incident CVD might be unavailable. Besides its representativeness, this study included large numbers of Māori, Pacific, Indian and Chinese participants. Previous studies of these ethnic-specific populations in New Zealand have been too small to enable reliable estimates of ethnicity effects in prediction models. The inclusion of ethnicity, SES, history of atrial fibrillation, family history of premature CVD and preventive treatment status,

in addition to the standard CVD risk predictors, are also strengths of this study. Most of these additional variables were shown to be important predictors which improve the new equations ability to discriminate better between high and low risk patients compared to previous equations.

One of the potential limitations of the study is that, like all modern patient cohorts, PREDICT included participants who were treated with CVD preventive medications, some of whom would have had this treatment initiated after the baseline assessment. New treatment, not accounted for at baseline, will lead to an underestimate of untreated risk, which is problematic because in clinical practice, a risk assessment is done to inform the question "what would happen to risk over the next five (or ten) years if there was no new treatment initiated?" However, preliminary analyses, reported in Appendix 9.1, suggested that new treatment initiated after baseline was limited and was therefore unlikely to be influential. This is in part due to the relatively short follow-up period (mean of approximately 4 years) and because many participants were already on treatment at baseline.

Another limitation is that information on other important CVD risk factors including levels of physical activity (141) and diet were not available in the study population. While these are potentially useful predictors, they are difficult to measure robustly in a busy general practice setting and are therefore not practical to include in risk equations.

Also, the PREDICT-1^{o} models were essentially internally validated (developed and assessed in the same dataset), whereas for the comparison PCEs models the validation was performed in a new, independently collected dataset. The performance of an existing model is often poorer than in the original development sample (21). Therefore, there is potentially an impact from comparing the results from the internal vs external validation.

Implications for research and practice

The PREDICT-1^{o} equations were developed for use in routine general practice. The purpose of this study was to investigate whether equations developed from contemporary primary care data would perform better than existing internationally recognised equations. The main findings of this study are that contemporary data are required to develop CVD risk prediction equations that are applicable to current populations, and equations that include additional sociodemographic and medical history information will improve prediction over and above the standard predictors included in most published equations. Rapidly declining incidence of CVD

and the changing distributions of risk factors in modern populations suggest that risk estimation tools will need to be regularly updated to reflect these changes. The increasing computerisation of health care records makes frequent updates and/or recalibration of risk estimation tools possible.

Future research should aim to externally validate the PREDICT-1^{o} models elsewhere. Also additional validation of these models in sub-populations of patients, such as ethnic groups, older people and patients with co-morbidities would be informative.

It is important to evaluate models' performance not only from a statistical perspective, but also from a clinical performance point of view. Net benefit analysis allows comparing clinical utility of models by examining the balance between true positives (patients correctly treated) and false positives (patients treated unnecessarily) across a range of treatment thresholds. Plotting net benefit using decision curves allows for an easy, intuitive interpretation, as higher net benefit over a clinically relevant range of decision thresholds implies that the model leads to better decision making. The relative clinical utility of models does depend on the definition of high risk, therefore it is important to assess the net benefit with consideration of thresholds currently recommended in clinical practice.

A focus on improving the quality of data available for developing and updating the CVD risk estimation models is also important. For example, the possible error in recording the self-reported family history of premature CVD in PREDICT study could potentially be addressed by modifying the PREDICT risk assessment template in consultations with primary care providers, or ideally linking to data from family members.

Further investigation of the effects of drug treatment status at risk assessment, especially why some preventive treatments at baseline are associated with the increased risk and others are not, would be informative. Also, the impact of preventive treatments initiated during follow up requires further research, particularly as the follow-up time of the PREDICT cohort increases. Modelling techniques that take into account treatment-covariate interactions and time-varying covariates might also be of use.

Summary

- Sex-specific models for predicting of CVD risk were developed using a contemporary representative population of primary care patients in New Zealand.

- The performance of these models was compared with existing models developed from a combination of Northern American cohorts and proved to be superior in all measures of statistical performance and clinical utility.

- This study supports the use of CVD risk prediction models derived from contemporary populations; it also supports addition of new risk predictors, both sociodemographic and clinical.

Chapter 7. Development and Internal Validation of Diabetes-Specific Models to Estimate CVD Risk

Introduction

Chapter 6 described the development and assessment of general population models that included diabetes as one of the risk predictors. This chapter will present the development and internal validation of models specifically designed for patients with type 2 diabetes (T2D).

T2D has been reported to double the risk of CVD over and above other known CVD risk factors (142, 143) and clinical guidelines in New Zealand and other countries recommend calculating CVD risk in these patients, to inform the intensity of preventive treatment to control blood pressure, blood lipids and levels of blood sugar (10, 144-146).

The majority of risk calculators have been developed over the last 20 years, and, although many include diabetes as a risk predictor, few have been designed specifically for patients with T2D (28, 33, 147-151). Comparing the performance of scores from different studies is difficult as outlined in Chapter 2 of this thesis, because of variation in exclusion criteria, duration of follow up, and definitions of endpoints and predictors. Comparisons are also difficult due to inherent differences in the distributions of risk factors in populations. Many people also receive medications to lower their risk factors (e.g. blood pressure, LDL cholesterol). A recent systematic literature review identified 45 prediction models which can be used for CVD prediction in T2D patients, of which only 12 models were designed specifically for patients with T2D (152). The most commonly used predictors in these models were age, sex, duration of diagnosed diabetes, HbA_{1c} and smoking, and they mostly predicted 5-year risk (152). Previous research has also shown that CVD risk scores developed in a general population are likely to underestimate CVD risk in individuals with diabetes (153). However, the evidence that the risk scores developed in patients with T2D estimate cardiovascular risk more accurately than those derived from general populations has been inconsistent (153).

In New Zealand, a set of T2D-specific models for predicting 5-year CVD risk were developed and published in 2010 (28). These models were derived from the New Zealand Diabetes Cohort Study which is a prospective open cohort using a routinely-collected data from national primary care program called “Get Checked”. This program was initiated in 2000 by the Ministry of Health, which funded primary health care to provide a free annual review for all patients with diabetes. The data collected between 2000 and 2006 were used to derive the models, which, in addition to predictors included in Framingham model (age, sex, smoking status, systolic blood

pressure and TC:HDL), incorporated new predictors such as ethnicity, duration of diabetes, micro- and macroalbuminuria, and haemoglobin A1c.

Aims and objectives

The aim of the research presented in this chapter was to establish whether a new risk score could be developed from the PREDICT-1^{o} T2D population that performs better than an existing, diabetes-specific risk score developed from the New Zealand Diabetes Cohort Study (NZDCS).

To achieve this aim, the objectives of this chapter were to:

- Evaluate the performance of NZDCS models in the PREDICT-1^{o} T2D population using standard performance metrics.
- Recalibrate the NZDCS model to the PREDICT-1^{o} T2D population, if necessary, and evaluate the re-calibrated model;
- Develop new sex-specific models using a subset of patients with T2D from PREDICT cohort;
- Evaluate the performance of new PREDICT-1^{o} T2D models and compare with the NZDCS model.

This chapter follows the Transparent Reporting of multivariable prediction models for Individual Prognosis Or Diagnosis (TRIPOD) recommendations as guidelines to report this multivariable prognostic modelling study (21). The sources of data, inclusion criteria, outcomes and predictors' definitions, treatment of missing data, and statistical analyses methods are specified in detail in Chapter 5.

Results

Participants

The full PREDICT cohort included 86,756 patients with T2D. After applying exclusions (Figure 7.1), there were 39,834 patients (51.5% men) aged 30-74 who experienced 3,295 first CVD events during 194,611 person-years of follow-up (mean 4.9 years). The risk profile data

were collected from August 2002 to October 2015. Outcomes data were collected up to 31 December 2015.

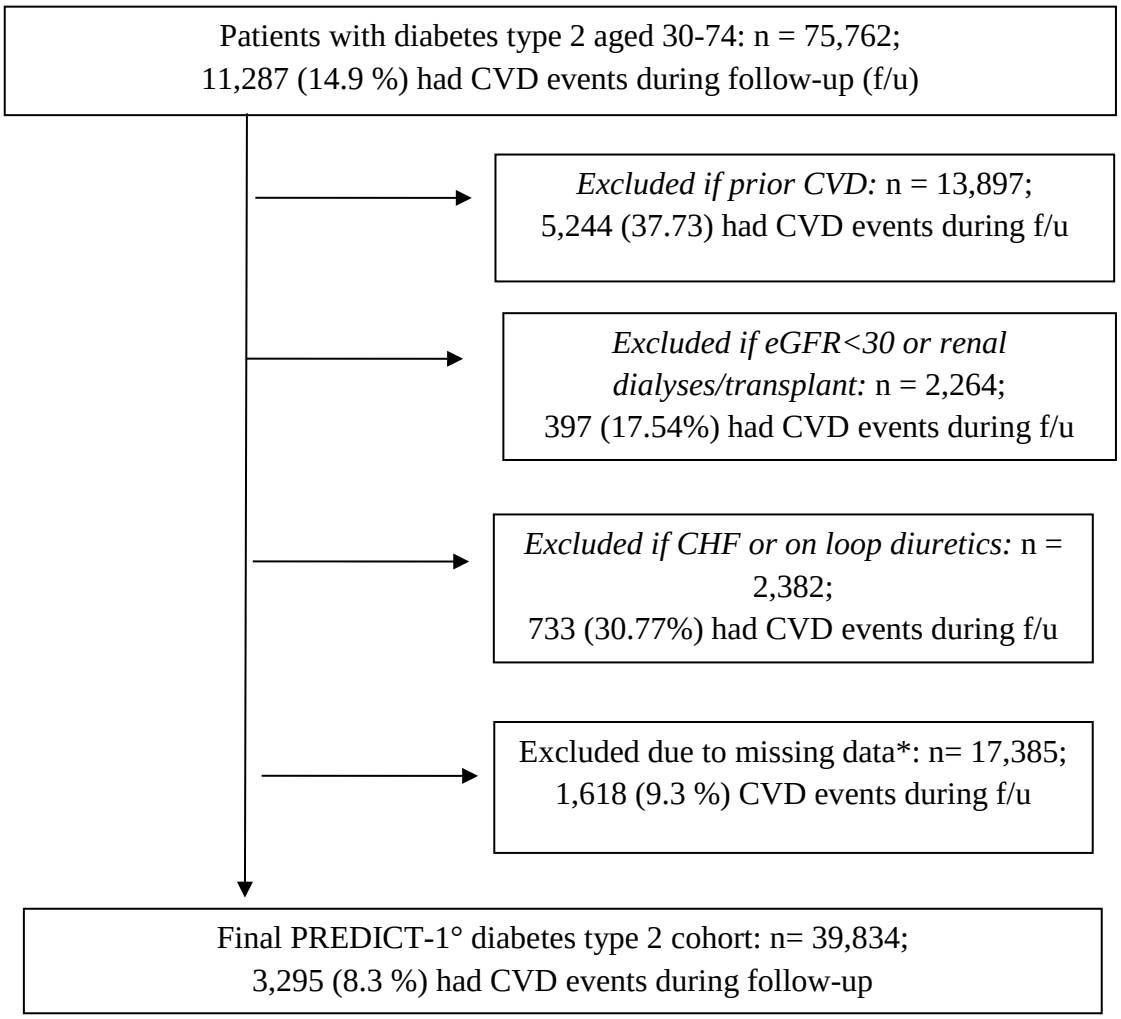


Figure 7.1. Flowchart of exclusions, PREDICT-1º T2D cohort aged 30-74 years

*The number of patients with missing data on each predictor: albumin to creatinine ratio – 15,295; years since T2D diagnosis – 5,249; HbA_{1c} 4,435; eGFR – 2,398; BMI – 2,264; TC:HDL – 27; nzdep – 3; smoking – 1.

The baseline description of the population is presented in Table 7.1. The crude annual rate of fatal or non-fatal CVD events was 14 per 1000 person-years in women and 20 per 1000 person-years in men.

Table 7.1. Baseline characteristics of the PREDICT-1º T2D cohort, aged 30-74 years

	Women	**Men**
Participants; n (% of total cohort)	19,340 (48.6)	20,494 (51.5)
Incident CVD events; n (% of sex-specific cohort)[a]	1,291(6.7)	2,004 (9.8)
Total person-years observed	93528.56	101081.9
Crude incidence of CVD (per 1000 per year)	13.8 (13.1, 14.6)	19.8 (19.0, 20.1)
Follow-up time in years	4.84 (2.5)[b]	4.93 (2.41)[b]
People with follow up ≥5 years	9135 (47.2)	9904 (48.3)
Age in years, y	54.2 (10.7)	53.9 (10.6)
Years since diagnosis, y	5.2 (5.5)	4.8 (5.2)
Self-identified ethnicity: %		
European	6,240 (32.3)	8,018 (39.1)
Māori	2,842 (14.7)	2,806 (13.7)
Pacific	5,765 (29.8)	4,691 (22.9)
Indian	3,234 (16.7)	3,547 (17.3)
Chinese	1,259 (6.5)	1,432 (7.0)
NZ Deprivation quintile: %		
1 (least deprived)	2,041 (10.6)	2,850 (13.9)
2	2,570 (13.3)	3,075 (15.0)
3	2,902 (15.0)	3,355 (16.4)
4	4,358 (22.5)	4,466 (21.8)
5 (most deprived)	7,469 (38.6)	6,748 (32.9)
Current smoker %	2,450 (12.7)	3,404 (16.6)
Family history of premature CVD %	1,970 (10.2)	1,838 (9.0)
History of atrial fibrillation %	217 (1.1)	390 (1.9)
Systolic blood pressure, mmHg	131.6 (17.7)	131.8 (16.6)
TC:HDL	4.0 (1.2)	4.4 (1.4)
eGFR, mL/min/1.73 m^2	89.1 (18.9)	88.8 (17.8)
HbA_{1c} mmol/mol	62.2 (20.6)	62.7 (20.9)
ACR, mg/g	12.2 (57.9)	12.8 (50.6)
BMI, kg/m^2	33.4 (8.1)	31.4 (6.8)
Medications at index assessment:		
On oral hypoglycaemic agents%	12,772 (66.0)	13,066 (63.8)
On insulin %	1,903 (9.8)	1,633 (8.0)
On blood pressure lowering medications %	11,515 (59.5)	11,452 (55.9)
On lipid lowering medications %	9,938 (51.4)	11,070 (54.0)
On antithrombotic medications %	7,216 (37.3)	8,452 (41.2)

[a] Values are means and standard deviations unless otherwise stated.

BMI = body mass index. ACR = urinary albumin to creatinine ratio. eGFR = estimated glomerular filtration rate; HbA_{1c} = haemoglobin A1c; TC:HDL = Total Cholesterol to HDL Cholesterol ratio; CVD = cardiovascular disease.

[b] Follow-up time ranged from one day to 13 years, in both men and women.

About 82% of the cohort were on one or more classes of medications at baseline assessment. Nearly 67% were on oral hypoglycaemic agents (OHG) or insulin, and 73% were on at least one of the lipid, blood pressure lowering or antithrombotic drugs.

Compared to men, there were higher proportions of women in the highest quintile of deprivation and more women were treated with oral hypoglycaemic agents, insulin or blood pressure lowering medications at baseline. Mean time since diagnosis of T2D was also longer among women.

Myocardial infarction was the most common outcome (34%), with coronary heart disease accounting for half of all events. Stroke and transient ischemic attacks accounted for 22% of events, congestive heart failure for 15.3%, and peripheral vascular disease for 9% (Table 7.2).

Table 7.2. Number and type of first CVD events in PREDICT-1^{o} T2D cohort, aged 30-74 years

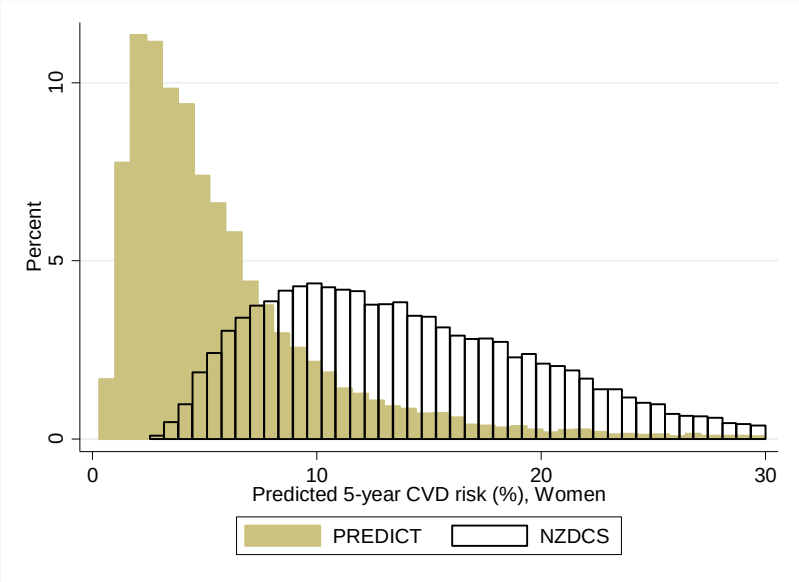
Outcome type	**Non-Fatal events, n**	**Fatal events, n[a]**	**Proportion of all CVD events, %**
Myocardial infarction	1,039	74	33.8
Unstable angina	499	1	15.2
Other coronary heart disease	72	41	3.4
Ischaemic stroke	414	29	13.4
Haemorrhagic stroke	71	41	3.4
Transient ischemic attack	170	0	5.2
Peripheral vascular disease	293	4	9.0
Congestive heart failure	482	22	15.3
Other CVD-related deaths	n/a	43	1.3
All CVD events (N = 3,295)[b]	3040	255	100

[a] If a participant died within 28 days of a non-fatal CVD event, the event was counted as fatal.

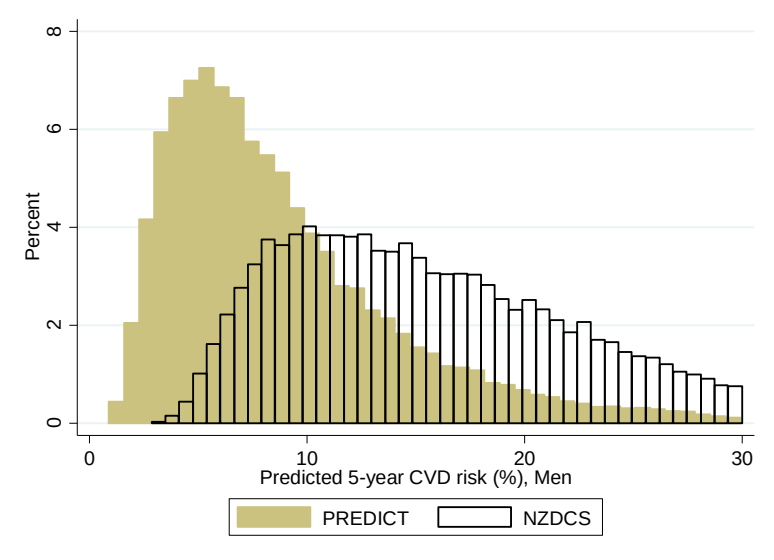
[b] If a participant had more than one type of CVD event, only the first was counted.

Validation and recalibration of the NZDCS model in the PREDICT-1^{o} T2D cohort

There were major differences between the predicted CVD risk distributions of T2D participants in the PREDICT-1^{o} T2D cohort estimated by the original NZDCS and the PREDICT-1^{o} T2D equations (Figure 7.2). The interquartile ranges and medians of the NZDCS scores were over twice as high as those of PREDICT-1^{o} T2D scores.



7.2a. Women

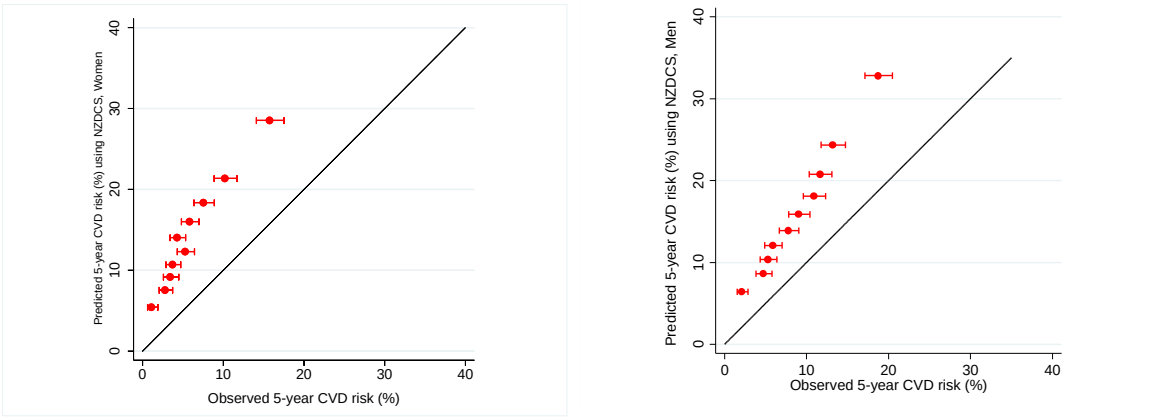


7.2b. Men

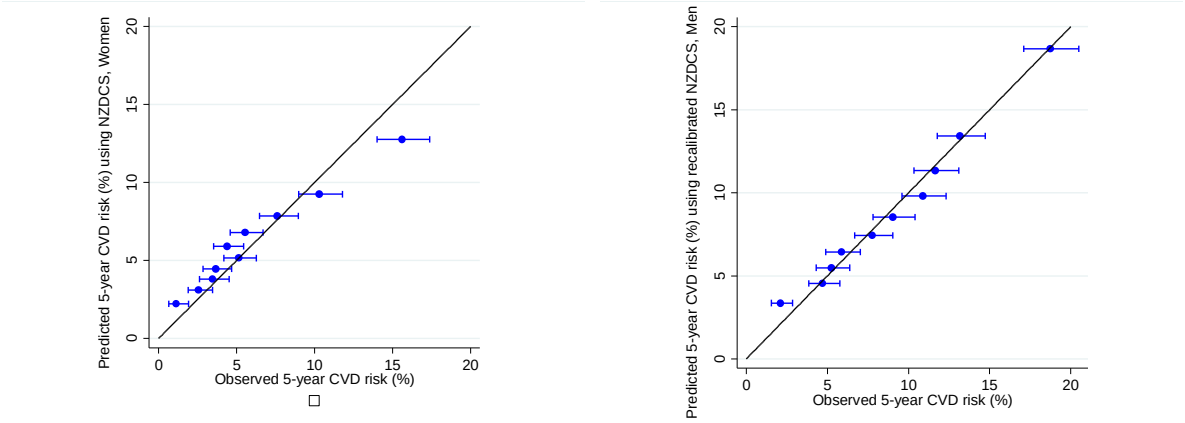
5-year CVD risk in women (PREDICT): mean = 6.1%, median = 4.5% (IQR: 2.7%, 7.4%)
5-year CVD risk in women (NZDCS): mean = 14.4%, median = 13.2% (IQR: 9.2%, 18.3%)
5-year CVD risk in men (PREDICT): mean = 9.2%, median = 7.5% (IQR: 5.0%, 11.5%)
5-year CVD risk in men (NZDCS): mean = 16.4%, median = 14.9% (IQR: 10.4%, 20.8%)

Figure 7.2. Distribution of PREDICT-1º T2D and NZDCS risk prediction scores, by sex

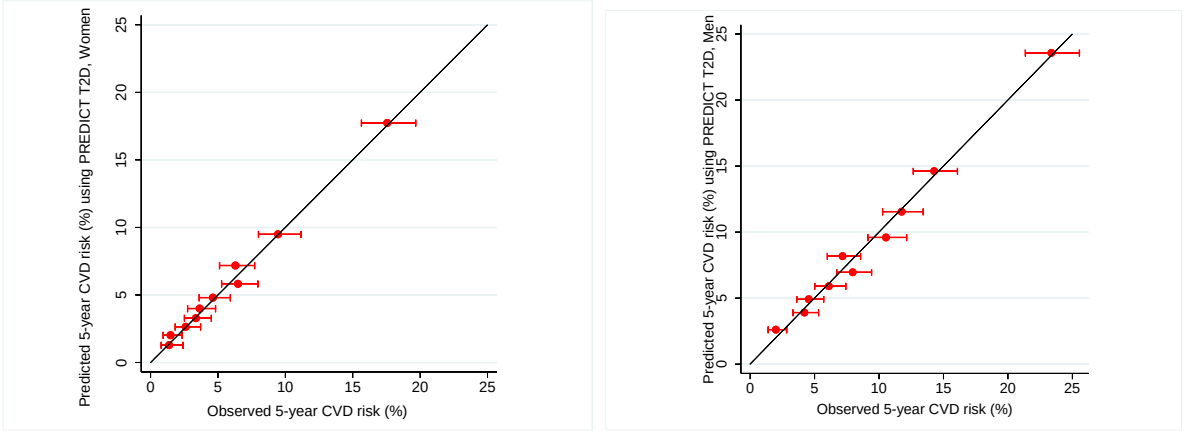
The calibration graphs (Figure 7.3a.i and 7.3b.i) also illustrate this significant over-prediction of risk in both sexes. Recalibration of NZDCS (Figure 7.3a.ii and 7.3b.ii) substantially improved the calibration plot in men, and, to a lesser extent, in women.



i. Original NZDCS



ii. Recalibrated NZDCS



iii. PREDICT-1º T2D models

Figure 7.3. Calibration of the original NZDCS, recalibrated NZDCS and PREDICT-1º T2D models, in women (left) and men (right) aged 30-74 years

The calibration slope suggested NZDCS model was underfitted in women (1.32 (1.22, 1.42)), but reasonably well fitted in men (1.04 (0.96, 1.12)).

Assessment of the additional predictors included in the PREDICT-1^{o} T2D models, but not in the NZDCS model, showed that SES, eGFR, BMI, atrial fibrillation, and treatment with insulin were statistically significant in both sexes; whereas lipid lowering treatment was significant in women only, and treatment with oral hypoglycaemic agents in men only (Table 7.3).

Table 7.3. Hazard ratios for the additional predictors from the PREDICT-1^{o} T2D models, appended to the NZDCS model

	Women	**Men**
NZDCS prognostic index*	1	1
BMI, kg/m^2	1.013 (1.005, 1.020)	1.017 (1.010, 1.024)
eGFR, mL/min/1.73 m^2	0.991 (0.988, 0.994)	0.997 (0.994, 0.999)
Atrial fibrillation	2.504 (1.828, 3.432)	1.546 (1.214, 1.969)
Family history	1.108(0.924, 1.329)	1.173 (1.006, 1.367)
Oral hypoglycemic drugs	0.991 (0.866, 1.134)	0.847 (0.763, 0.940)
Insulin	1.338 (1.130, 1.584)	1.354 (1.172, 1.564)
Lipid lowering treatment	0.836 (0.731, 0.956)	0.901 (0.810, 1.002)
Antiplatelet/anticoagulants	1.005 (0.885, 1.141)	0.976 (0.882, 1.081)
NZ Deprivation index (per 1 quintile)	1.086 (1.038, 1.136)	1.065 (1.030, 1.101)

*NZDCS prognostic index included in Cox regression models as an offset term.

PREDICT-1^{o} T2D model development and specification

The final sex-specific PREDICT-1^{o} T2D models included 16 predictors:

- Age at risk assessment (in years, continuous)
- Ethnicity (European, Indian, NZ Māori, Pacific, Chinese)
- Current smoking (yes/no)
- Albumin to creatinine ratio (continuous)
- Duration of diabetes (in years, continuous)
- eGFR (continuous)
- HbA_{1c} (continuous)

- SBP (continuous, average of two measures)
- TC:HDL (continuous)
- BMI (kg/m^2 continuous)
- Deprivation quintile (ordinal, treated as continuous)
- Binary treatment indicators: oral hypoglycaemic, insulin, blood pressure lowering, lipid lowering treatments, and treatment with antithrombotic medications.

No interactions were identified between variables in the final models, however, based on MFP analyses, a logarithmic transformation was applied to ACR.

After adjusting for all other covariates, surprisingly, Pacific ethnicity in women was associated with 24% reduction of risk compared with European women. The effect estimates in the final sex-specific models provided no evidence of increased risk associated with Māori ethnicity in either sex, and with Pacific ethnicity in men. However, Indian ethnicity was associated with 20% higher risk in women and 38% higher risk in men, relative to Europeans. Also, Chinese and other Asian peoples were at 30% lower risk than Europeans. Treatment with insulin was associated with 30% increase in risk in women, and 28% increase in men, while treatment with oral hypolycaemic agents was associated with 12% decrease in risk in men only. Family history of premature CVD was associated with 19% higher risk in men only, and history of atrial fibrillation was associated with more than twofold increase in risk in women, and 66% increase in men. Estimated risk increased by 7% in women and by 6% in men per quintile of the social deprivation score. Current age, duration of known T2D, being a smoker, systolic blood pressure, TC:HDL, ACR, BMI, eGFR, HbA_{1c}, were all statistically significant predictors, as was blood pressure lowering treatment, but not lipid lowering or antithrombotic treatments.

The hazard ratios for the final models are presented in Table 7.4.

Table 7.4. Adjusted[a] hazard ratios for first cardiovascular event, by sex

	Adjusted hazards ratios (95% CI)	
	Women	Men
Current age, y	1.039 (1.031, 1.047)	1.042 (1.036, 1.049)
Years since diagnosis of T2D, y	1.014 (1.004, 1.024)	1.012 (1.004, 1.021)
Ethnicity:		
European/other	1	1
Māori	1.016 (0.850, 1.216)	1.049 (0.906, 1.214)
Pacific	0.761(0.642, 0.901)	0.894 (0.779, 1.027)
Indian	1.204 (1.003, 1.445)	1.377 (1.199, 1.583)
Chinese	0.702 (0.496, 0.995)	0.728 (0.562, 0.942)
NZ Deprivation quintile (per 1 quintile)	1.067 (1.017, 1.119)	1.060 (1.022, 1.100)
Smoking:		
Non-smoker	1	1
Current smoker	1.630 (1.398, 1.901)	1.294 (1.152, 1.452)
Family history of premature CVD	0.991 (0.834, 1.177)	1.186 (1.025, 1.374)
History of atrial fibrillation	2.375 (1.731, 3.258)	1.660 (1.314, 2.097)
SBP, mmHg	1.011 (1.007, 1.014)	1.007 (1.004, 1.010)
TC:HDL	1.130 (1.087, 1.175)	1.074 (1.043, 1.106)
eGFR, mL/min/1.73 m^2	0.991 (0.988, 0.994)	0.995 (0.993, 0.998)
ACR, mg/g [b]	1.210 (1.165, 1.248)	1.168 (1.136, 1.201)
HbA_{1c}, mmol/mol	1.008 (1.006, 1.011)	1.007 (1.005, 1.009)
BMI, kg/m^2	1.009 (1.001, 1.017)	1.010 (1.003, 1.018)
Treatment status at index assessment:		
On oral hypoglycaemic medications	0.994 (0.866, 1.141)	0.882 (0.791, 0.983)
On insulin	1.300 (1.095, 1.543)	1.276 (1.100, 1.481)
On blood pressure lowering medications	1.193 (1.027, 1.386)	1.130 (1.008, 1.267)
On lipid lowering medications	0.862 (0.756, 0.984)	0.977 (0.878, 1.088)
On antithrombotic medications	1.036 (0.917, 1.171)	0.999 (0.904, 1.103)

[a] Hazard ratios are adjusted for all other variables included in the model.
[b] The hazard ratios for ACR are for scaled (/1000), centered and log transformed variable (natural log).

The model equation and a calculation example are supplied in Table 7.5.

Table 7.5. Derivation of absolute risk, with example calculation

Variable	Coefficient (Women)	Coefficient (Men)	EXAMPLE CALCULATION[a]		
			Woman	Coefficient x Variable[#]	Product
Current age	0.0382644	0.0415537	55 years	0.0382644x c.Age[#]	0.03062340
			European		
Māori	0.0162543	0.047768		0	
Pacific	-0.2734977	-0.1115183		0	
Indian	0.1853238	0.3201267		0	
Chinese	-0.3533301	-0.3174462		0	
NZDep quintile	0.0645955	0.058332	Quintile 3	0.0645955 x c.NZDep[#]	-0.0422309
Current smoker	0.4887195	0.2574087	No	0	
Fam. hist. CVD	-0.0089665	0.1709664	No	0	
AF	0.8650354	0.5068463	No	0	
Years since diagnosis of T2D	0.0137565	0.0121146	10	0.0137565 x c.YST2D	0.06567555114
SBP	0.0107228	0.0067902	140mmHg	0.0107228 x c.SBP[#]	0.03419985483
TC:HDL	0.1221713	0.0716341	5 units	0.1221713 x c.TC:HDL[#]	0.1233910106
eGFR	-0.009094	-0.0045729	90	-0.009094 x c.eGFR	-0.00805623
ACR	0.1869642	0.1552225	2.5	0.1869642 x ln c.ACR	-0.29561806
HbA_{1c}	0.0083445	0.0070353	64	0.0083445 x c.HbA_{1c}	0.01460787995
BMI	0.0088516	0.0103602	30	0.0088516 x c.BMI	-0.03004606
BPL	0.1768606	0.122079	Yes	0.1768606 x 1 (OBPLM)	0.1768606
LL	-0.1479567	-0.0228048	No	0	
Antithrombotic	0.0352237	-0.0013748	No	0	
Oral hypogl	-0.0061383	-0.1253885	Yes	-0.0061383 x1 (OG)	-0.0061383
Insulin	0.2625742	0.2440459	No	0	
			Sum Coefficients x Variables		0.06326876543

Means for centering			[#]Calculated centred variable for example*	
Age	54.19968976	53.88225822	c.Age = 55 – 54.19968976	0.80031054
NZDep Quintile	3.65377456	3.448277545	c.NZDep = 3 – 3.65377456	-0.6537746
SBP	131.8105481	132.0015126	c.SBP = 135 – 131.8105481	3.1894519
TC:HDL	3.990016388	4.376096985	c.TC:HDL = 5 - 3.990016388	1.00998361
eGFR	89.11411629	88.83651737	c.eGFR=90 - 89.11411629	0.88588371
HbA1c	62.24940021	62.67839856	c. HbA1c = 64 - 62.24940021	1.75059979
Yrs since T2D	5.225853154	4.848833805	c. Yrs since T2D= 10 - 5.225853154	4.77414685
BMI	33.39442089	31.41335025	c.BMI= 30 - 33.39442089	-3.3944209
ACR	ln((ACR+0.009999999776482 6)/1000)+4.406324583	ln((ACR+0.009999999776482 6)/1000)+4.355670634	ACR=ln((2.5+0.0099999997764826)/1000) +4.406324583	-1.58114794
Baseline survival function (at 5 years)			Calculating 5-year CVD risk (1-Baseline surv $^{\exp(\text{sum of coefficients x variables})}$) x 100	
For women = 0.9415411			= (1- 0.9415411 $^{\exp(0.06326876543)}$) x 100 = 6.22%	
For men = 0.9084156				

[a] Example 5-year CVD risk calculation for a 55-year European women, in the middle social deprivation quintile (i.e. quintile 3), no family history of premature CVD, a non-smoker, no atrial fibrillation, systolic blood pressure = 140mmHg, TC:HDL = 5, 10 years since diagnosis T2D, eGFR=90, ACR=2.5, HbA_{1c}=64, BMI=30, on metformin and blood pressure lowering treatment.

[b] Abbreviations: BMI = body mass index. ACR = urinary albumin to creatinine ratio. eGFR = estimated glomerular filtration rate; HbA_{1c} = haemoglobin A1c; TC:HDL = Total Cholesterol to HDL Cholesterol ratio; CVD = cardiovascular disease; AF = atrial fibrillation

Models' performance

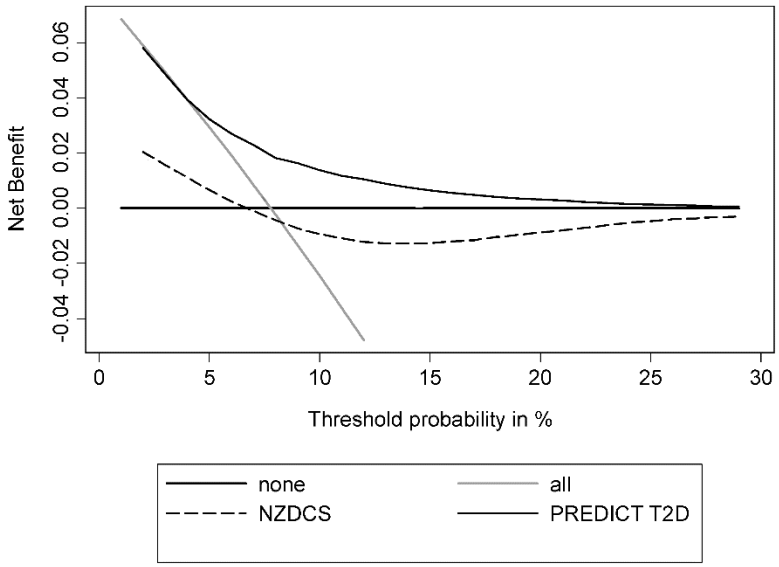
Both calibration and discrimination performance of PREDICT-1º T2D models were superior to the NZDCS model. The calibration plots for PREDICT-1º T2D models (Figure 7.3) showed that estimated event rates were close to observed, in both men and women. All standard performance statistics were statistically significantly better for the PREDICT-1º T2D models (Table 7.6).

Table 7.6. Performance statistics for the PREDICT-1° T2D and NZDCS models, by sex

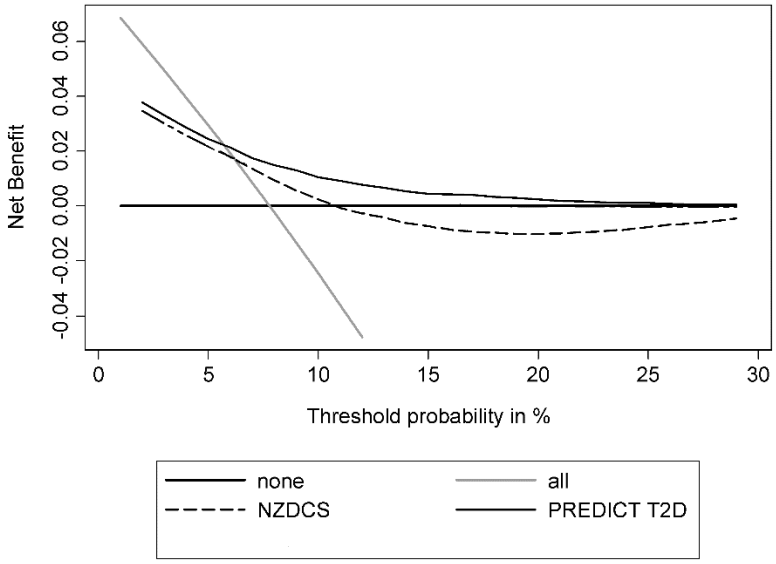
	PREDICT-1º T2D	**NZDCS**
Women		
R^2	30 (27, 33)*	22 (19, 24)*
Harrell's C statistic	0.72 (0.70, 0.73)	0.67 (0.66, 0.69)
Royston's D statistic	1.346(1.254, 1.438)*	1.072(0.988, 1.156)*
Calibration slope	1 (0.93, 1.07)	1.32 (1.22, 1.42)
Men		
R^2	22 (20, 24)*	15 (13, 17) *
Harrell's C statistic	0.68 (0.67, 0.69)	0.65 (0.64, 0.66)
Royston's D statistic	1.082 (1.009, 1.155)	0.867 (0.802, 0.932)*
Calibration slope	1 (0.94, 1.06)	1.04 (0.96, 1.12)

*95% confidence intervals were calculated using 5000 bootstrap replicates.

The decision curve graphs (Figure 7.4) show a net benefit of using PREDICT-1º T2D compared to treating everyone over a treatment threshold range of 5% to 22% in women and 7% to 20% in men. There was no benefit from using the NZDCS in women whereas for men there was a small net benefit in the treatment threshold range from 7% to 11%.



7.4a. Women



7.4b. Men

Figure 7.4. Net benefit curves for PREDICT-1º T2D and NZDCS models

The results of the sensitivity analyses where predictors with missing data (ACR, eGFR, years since diagnosis of diabetes) were categorised (categories defined by clinical considerations) and include 'unknown' category, are presented in Appendices 7.1-7.4. The effects estimates and performance of the models with categorised predictors were comparable with those of the main models, although the proportion of explained variation, D-statistic and Harrell's C were slightly lower in the models with categorised predictors.

Discussion

Summary of findings

The new models for estimating CVD risk presented in this chapter are derived from a large and diverse cohort of adults with T2D, typical of those being seen in routine primary care practice. Half of the cohort were followed up for 5 years or longer. The study population included the entire age range eligible for CVD risk assessment according to New Zealand guidelines (10), with both newly diagnosed and long-term T2D patients. The new equations build on a previous equation developed in the New Zealand T2D registry from the early 2000s and include predictors representing duration of diagnosed diabetes, renal function, body composition, socioeconomic status, ethnicity, history of atrial fibrillation and medications to control blood sugar, lipids and blood pressure. The new PREDICT-1^{o} T2D models compared favourably with the NZDCS model across all performance indicators when validated in the PREDICT-1^{o} T2D cohort: discrimination, calibration and clinical utility. The most striking observation was the substantial over-prediction of risk when the NZDC equation was applied to the PREDICT-1^{o} T2D cohort.

Interpretation of findings

Applying the NZDCS to the PREDICT-1^{o} T2D cohort showed that the NZDCS score overestimated risk across all deciles, in both men and women. While this is not surprising given the differences in population characteristics of the two cohorts, the degree of over-prediction was greater than expected. Compared with PREDICT-1^{o} T2D, the NZDCS cohort was substantially older, included people with advanced chronic kidney disease, had fewer exclusions related to history of CVD, and therefore had a higher underlying risk. The fact that recalibration to the PREDICT-1^{o} T2D baseline survival function improved NZDCS calibration supports this explanation.

Māori and Pacific people are known to be at increased risk of developing CVD and, in the general population equations, Māori or Pacific ethnicities were independent predictors of risk (Chapter 5). Yet no increased risk was observed in people with T2D in these ethnic groups, after adjusting for the other risk predictors included in the PREDICT-1^{o} T2D models. This suggests that their increased risk can be accounted for by the additional predictors in the PREDICT-1^{o} T2D compared to the general population models (i.e. renal function, HbA1c, BMI, NZDep). This explanation is supported by the initial exploratory analyses where age-adjusted models were fitted with ethnicity as the only variable, and both Pacific and Māori ethnicities were associated with increased risk. However, in the final PREDICT-1^{o} T2D models, Indian and Chinese ethnicities were associated with increased and decreased risk, respectively, therefore having ethnicity as predictor in the models was justified.

Exclusion of patients with missing values may lead to biased results. However, the sensitivity analysis, which included participants with missing data via an 'unknown' category, produced very similar models, in terms of predictor weights and performance (Appendices 7.1-7.4). Therefore the complete case final models presented here are unlikely to be significantly biased.

Findings in the context of existing research

Unlike other equations, these new equations included two manifestations of kidney disease, albuminuria and eGFR , which represent renal damage and renal function, and are independent risk factors for CVD (154). It is therefore appropriate to include both in risk prediction models although a recent systematic literature review (152) did not identify any models including both measures. Also, no other diabetes-specific models have included treatment with oral hypoglycaemic drugs and insulin, and very few have included blood pressure lowering treatment (3 out of 12), or lipid lowering treatment (1 out of 12). Furthermore, while only three out of twelve models included BMI, the results presented here show that body composition is an independent predictor of CVD risk.

In contrast to many other studies, data-driven approaches to variable selection were not used and predictors were chosen *a priori*, based on prior research and clinical judgement. This reduces the chance of accidental findings that are not generalisable to other populations (38).

Like in all modern T2D cohorts used to derive CVD risk prediction equations, the majority of participants were taking CVD preventive medications, and the choices of treatment are potentially important predictors in their own right. Therefore, these baseline medications were

included as predictors. As most patients were already treated at baseline, the potential impact of new treatments initiated during follow-up on risk, was likely to be relatively small.

Strengths and limitations

One of the main strengths of this study is the contemporary and representative population of patients eligible for CVD risk assessment that was used to derive the models. In 2012, the Ministry of Health introduced The More Heart and Diabetes Checks health target, which included a cardiovascular risk assessment and a blood test for diabetes (HbA_{1c}), delivered in primary care settings (90). The goal was for 90 percent of people in specified age and ethnicity cohorts to have had their risk assessed in the past five years, and this goal was achieved in September 2016. Another advantage of this study was that about half the participants had been followed up for 5 years or longer. Also, this study had the benefit of a large number of events per predictor variable: 65 in women and 100 in men, therefore far exceeding the 'rule of thumb' of 10 events per predictor (155). Furthermore, the PREDICT-1^{o} T2D models included additional predictors such as medication indicators, history of atrial fibrillation, BMI, eGFR, ethnicity and socioeconomic status, which are not commonly included in similar models.

This study also has a number of limitations. Firstly, information on diet and physical activity was not available, though these aspects of lifestyle do influence CVD risk and it would be preferable to include them as predictors. Of note, approximately 30% of patients in the cohort were not receiving a diabetes-related medication, implying that their diabetes was being managed by lifestyle. The relatively large proportion in this category is most likely due to the introduction of universal screening for diabetes as part of a CVD risk assessment in New Zealand from about 2010. As a result, many people with diabetes are now diagnosed early and this could account in part for the over-prediction of CVD risk using the older NZDCS equation. This older equation was derived in a population of people whose diabetes was diagnosed before universal screening and who were therefore more likely to have been diagnosed later in the progression of their diabetes.

Secondly, T2D was defined via diagnoses recorded in primary care, as well as in hospital records and in drug dispensing records. This means that the study population may include patients with type 1 diabetes (both due to misdiagnosis and recording error), although proportion of these is likely to be small.

Finally, missing data in some of the predictors resulted in exclusion of a substantial proportion of participants. However, sensitivity analysis was performed, which included patients with missing data, and showed that these alternative models were very similar.

Implications for research and practice

These CVD risk models were developed to support clinical decision making on primary prevention of CVD in patients with T2D, in routine primary care practice. They use diabetes-relevant information that is already routinely-collected in primary care and, as computerised risk calculators and electronic health records are now widely used in New Zealand, risk calculation can be automated thus reducing practitioner time burden. The full details of the sex-specific models have been provided to enable external validation. Validation of another New Zealand model, the NZDCS model, developed in a cohort of people from a Diabetes register recruited in the early 2000s, showed that differences in population characteristics, and when a cohort is recruited, can have a major influence on model performance.

In conclusion, well-performing diabetes-specific CVD risk prediction models can be developed from a representative, contemporary cohort of primary care patients with T2D. Models developed from older cohorts (both chronologically and in terms of participants' age), with higher underlying risk, such as the NZDCS score, do not perform well in a lower risk general patient population, even after recalibration.

Summary

- Sex-specific models for estimation of CVD risk were developed using a modern and representative population of primary care patients with type 2 diabetes.
- Performance of these models was compared with that of an existing diabetes-specific model developed from diabetes registry data and proved to be superior, in all measures of statistical performance and clinical utility.
- This study supports use of CVD risk estimation models derived from populations that are similar to the populations the models will be applied to.

Chapter 8. Comparison of General Population and Diabetes-Specific Models to Estimate CVD Risk

Introduction

Chapter 7 described the development and assessment of diabetes-specific models to estimate CVD risk; it also compared performance of this model with the performance of an alternative diabetes-specific model derived from the New Zealand Diabetes Cohort Study using New Zealand diabetes registry data. This chapter will compare the performance of diabetes-specific models presented in Chapter 7 and the general population models presented in Chapter 6.

Diabetes is an important risk factor for CVD (156) and influences the risk of CVD through several mechanisms. In addition to the atherosclerotic pathways of CVD, there is the non-atherosclerotic diabetic cardiomyopathy component in cardiovascular outcomes (156-158). Also, there is evidence that dyslipidaemia in diabetes is qualitatively different from dyslipidaemia in the general population (159). Lipid metabolism is often affected by insulin resistance and, in conjunction with diabetes-associated obesity, leads to formation of fatty deposits in the arteries and increased tendency to oxidation (160). Diabetes is associated with abnormal haemostasis (blood clotting) that could be responsible for additional CVD risk, including diminished fibrinolytic activity, increased platelet aggregation and adhesiveness, and raised concentrations of fibrinogen (161-167). Theoretically, there is a possibility of additional biological mechanisms for increased risk of CVD, and of different roles of established CVD risk factors, in people with T2D. For example, a recent meta-analysis of studies in the Emerging Risk Factors Collaboration investigated associations between diabetes and HRs for vascular disease. The authors found that the effect size was significantly greater in some groups at lower absolute risk of CVD: in women, younger ages, non-smokers and those with lower than average blood pressure (168).

The varying effect of diabetes in men and women was recognised in early Framingham models (9), which included an interaction term between diabetes and sex indicating increased risk in women (who are usually at lower CVD risk relative to men) if they have diabetes. The approach has since been to build sex-specific models as the possibility of other sex differences cannot be ruled out. However, as discussed above, T2D can also modify the effects of other known (and new) predictors.

Developing general models with interaction terms versus sub-population-specific models

By definition, risk factors interact when the presence of one factor modifies the effect that another factor has on the occurrence of disease. The term ‘effect modification’ is often used in

reference to such interactive phenomena (169). An analysis that includes the assessment of interactions allows a fuller description of how event or outcome incidence varies as a function of risk factors, than would an analysis that only considers overall or average effects of risk factors. Therefore, the inclusion of interaction terms can improves one's ability to predict disease on the basis of an individual's profile for a set of risk factors (169). At the same time, there are several practical and theoretical reasons why it is often impossible to discern from data how and even whether two risk factors interact in biologically meaningful ways. Some of these are the unmeasured intervening variables, numerous possibilities of independent, antagonistic or synergetic effects given the typically unknown nature of underlying pathological process, and various possible functional forms of relationships (169). Different mechanisms of interaction may predict the same patterns of disease. Also, measurement errors, including random errors, may significantly impede interaction assessment (170).

As has been noted, for predictive models, the rules of model-building and evaluation of interaction terms are not necessarily restrained by considerations of biological mechanisms or interpretation of individual terms in the model (171). And yet, prediction may be more accurate if the mathematical model corresponds well to the biological process. For this reason, the biological accuracy of a model should be considered along with statistical convenience and simplicity, even for models used exclusively for prognostic purposes (172).

In regression models, the common statistical approach is to define interactions as multiplication products in the model. The limitations of such approaches have been described in detail (171-173). In brief, these approaches lack any consideration of what constitutes interaction on the biological level. Also, it renders the presence or absence of interaction entirely dependent on the form of the statistical model chosen: for the same data, interactions may appear to be present when using one model but absent when using another.

An alternative approach to dealing with interactions is to stratify the population by the effect modifying variable, such as developing CVD risk prediction models using a sub-population of primary care patients with T2D. In addition to the study population being homogeneous with regard to the diabetes status, there is often considerable accompanying diabetes-specific data (not usually available for non-diabetic patients) that can be useful for CVD risk prediction in this sub-population.

As discussed in the previous chapter, over the last few decades, many risk calculators have been developed, with only a small fraction intended specifically for patients with type 2 diabetes, although many include diabetes as one of predictors (153). According to a recent systematic literature review (152), the most commonly used predictors in models developed in diabetes populations were age, sex, duration of diagnosed diabetes, HbA_{1c} concentration and smoking, and they mostly predicted 5-year risk, while in models developed from general populations, in addition to diabetes, the most frequent predictors were age, sex, systolic blood pressure, smoking and cholesterol. Unlike the diabetes-specific models, most of these predicted 10-year risk (152).

There is only limited evidence that the risk scores developed specifically in patients with diabetes estimate cardiovascular risk more accurately. Studies comparing the predictive ability of CVD risk scores developed in diabetic populations, with those developed in general populations, showed that their performance varied considerably (152, 153).

However drawing definite conclusions about model performance is problematic when comparing models developed from different studies because of between-study variation in exclusion criteria, duration of follow up, definitions of endpoints and predictors, distributions of risk factors and their treatment (including temporal trends in these) and differences in background risk of specific populations. Also, inconsistency of methods used to evaluate models in different studies makes it difficult to compare the predictive ability of different equations (153). A large number of patients with diabetes in the PREDICT cohort provides an opportunity to compare models derived from the same patient population, using a standard set of performance measures, therefore resulting in a more valid comparison than has been possible previously.

Aims and objectives

The aim of this chapter was to establish whether the PREDICT diabetes-specific or PREDICT general population models performed better in primary care patients with type 2 diabetes.

To achieve this aim, the objectives of this chapter were to:

- Evaluate the performance of PREDICT-1^{o} (general population) models in the PREDICT sub-population with type 2 diabetes (PREDICT-1^{o} T2D).

- Evaluate the performance of PREDICT-1º T2D models in the same population.
- Compare the performance indicators of PREDICT-1º models with the performance indicators for the PREDICT-1º T2D models.

Methods

The development of general population and diabetes-specific models was described in detail in Chapters 6 and 7. In brief, primary care patient data were used to generate two sets of risk scores for men and women. The general population models included diabetes as one of predictors, in addition to other established risk factors (age, sex, ethnicity, SES, systolic blood pressure, TC:HDL, AF, BPL and LL treatments, family history of premature CVD). The diabetes-specific models, using data from T2D patients only, included several additional diabetes-related predictors (duration of known diabetes, HbA1c, treatment with insulin and oral hypoglycaemic agents, estimated glomerular filtration rate (eGFR), and albumin to creatinine ratio (ACR)). Both sets were developed using sex-specific Cox regression models.

As in Chapters 6 and 7, the performance of the models is compared with regards to: discrimination, calibration and explained variation. Discrimination metrics included Harrell's C-index of concordance (61) and the D statistic, (64). Calibration was assessed in terms of the 'calibration slope' (regression on the prognostic index) (120) and graphically, by grouping the estimated 5-year risk into deciles and plotting mean predictions from the model against the mean Kaplan-Meier survival probabilities (55). Explained variation was assessed using Royston's (72) modification of O'Quigley, Xu & Stare's (74) modification of Nagelkerke's (73) R-squared statistic for proportional hazards (PH) models for censored survival data.

Clinical utility across various risk thresholds was assessed via decision curve analysis (83).

All analyses were performed using Stata13 software (122).

Results

A baseline comparison of the PREDICT-1º T2D and PREDICT-1º cohorts is presented in Table 8.1.

Table 8.1. Comparison of the PREDICT-1º T2D and PREDICT-1 º cohorts, aged 30-74 years

	PREDICT-1º T2D Cohort		**PREDICT-1º Cohort**	
	Women	Men	Women	Men
Participants; n (% of total cohort)	19,340 (48.6)	20,494 (51.5)	175,699 (43.7)	226,053 (56.3)
Incident CVD events; n (% of sex-specific cohort)[a]	1,291(6.7)	2,004 (9.8)	5,650 (3.2)	9,736 (4.3)
Total person-years observed	93,528.56	101,081.9	743,640	941,881
Crude incidence of CVD (per 1000 per year)	13.8 (13.1, 14.6)	19.8 (19.0, 20.1)	7.6 (7.4, 7.8)	10.3 (10.1, 10.5)
Follow-up time in years	4.84 (2.5)[b]	4.93 (2.41)[b]	4.2 (2.7)[b]	4.2 (2.7)[b]
People with follow up ≥5 years	9,135 (47.2)	9,904 (48.3)	58,493 (33.3)	72,417 (32.0)
Age in years, y	54.2 (10.7)	53.9 (10.6)	56 (8.9)	51.8 (9.9)
Years since diagnosis, y	5.2 (5.5)	4.8 (5.2)	NA	NA
Self-identified ethnicity: %				
European	6,240 (32.3)	8,018 (39.1)	96,032 (54.7)	128,503 (56.9)
Māori	2,842 (14.7)	2,806 (13.7)	23,853 (13.6)	27,573 (12.2)
Pacific	5,765 (29.8)	4,691 (22.9)	22,537 (12.8)	28,073 (12.4)
Indian	3,234 (16.7)	3,547 (17.3)	14,188 (8.1)	20,232 (9.0)
Chinese	1,259 (6.5)	1,432 (7.0)	19,089 (10.9)	21,672 (9.6)
NZ Deprivation quintile: %				
1 (least deprived)	2,041 (10.6)	2,850 (13.9)	38,523 (21.9)	50,379 (22.3)
2	2,570 (13.3)	3,075 (15.0)	34,230 (19.5)	44,609 (19.7)
3	2,902 (15.0)	3,355 (16.4)	31,808 (18.1)	40,684 (18.0)
4	4,358 (22.5)	4,466 (21.8)	32,626 (18.6)	41,553 (18.4)
5 (most deprived)	7,469 (38.6)	6,748 (32.9)	38,512 (21.9)	48,828 (21.6)
Current smoker %	2,450 (12.7)	3,404 (16.6)	21,703 (12.4)	37,058 (16.4)
Family history of premature CVD %	1,970 (10.2)	1,838 (9.0)	22,996 (13.1)	24,495 (10.8)
History of atrial fibrillation %	217 (1.1)	390 (1.9)	1,777 (1.0)	3,680 (1.6)
Diabetes (type1, type2 or type unknown)	NA	NA	27,377 (15.6)	30,942 (13.7)
Systolic blood pressure, mmHg	131.6 (17.7)	131.8 (16.6)	129 (17.7)	129 (16.2)
TC:HDL	4.0 (1.2)	4.4 (1.4)	3.7 (1.1)	4.4 (1.3)
eGFR, mL/min/1.73 m^2	89.1 (18.9)	88.8 (17.8)	NA	NA
HbA_{1c} mmol/mol	62.2 (20.6)	62.7 (20.9)	NA	NA
ACR, mg/g	12.2 (57.9)	12.8 (50.6)	NA	NA
BMI, kg/m^2	33.4 (8.1)	31.4 (6.8)	NA	NA
Medications at index assessment:				
On oral hypoglycaemic %	12,772 (66.0)	13,066 (63.8)	NA	NA
On insulin %	1,903 (9.8)	1,633 (8.0)	NA	NA
On blood pressure lowering medications %	11,515 (59.5)	11,452 (55.9)	45,973 (26.2)	43,253 (19.1)
On lipid lowering medications %	9,938 (51.4)	11,070 (54.0)	27,540 (15.7)	33,372 (14.8)
On antithrombotic medications %	7,216 (37.3)	8,452 (41.2)	17,831 (10.2)	21,723 (9.6)

[a] Values are means and standard deviations unless otherwise stated.

BMI = body mass index. ACR = urinary albumin to creatinine ratio. eGFR = estimated glomerular filtration rate; HbA_{1c} = haemoglobin A1c; TC:HDL = Total Cholesterol to HDL Cholesterol ratio; CVD = cardiovascular disease.

[b] Follow-up time ranged from one day to 13 years, in both men and women.

The proportion of women was higher in PREDICT-1º T2D population (49%) than in the PREDICT-1° study population (44%). In the PREDICT-1º T2D cohort, the crude annual CVD event rate was approximately 14 per 1000 person-years in women and 20 per 1000 person-years in men, compared to approximately 8 per 1000 person-years in women and 10 per 1000 person-years in men in the PREDICT-1º cohort. Less than 40% of the PREDICT-1º T2D cohort, but more than 50 % of the PREDICT-1º cohort, were Europeans. On average, blood pressure was higher in both sexes in the PREDICT-1º T2D population, and TC:HDL was higher in women with T2D. CVD preventive medications were dispensed 2-4 times more frequently in the PREDICT-1º T2D cohort compared to the PREDICT-1º cohort.

In terms of the CVD events in the two cohorts, the proportions by event subtype were relatively similar, with the exception of higher incidence of peripheral vascular disease hospitalisations in the PREDICT-1º T2D population (9%) compared with PREDICT-1° population (5.5%) (Table 8.2).

Table 8.2. Number and type of first CVD events in the PREDICT-1º T2D and PREDICT-1° cohorts, aged 30-74 years

	PREDICT-1º T2D cohort			PREDICT-1º cohort		
Outcome type	**Non-Fatal events, n**	**Fatal events, n[a]**	**Proportion of all CVD events, %**	**Non-Fatal events, n**	**Fatal events, n[a]**	**Proportion of all CVD events, %**
Myocardial infarction	1,039	74	33.8	4,984	188	33.6
Unstable angina	499	1	15.2	2,275	11	14.9
Other coronary heart disease	72	41	3.4	343	436	5.1
Ischaemic stroke	414	29	13.4	2,124	156	14.8
Haemorrhagic stroke	71	41	3.4	445	205	4.2
Transient ischemic attack	170	0	5.2	1,123	0	7.3
Peripheral vascular disease	293	4	9.0	790	62	5.5
Congestive heart failure	482	22	15.3	1,795	113	12.4
Other Ischaemic CVD-related deaths	n/a	43	1.3	n/a	336	2.2
Total CVD events (N = 15,386)[b]	3,040	255	100	13,879	1,507	100

[a] If a participant died within 28 days of a non-fatal CVD event, the event was counted as fatal.
[b] If a participant had more than one type of CVD event, only the first was counted.

The adjusted HRs from models derived in the PREDICT-1° cohort and the PREDICT-1º T2D cohort are presented in Table 8.3.

Table 8.3. Adjusted[a] hazard ratios for first cardiovascular event, in PREDICT-1º and PREDICT-1º T2D models

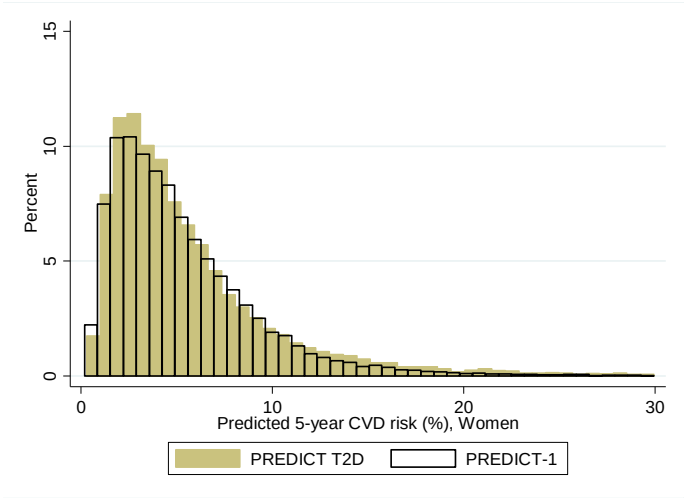
	PREDICT-1º T2D model		**PREDICT-1º model**	
	Adjusted hazards ratios (95% CI)			
	Women	Men	Women	Men
Current age, y	1.04 (1.03, 1.05)	1.04 (1.04, 1.05)	1.08 (1.07, 1.08)	1.07 (1.07, 1.07)
Years since diagnosis of T2D, y	1.01 (1.00, 1.02)	1.01 (1.01, 1.02)	N/A	N/A
Diabetes	N/A	N/A	1.72 (1.61, 1.85)	1.75 (1.66, 1.85)
Ethnicity:				
European	1	1		
Māori	1.02 (0.85, 1.22)	1.05 (0.91, 1.21)	1.48 (1.37, 1.60)	1.34 (1.26, 1.42)
Pacific	0.76 (0.64, 0.90)	0.89 (0.78, 1.03)	1.22 (1.12, 1.33)	1.19 (1.12, 1.27)
Indian	1.20 (1.00, 1.45)	1.38 (1.20, 1.58)	1.13 (1.00, 1.27)	1.34 (1.24, 1.45)
Chinese	0.70 (0.50, 0.99)	0.73 (0.56, 0.94)	0.75 (0.66, 0.85)	0.67 (0.61, 0.74)
NZ Deprivation quintile (per 1 quintile)	1.07 (1.02, 1.12)	1.06 (1.02, 1.10)	1.11 (1.09, 1.14)	1.08 (1.07, 1.10)
Smoking:				
Non-smoker	1	1		
Current smoker	1.63 (1.40, 1.90)	1.29 (1.15, 1.45)	1.86 (1.73, 2.00)	1.66 (1.57, 1.75)
Family history of premature CVD	0.99 (0.83, 1.18)	1.19 (1.02, 1.37)	1.05 (0.97, 1.12)	1.14 (1.08, 1.21)
History of atrial fibrillation	2.38 (1.73, 3.26)	1.66 (1.31, 2.10)	2.44 (2.12, 2.81)	1.80 (1.62, 2.00)
SBP, mmHg	1.01 (1.01, 1.01)	1.01 (1.00, 1.01)	1.02 (1.01, 1.02)	1.02 (1.02, 1.02)
TC:HDL	1.13 (1.09, 1.18)	1.07 (1.04, 1.11)	1.13 (1.11, 1.15)	1.14 (1.12, 1.15)
eGFR, mL/min/1.73 m^2	0.99 (0.98, 0.99)	0.99 (0.99, 0.99)	N/A	N/A
ACR[b], mg/g	1.21 (1.17, 1.25)	1.17 (1.14, 1.20)	N/A	N/A
HbA$_{1c}$, mmol/mol	1.008(1.006,1.011)	1.007(1.005,1.009)	N/A	N/A
BMI, kg/m^2	1.009(1.001,1.017)	1.010(1.003,1.018)	N/A	N/A
Treatment status at index assessment:				
On oral hypoglycaemic medications	0.99 (0.87, 1.14)	0.88 (0.79, 0.98)	N/A	N/A
On insulin	1.30 (1.10, 1.54)	1.28 (1.10, 1.48)	N/A	N/A
On blood pressure lowering medications	1.19 (1.03, 1.37)	1.13 (1.01, 1.27)	1.40 (1.31, 1.50)	1.34 (1.27, 1.42)
On lipid lowering medications	0.86 (0.76, 0.98)	0.98 (0.88, 1.09)	0.94 (0.88, 1.01)	0.95 (0.90, 1.00)
On antithrombotic medications	1.04 (0.92, 1.17)	0.99 (0.90, 1.10)	1.12 (1.04, 1.21)	1.10 (1.03, 1.17)

[a] Hazard ratios are adjusted for all other variables included in the model.
[b] The hazard ratios for ACR are for scaled (/1000), centered and log transformed variable (natural log).

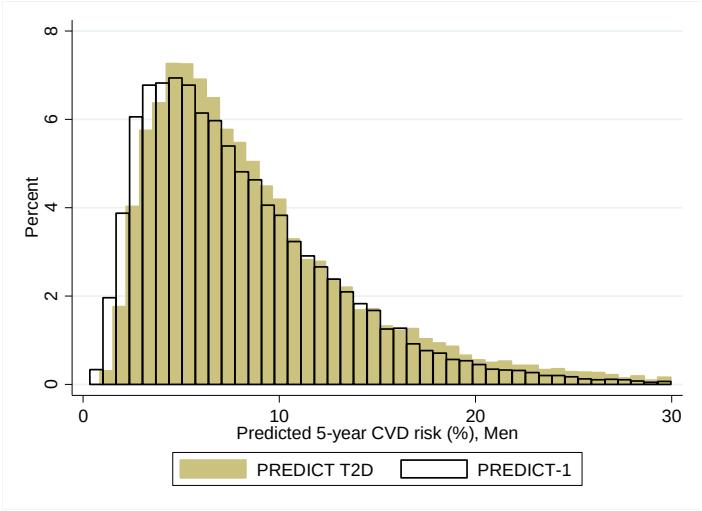
The two sets of models had overlapping predictors, with the PREDICT-1º T2D models including additional predictors relevant to diabetes. The additional variables in the PREDICT-1º T2D models were duration of known diabetes, HbA1c, eGFR, ACR, treatment with insulin and oral hypoglycaemic agents, and body mass index (BMI). Inclusion of these additional predictors influenced the weights attributed to the other predictors in the PREDICT-1° models. For example, hazard ratios (HRs) for Pacific and Māori ethnicities were associated with increased risk in PREDICT-1° models, but this effect was not apparent in the PREDICT-1º T2D models. An additional sensitivity analysis performed in PREDICT-1º T2D population showed that after removal of additional predictors from PREDICT-1º T2D models (eGFR, ACR, duration of known diabetes, treatment with insulin, BMI and deprivation), the HR estimates increased for Māori and Pacific peoples in both sexes: 1.45 for Māori and 1.31 for Pacific men, 1.45 for Māori and 1.26 for Pacific women.

Furthermore, the effects of current age and systolic blood pressure in the PREDICT-1º T2D models were attenuated compared to the PREDICT-1° cohort models in both men and women with the addition of diabetes-related predictors. The effect of TC:HDL also weakened in men in the PREDICT-1º T2D model as did the increased risk associated with blood pressure lowering treatment in both sexes with the addition of diabetes-specific variables.

The distributions of PREDICT-1º and PREDICT-1º T2D absolute risk scores were similar although the distribution (means, medians and IQRs) of the PREDICT-1º scores were consistently lower than those of PREDICT-1º T2D models (Figure 8.1).



8.1a. Women

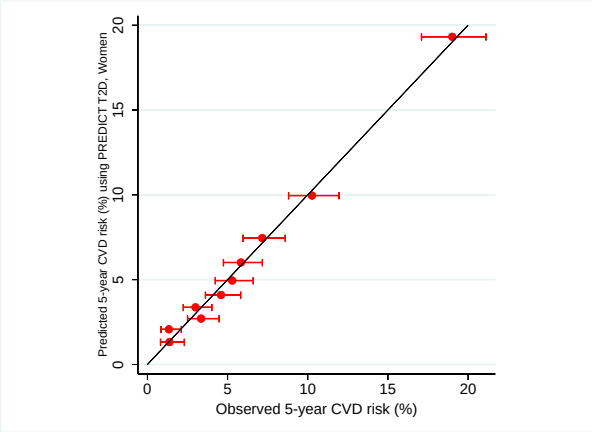


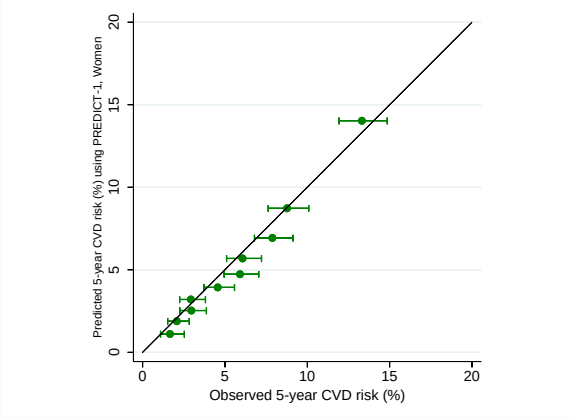
8.1b. Men

5-year CVD risk in women (PREDICT-1º T2D): mean = 6.1%, median = 4.5% (IQR: 2.7%, 7.4%)
5-year CVD risk in women (PREDICT-1º): mean = 5.3%, median = 4.3% (IQR: 2.5%, 6.9%)
5-year CVD risk in men (PREDICT-1º T2D): mean = 9.2%, median = 7.5% (IQR: 5.0%, 11.5%)
5-year CVD risk in men (PREDICT-1º): mean = 8.1%, median = 6.9% (IQR: 4.3%, 10.6%)

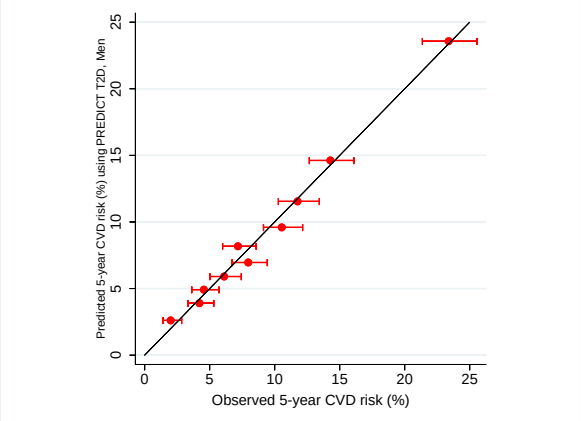
Figure 8.1. Distributions of PREDICT-1º and PREDICT-1º T2D risk prediction scores, by sex

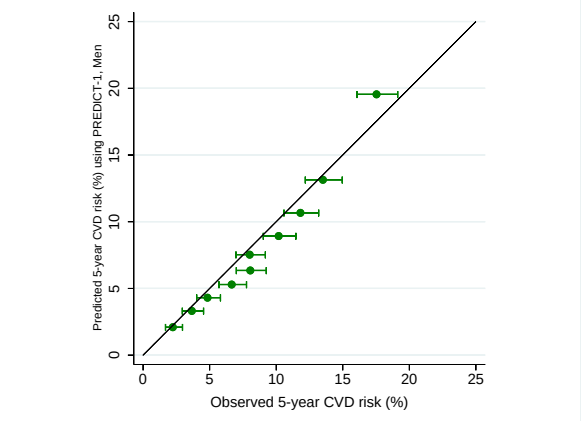
Calibration graphs (Figure 8.2) suggest that both PREDICT-1º T2D models were better calibrated compared with the PREDICT-1° models. In men, the PREDICT-1° model underestimated risk across all deciles except for the top two; in women, the PREDICT-1° model underestimated risk in deciles 1, 5, 6 and 8.





8.2a. Women





8.2b. Men

Figure 8.2. Calibration plots: PREDICT-1º T2D (left) and PREDICT-1° (right)

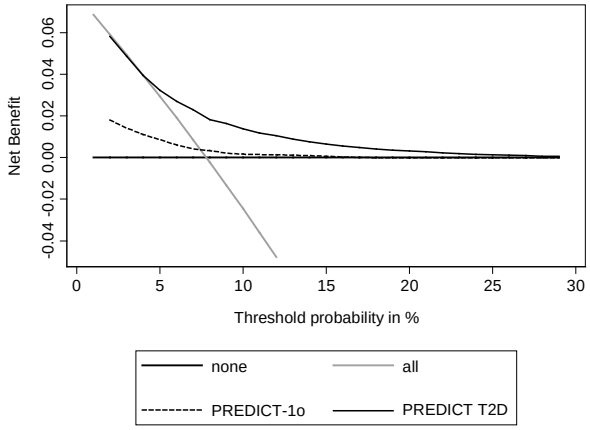
All other performance statistics indicated than PREDICT-1º T2D model was superior to PREDICT-1º models when assessed in the T2D population (Table 8.4).

Table 8.4. Performance statistics for the PREDICT-1° T2D and PREDICT-1° models, by sex

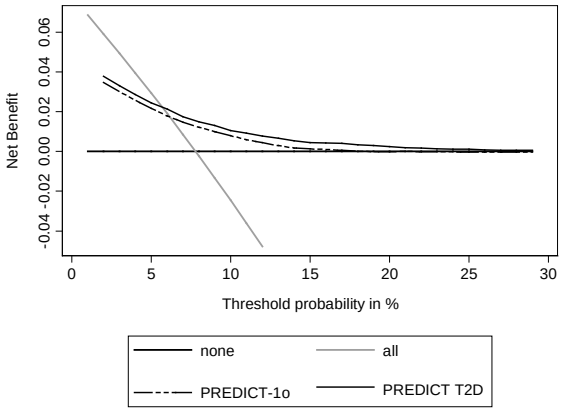
	PREDICT-1° T2D	**PREDICT-1°**
Women		
R^2	30 (27, 33)*	19 (17, 21)*
Harrell's C statistic	0.72 (0.70, 0.73)	0.67 (0.66, 0.69)
Royston's D statistic	1.346(1.254, 1.438)*	1.007 (0.929, 1.085)
Calibration slope	1 (0.93, 1.07)	0.89 (0.82, 0.96)
Men		
R^2	22 (20, 24)*	15 (13, 16)*
Harrell's C statistic	0.68 (0.67, 0.69)	0.65 (0.64, 0.66)
Royston's D statistic	1.082 (1.009, 1.155)	0.846 (0.787, 0.905)
Calibration slope	1 (0.94, 1.06)	0.81 (0.75, 0.87)

*95% confidence intervals were calculated using 5000 bootstrap replicates.

The net benefit of using PREDICT-1º T2D models was greater compared with the PREDICT-1º ones across the range of clinically meaningful threshold definitions of high risk (10, 15, and 20% 5-year risk). In women, below a 5% and above a 22% threshold definition, the PREDICT-1º T2D model provided no additional benefit beyond a treat-all or treat-none approach, respectively (Figure 8.3a). The PREDICT-1° model provided no benefit below 7% and above 12% (Figure 8.3a). In men, the PREDICT-1º T2D model was superior in the range between 6% and 20% thresholds and the PREDICT-1° model was superior between 6% and 15% risk thresholds (Figure 8.3b).



8.3a. Women



8.3b. Men

Figure 8.3. Net benefit curves for PREDICT-1º T2D and PREDICT-1º models

Discussion

Summary of findings

This chapter compared four sex-specific models for estimating CVD risk in patients with T2D, derived from the same primary care population: two PREDICT-1^{o} models, derived from the cohort including patients with and without T2D, and two PREDICT-1^{o} T2D models, derived from the sub-cohort restricted to those with T2D. While the two cohorts differed with regards to distributions of the risk factors and the incidence of CVD, they had identical predictor and outcome definitions, the same inclusion criteria (other than diabetes status), as well as the same study timeframes and methods of data collection. In addition to the predictors included in the general population PREDICT-1^{o} models, the PREDICT-1^{o} T2D models also contained several diabetes-relevant predictors. Comparative evaluation of models' performance showed that the diabetes-specific models performed better in T2D patients, compared with the general population models, across all the performance indicators.

Interpretation of findings

Important clinical and demographic differences between cohorts were to be expected. Distributions of risk factors and incidence of CVD are likely to differ in a diabetes population compared with a general population, with more adverse risk profiles in people with diabetes. It is not surprising therefore that the PREDICT-1^{o} models underestimated risk in people with diabetes since the incidence of CVD in the PREDICT-1^{o} T2D cohort was twice as high as that in the general cohort. The predictive performance of models is influenced by how similar the incidence and the distributions of the risk factors in the derivation and the validation cohorts are, and often re-calibration is undertaken to improve the agreement between the observed and predicted outcomes. Of note, the calibration of the general PREDICT-1^{o} models in the PREDICT-1^{o} T2D population was much better than for the NZDCS models described in Chapter 6.

The comparatively better discrimination of the PREDICT-1^{o} T2D models is likely to be explained by the presence of the additional diabetes-relevant predictors, as increased risk associated with diabetes depends in part on the level of glycaemic control, duration of diabetes, and renal health. This explanation is also supported by the increase in the proportion of explained variation between PREDICT-1^{o} and PREDICT-1^{o} T2D models, from 19% to 30% in women, and from 15% to 22% in men. Still, the discrimination of both sets of models was

moderate when assessed using the C-statistic, which would be influenced by the more homogeneous population of T2D patients. Furthermore, only part of CVD risk in patients with diabetes will be accounted for by the risk factors included, and other mechanisms, as well as measurement error are likely to be important.

Findings in the context of existing research

The results are consistent with the report by Guzder et al. (174) who compared Framingham (9) and UKPDS equations (175) and observed a generally superior performance of the diabetes-specific UKPDS models for prediction of CHD in the UK primary care patients with newly diagnosed T2D (176). The findings of improved discrimination after the addition of diabetes-relevant predictors are also consistent with several published reports (149, 177, 178). For example, a study that investigated substitution of diabetes with glycated haemoglobin (HbA_{1c}) in Cox models containing the original Framingham variables (age, total and HDL cholesterol, systolic blood pressure, smoking status and diabetes), using data from EPIC-Norfolk community cohort, found that replacing diabetes with HbA_{1c} improved the model's discrimination (177, 178). Similarly, the Atherosclerosis Risk in Communities (ARIC) investigators observed that adding a battery of diabetes-relevant predictors to the conventional set improved discrimination of models to estimate risk of CHD in the US community cohort (149). At the same time, an investigation by the Emerging Risk Factor Collaboration found that, in people without a history of diabetes, information about fasting glucose concentration or impaired glucose status did not significantly improve metrics of CVD prediction when added to several conventional risk factors (168). This may be due to error introduced when recording fasting or non fasting status or due to this measure being cross-sectional compared with HbA_{1c} being the measure of overall glycaemic control over 8-12 weeks. Alternatively, this might indicate that predictors relevant to the T2D population may not necessarily be useful for those without diabetes. A number of other publications also report that general population models tend to underestimate CVD risk in individuals with diabetes (153, 179-181) and consistent with this analysis, only moderate discrimination performance was observed by two other studies which validated several CVD risk scores in diabetic populations (174, 182).

The candidate is not aware of other studies comparing the two approaches by deriving both general population and diabetes-specific CVD risk models in the same study population. The two recent systematic literature reviews that investigated performance of both diabetes-specific and general population models (152, 153) considered models derived from different study

populations, focused mostly on external validation results (also mostly in different study populations) and found that performance varied considerably. It is not surprising that the results of such comparisons were inconclusive, since performance of prognostic models does depend on the characteristics of the validation population. Also, it is impossible to disentangle whether differences in performance are due to the differences in model-building approaches, or due to varying characteristics of derivation and validation populations.

Strengths and limitations

A major strength of this study was the comparison of models derived from the same primary care cohort. As a result, the models were similar with regards to measurement error and any unmeasured characteristics, as well as temporal trends in the distributions of risk factors and treatment practices. Another advantage of this study was the use of a standard approach to evaluate performance across models. Having identical definitions of outcomes in models is an advantage as well, since apparent performance can be impaired by variations in outcome definitions. Similar proportions of events by subtype in the two cohorts is potentially an advantage as well. In prospective observational studies of people with diabetes, adverse lipid profiles and microalbuminuria have been shown to have a substantial influence on all forms of atherosclerotic diseases, yet the impact of these risk factors on CHD, PVD and cerebrovascular events varied (168).

A large proportion of the T2D patients in this study population were on CVD preventive medications, as well as diabetes-related medications. However, predicting CVD risk is still very relevant, as many patients are under-treated and an assessment of current CVD risk can help inform further treatment decisions. Both PREDICT-1° and PREDICT-1° T2D models included CVD preventive drug treatment status at baseline as potential predictors; PREDICT-1° T2D models also included treatment with oral hypoglycaemic agents and insulin, along with the measure of glycaemic control.

Implications for research and practice

The general population and diabetes-specific models were developed for use in primary care, and it is important to establish whether a sub-population-specific approach improves estimation of risk. For clinical purposes, it is essential for models to have both good discrimination and good calibration. However, as discussed in Chapter 3, these are two independent aspects of the model. In case of miscalibration, recalibration is possible; but poor discrimination cannot be

fixed. This study showed that using additional information recorded in primary care in relation to diabetes which would not normally (or at all) be collected in those without diabetes, significantly improved discrimination of the CVD risk prediction models. Since the additional, diabetes-specific predictors improved the PREDICT-1º T2D models' performance indicators, the findings support the hypothesis that a sub-population-specific approach is superior. However, to confirm these conclusions, the results need to be replicated in other populations. Future research should aim to assess the performance of general population and diabetes-specific models derived in other populations. Also, exploration of additional risk predictors that could improve the models' discrimination in T2D populations further should be considered.

In conclusion, this analysis supports the hypothesis that CVD risk prediction models derived in a T2D cohort can estimate CVD risk more accurately in patients with type 2 diabetes than models derived in a cohort including people with and without type 2 diabetes. Because the performance of risk models is strongly influenced by the background risk of a specific population to which it is applied, as well as other population characteristics, measured and unmeasured, it may be more practical to develop or recalibrate models specifically in the population of interest, rather than search for one equation that suits all populations.

Summary

- The PREDICT-1º T2D cohort has a more adverse risk profile compared with the general PREDICT-1º cohort.

- Comparative assessment of PREDICT-1º and PREDICT-1º T2D models in the cohort of primary care patients with diabetes showed that the later perform better on all measures of statistical performance and clinical utility.

- This study supports the hypothesis that diabetes-specific models, including additional predictors specifically relevant to diabetes, are superior to the general population models that include diabetes status along with other conventional CVD risk predictors.

Chapter 9. Conclusion

Risk prediction models play an important role in clinical practice by providing information about a patient's likelihood of future health-related outcomes. Therefore, clinical access to good quality prediction models is highly advantageous. The performance of new models is influenced by the methodological approaches to model development, and the similarity between the study populations used to develop the models and the populations to which they are applied.

This research focused on the development of new CVD risk prediction models using contemporary data from the large PREDICT primary care patient cohort in New Zealand, and on comparing the performance of new general population and diabetes population models. The approach taken drew on the methods used by established predictive modelling research groups and the methodological recommendations from the TRIPOD group (3, 21).

The first part of this thesis covered the rationale and background to this research, followed by a literature review of the range of methodological approaches applied by influential research groups in developing CVD risk prediction models (Chapter 2). In Chapter 3, the statistical concepts and methods of model construction and assessment of model performance were presented. Next, the PREDICT cohort study design was described, along with how the dataset was created, and a discussion of the strengths and limitations of the PREDICT data. Chapter 5 presented the specific methodological procedures used for deriving and assessing the performance of the new prediction models, and for the external validation and comparison with current US and NZ models. Chapter 6 described the development and evaluation of the general population models, and Chapter 7 presented the diabetes-specific models. Finally, these two sets of models were compared in Chapter 8.

Statement of principal findings; findings in relation to other studies

This research has demonstrated that it is possible to generate a very large CVD risk prediction cohort study in a routine primary care setting and develop prediction models that perform measurably better than currently available models. Both the general population and the diabetes-specific models were superior across all the standard performance indicators than models developed in cohorts recruited a number of years ago. This research also confirmed that the current New Zealand primary care population is mostly at low risk, and that predictive models developed from older cohorts now significantly overestimate CVD risk. Therefore,

models should be regularly updated to ensure optimal performance, since the effects of prognostic factors and the underlying risk in populations are likely to change with time, due to rapidly declining incidence of CVD and the changing distributions of risk factors in modern patient populations.

Another key finding was consistent with the idea that sub-population-specific models would generally perform better than generic models within the particular sub-population. This is likely to be because models based on a dataset derived by stratifying on a potentially effect-modifying characteristic allows for more accurate modelling of the effects of the traditional predictors, and also the addition of predictors specifically relevant to the sub-population of interest.

This research also showed that the addition of routinely available measures of socioeconomic status, self-identified ethnicity and several other regularly collected clinical measures, allows for the detection of high-risk patient groups whose risk would be underestimated otherwise. For example, in the PREDICT study population, Māori, Pacific and Indian patients with low socioeconomic status had estimated CVD risk twice as high as that of European or Chinese patients with high socioeconomic status.

The finding that CVD risk in the PREDICT cohort was considerably overestimated by the Northern American PCEs is consistent with other validation studies. A review of 15 external validation studies of the PCEs reported that observed risk was almost always overestimated (139). However, participants in these studies, including several randomised trials, were largely volunteers and the authors admitted the possibility of healthy-volunteer bias. As PREDICT participants were automatically recruited in routine practice (rather than individually consented) and represented contemporary primary care patients, the findings of this research provide the most conclusive evidence to date that the PCEs overestimate risk.

Most published CVD risk prediction equations include limited numbers of predictors, typically the traditional risk factors described in the Framingham cohort: sex, age, smoking, diabetes, blood pressure and blood lipids (135). Yet additional predictors exist that can improve risk stratification. Sociodemographic variables (self-identified ethnicity, socioeconomic status) or routinely-collected sub-population-specific biomarkers, such as measures of glycaemic control or renal health if a patient has diabetes, are examples of additional predictors. The recent UK QRISK3 equations (35) are the most comprehensive CVD prediction equations, including 22

variables; however, many of these may not be readily available during a CVD risk assessment in routine practice.

Strengths and limitations of the thesis

A major strength of this study is that it was large, contemporary and representative. The high level of data completeness, rare for a study of this size, along with sufficient accumulated follow-up time and a large number of observed events, make the data appropriate for development of CVD risk prediction models.

This study was strengthened by its methods of model development and validation, which are in line with the current best-practice recommendations (38).

This is the first study where the comparison of general population and diabetes-specific models for predicting CVD risk has been made in the same primary care population. This approach is advantageous when comparing models, since any differences in their performance cannot be attributed to variations between the datasets used, such as different inclusion criteria, definitions of outcomes and predictors, known and unknown measurement error, or the modelling approach. Hence the performance differences can be attributed to the predictors modelled with more certainty.

Additionally, using standard ICD codes to identify outcome events in this study is an advantage because this can facilitate international comparisons. A recent systematic review of CVD risk prediction models identified 363 models, most developed in Europe and North America (135). The authors reported substantial variation in outcome definitions and recommended use of more uniform definitions, preferably ICD-coded events, as was done in this research. While the accuracy of ICD coding for specific diagnoses can be unreliable, the broader CVD definition is likely to be more reliable and high sensitivities and positive predictive values have been reported for ICD-coded CVD events in national datasets (109).

This research also has a number of potential limitations. The dataset was derived largely from health data, not specifically collected for research purposes, and much of it was collected in a primary care practice setting, so there is a possibility of misclassification bias and missing information. However, the quality and completeness of data was improved by integrating the PREDICT risk assessment software into the electronic health record, using standard predictor definitions and built-in quality checks. Nevertheless a certain degree of misclassification and

measurement error in predictors and other patient characteristics relevant for these analyses cannot be ruled out. Furthermore, reliance on collecting data in routine practice meant that a balance had to be maintained between the primary care providers' workload and the amount of information collected. Thus, information on some relevant predictors, for example, diet and physical activity, was not available in the PREDICT cohort. Another limitation was that the outcome events were not individually adjudicated, which was not possible given the use of anonymised data.

A potential weakness of developing CVD risk prediction equations in contemporary cohorts like PREDICT, is the likelihood that some participants will initiate new CVD preventive treatment during follow-up. However, in part due to the relatively short follow-up time and due to the relatively high level of treatment at baseline, the net proportion of person-time in the PREDICT cohort spent on medications initiated after the baseline assessment was small (only about 12% of person-time during follow-up - Appendix 9.1). Therefore it is unlikely that new treatment explains more than a small proportion of the over-prediction of risk by the older equations derived from largely untreated cohorts, as was suggested previously (183). Nevertheless, the high level of preventive medication use (approximately one-third of the PREDICT cohort were on treatment at baseline) is one of the reasons for the low average risk in the cohort. To account for this, baseline medications were included as variables in the equations. Several other equations (12, 32, 35) have included baseline blood pressure-lowering treatment, and so, for completeness, lipid-lowering and antithrombotic medications were included in the general population PREDICT-1^{o} equations, and insulin and oral antiglycaemic medications were included in the PREDICT-1^{o} T2D equations. As preventive treatment is seldom optimal, patients who remain at high predicted risk despite treatment (often monotherapy) will be candidates for additional interventions. Therefore, treated participants should always be included in risk prediction cohorts. There is currently no methodological consensus as to how new treatment initiated during follow-up should be addressed, and no CVD prediction studies have adequately adjusted for new treatment.

Implications of the thesis

The new equations presented in this thesis were developed to support clinical decision-making. They utilise information that is already routinely collected in primary care. Integrated

electronic calculators are now widely used in NZ primary care. These automatically populate many variables from patients' records, so implementing the new equations will require minimal additional effort from health professionals.

Until this year, the national recommendation for CVD risk assessment in New Zealand primary care has been to apply the New Zealand-adjusted Framingham equation in the general population without prior CVD (9, 10). The New Zealand Diabetes Cohort Study equation has been available for risk assessment in patients with diabetes (28), although the diabetes equation has not been widely used because it was never integrated into primary care patient management systems. The equations presented here were designed to replace both these sets of equations and the 2018 New Zealand guidelines for Cardiovascular Risk Assessment and Management in primary care have already recommended using the new PREDICT-1^{o} general population and PREDICT-1^{o} T2D equations, and implementation plans are under way for integration into primary care patient management systems (184).

As the underlying CVD risk in populations changes over time, the role of predictors may also change, so it will be important to update the prediction models. Because PREDICT is an ongoing open cohort, with regular linkages to update outcome and predictor information, it will be able to provide contemporary data for such regular updates.

As the PREDICT population grows larger and the duration of follow up grows, opportunities will emerge for developing other models tailored to sub-populations, such as ethnicity-specific models, models for older and/or younger age groups, etc.

Opportunities for future research

Comparisons of the performance of sub-population-specific models with general population models is a high priority for future research. The priority sub-populations will be those known to be at increased CVD risk, including Māori, Pacific and South Asian populations, socioeconomically deprived populations, older age groups, those with abnormal biomarkers, and groups with comorbidities.

The investigation of whether new predictors significantly improve model performance is also a priority for future research. As discussed above, these research priorities will also require an international discussion on new ways to assess improvements in the performance of models that go beyond the relatively insensitive statistical metrics currently used. The standard

performance metrics (e.g. concordance statistics) are global measures which are relatively insensitive to the addition of new variables that may have clinically relevant predictive effects for sub-populations, as assessed by the hazard ratio of a new predictor and its prevalence (185, 186).

The impact of new or changing treatments during follow-up on risk models, needs to be investigated, particularly in studies with extended follow-up. The current PREDICT cohort has a mean and median follow-up of less than 5 years, but as the follow-up time increases, the impact of new and changing treatment will have an increasing impact on CVD outcomes and therefore on the risk models. Approaches including updated covariates (treatment variables and the biomarkers which may change as a result of treatment) have been proposed, but to date, there do not appear to be any published CVD risk models that have adequately accounted for new treatment initiated after baseline assessments.

As routinely-collected health data become more available internationally, they will be increasingly used in the development of prediction models. A substantial range of information can be accumulated in the process of health care, and these datasets are bound to be very large, yet more data does not necessarily provide more meaningful/relevant information and data quality issues are difficult to assess in large datasets. Furthermore, as datasets increase in size, conventional statistical methods become less suitable for assessing model performance and the high precision of estimates and the unrealistically small standard errors might not produce realistic measures of uncertainty. These large datasets with repeated measures and frequently updated or real-time information may be suitable for data-hungry analytical approaches, such as machine learning (187, 188) or dynamic prediction modelling (189, 190). However, it is unclear whether such approaches will outperform the conventional regression modelling in terms of meaningful predictive ability. An assessment of the performance of models developed using these new analytical approaches is another priority for research using datasets like the PREDICT cohort.

As pointed out by the authors of the TRIPOD statement, methodological issues in prognostic and prediction modelling are constantly evolving (21). Therefore it will be of benefit to revise and update the prognostic modelling approaches, in order to incorporate the new evidence and address the weaknesses of current approaches.

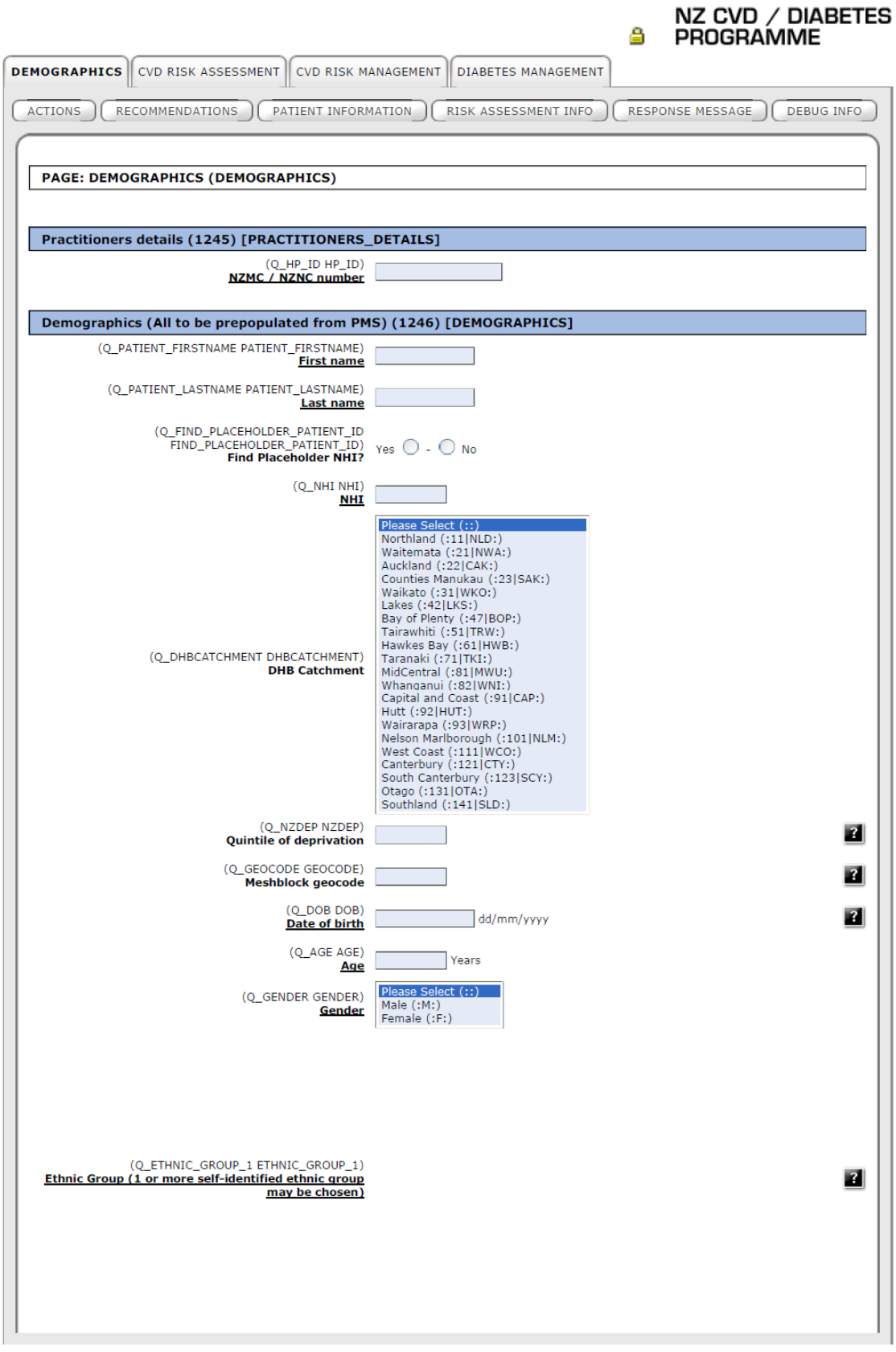
Multiple analytical options are available when developing new prediction models and assessing their performance. In CVD prediction modelling, it has not been common practice for published reports to give a complete account of the process of model development, including the rationale for the choices made. This thesis is one of the first studies to report the whole process of model development and assessment, to an extent that it can be replicated by others. Furthermore, few CVD prediction studies have assessed new predictors by fitting models with the prognostic index from an existing model along with the new predictors, as was done in this thesis. This approach is useful to assess new predictors' ability to identify cases at high risk, over and above the known risk. Using this approach, the size of the effect estimates (such as hazard ratios), prevalence and distributions of new predictors, as well as the cost of their measurement, may be considered when deciding whether to add them during development of new models. In contrast, the addition of new predictors does not usually lead to a noticeable improvement in model's performance as measured by standard indices (such as the C-statistic).

Conclusion

This research project has demonstrated that large-scale high-quality risk prediction research can be undertaken effectively and efficiently using linked routinely-collected computerised health data. The New Zealand experience with the PREDICT research programme provides a model for a wide range of future research opportunities. While health data are increasingly collected electronically, the most important requirement for successful research in this field is the availability of a unique personal identifier on all relevant records to enable accurate linkage of individual patient information. By using robust encryption methodology, patient confidentiality can be maintained, while enabling multiple individual records to be linked. Many countries have versions of personal health identifiers but have not allowed them to be used for health research. This project has demonstrated some of the benefits of allowing researchers to use linked health data on a large scale.

APPENDICES

Appendix 4.1 PREDICT templates: demographics, CVD risk assessment, CVD risk management, and diabetes management



Not Stated (::)
New Zealand European (:11:)
Other European (:12:)
New Zealand Maori (:21:)
Samoan (:31:)
Cook Island Maori (:32:)
Tongan (:33:)
Niuean (:34:)
Tokelauan (:35:)
Fijian (:36:)
Other Pacific Islands (not listed) (:37:)
Pacific Island not further defined (:30:)
Indian (:43:)
Sri Lankan (:441:)
Pakistani (:44414:)
Bangladeshi (:44412:)
Afghani (:44411:)
Nepalese (:44413:)
Tibetan (:44415:)
Chinese (:42:)
Japanese (:442:)
Korean (:443:)
Southeast Asian (:41:)
Other Asian (Code 44) (:44:)
Other Asian (Code 444) (:444:)
Asian not further defined (:40:)
Middle Eastern (:51:)
Latin American / Hispanic (:52:)
African (:53:)
Other (:54:)
European Not Further Defined (:10:)

(Q_ETHNIC_GROUP_2 ETHNIC_GROUP_2)
Ethnic Group 2

Not Stated (:99:)
New Zealand European (:11:)
Other European (:12:)
New Zealand Maori (:21:)
Samoan (:31:)
Cook Island Maori (:32:)
Tongan (:33:)
Niuean (:34:)
Tokelauan (:35:)
Fijian (:36:)
Other Pacific Islands (not listed) (:37:)
Pacific Island not further defined (:30:)
Indian (:43:)
Sri Lankan (:441:)
Pakistani (:44414:)
Bangladeshi (:44412:)
Afghani (:44411:)
Nepalese (:44413:)
Tibetan (:44415:)
Chinese (:42:)
Japanese (:442:)
Korean (:443:)
Southeast Asian (:41:)
Other Asian (Code 44) (:44:)
Other Asian (Code 444) (:444:)
Asian not further defined (:40:)
Middle Eastern (:51:)
Latin American / Hispanic (:52:)
African (:53:)
Other (:54:)
European Not Further Defined (:10:)

(Q_ETHNIC_GROUP_3 ETHNIC_GROUP_3)
Ethnic Group 3

Not Stated (:99:)
New Zealand European (:11:)
Other European (:12:)
New Zealand Maori (:21:)
Samoan (:31:)
Cook Island Maori (:32:)
Tongan (:33:)
Niuean (:34:)
Tokelauan (:35:)
Fijian (:36:)
Other Pacific Islands (not listed) (:37:)
Pacific Island not further defined (:30:)
Indian (:43:)
Sri Lankan (:441:)
Pakistani (:44414:)
Bangladeshi (:44412:)
Afghani (:44411:)
Nepalese (:44413:)
Tibetan (:44415:)
Chinese (:42:)
Japanese (:442:)
Korean (:443:)
Southeast Asian (:41:)
Other Asian (Code 44) (:44:)
Other Asian (Code 444) (:444:)
Asian not further defined (:40:)
Middle Eastern (:51:)
Latin American / Hispanic (:52:)
African (:53:)
Other (:54:)
European Not Further Defined (:10:)

NEXT ...

PAGE: CVD RISK ASSESSMENT (CVD_RISK_ASSESSMENT)

This page should be completed for all patients. All underlined items are required.

After submitting this form, additional follow up management forms become available to you. The secondary Diabetes management form will become available dependant upon the status of the Diabetes field on this form.

NOTE: It is inappropriate to do CVD risk assessment in pregnancy.

ASSUME NEGATIVE DEFAULTS

Clinical History (1248) [CLINICAL_HISTORY]

(Q_FAMILYHISTORY FAMILYHISTORY)
Family History of Premature CVD Yes - No

(Q_IHD IHD)
Angina/MI Yes - No

(Q_ANGINA ANGINA)
Angina Yes - No

(Q_MI MI)
MI Yes - No

(Q_PTCA_CABG PTCA_CABG)
PCI/CABG Yes - No

(Q_STROKE_TIA STROKE_TIA)
Ischaemic Stroke or Transient Ischaemic Attack (TIA) Yes - No

(Q_STROKE STROKE)
Ischaemic Stroke Yes - No

(Q_TIA TIA)
Transient Ischaemic Attack (TIA) Yes - No

(Q_PVD PVD)
PVD Yes - No

(Q_DIABETES DIABETES)
Diabetes
- Please select (::)
- None (:0:)
- Type 1 (:1:)
- Type 2 (incl Type 2 on insulin) (:2:)
- Type unknown (:3:)
- Current gestational diabetes (:4:)

(Q_ATRIAL_FIBRILLATION ATRIAL_FIBRILLATION)
ECG confirmed Atrial Fibrillation Yes - No

(Q_GEN_LIPID GEN_LIPID)
Diagnosed Genetic Lipid Disorder
- Please select (::)
- None (:0:)
- Familial hypercholesterolaemia (:1:)
- Familial defective apoB (:2:)
- Familial combined dyslipidaemia (:3:)
- Other genetic lipid disorder (:4:)

(Q_METABOLIC_SYNDROME METABOLIC_SYNDROME)
Diagnosed metabolic syndrome Yes - No

(Q_SMOKING SMOKING)
Smoking History
- Please select (::)
- No - never (:0:)
- No - quit over 12 months ago (:1:)
- No - recently quit (within 12 months) (:2:)
- Yes - up to 10 / day (:3:)
- Yes - 11 - 19 / day (:4:)
- Yes - 20+ / day (:5:)

(Q_PREGNANT PREGNANT)
Pregnant? Yes - No

Examination (1249) [RA_EXAMINATION]

(Q_BPS BPS)
Most recent BP (Sitting) / mmHg

(Q_BPS2 BPS2)
Previous BP (Sitting) / mmHg

(Q_TCHDL_RATIO TCHDL_RATIO)
TC/HDL ratio - Date: dd/mm/yyyy

(Q_TCL TCL)
Total Cholesterol mmol/L - Date: dd/mm/yyyy

Diabetes Screening (2113) [DM_SCREENING]

(Q_RA_GLUCOSE RA_GLUCOSE)
Fasting glucose (for diabetes screening) mmol/L - Date: dd/mm/yyyy

(Q_RA_HBA1C RA_HBA1C)
HbA1c (for diabetes screening) % - Date: dd/mm/yyyy

For diabetic patient (1250) [FOR_DIABETIC_PATIENT]

(Q_DIABETES_YR DIABETES_YR)
Diabetes: year of diagnosis

(Q_RENAL RENAL)
Renal disease

- Please select (::)
- No nephropathy (:0:)
- Confirmed microalbuminuria (:1:)
- Overt diabetic nephropathy (:2:)
- Non-diabetic nephropathy (:3:)

(Q_HBA1C HBA1C)
HbA1c % - Date: dd/mm/yyyy

(Q_DataReal_1 DataReal_1)
This data is the patient's real clinical information Yes - No

SUBMIT RISK ASSESSMENT Or PARK ONLY

'WHAT IF' / DEMONSTRATION STYLE RISK ASSESSMENT

PAGE: CVD RISK MANAGEMENT (CVD_RISK_MANAGEMENT)

Note the BMI calculator on this page calculates the BMI value automatically from height and weight. All underlined items are required.

Examination (1252) [CVD_EXAMINATION]

(Q_HEIGHT HEIGHT)
Height cm

(Q_WEIGHT WEIGHT)
Weight kg - Date: dd/mm/yyyy

(Q_BMI BMI)
BMI (Auto-calculated) kg/m²

(Q_WAIST WAIST)
Waist circumference cm

CVD medications (1253) [CVD_MEDICATIONS]

CAUTION: Please note that all medications default to "No". Please review carefully before proceeding.

UPDATE CVD MEDICATIONS FROM MEDTECH...

(Q_ASPIRIN ASPIRIN)
Aspirin

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)
- Don't know (:3:)

(Q_CLOPIDOGREL CLOPIDOGREL)
Clopidogrel

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_WARFARIN WARFARIN)
Warfarin

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_ACE_INHIBITOR ACE_INHIBITOR)
ACE Inhibitor

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_AT2 AT2)
Angiotensin II Receptor Blocker

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_BETA_BLOCKER BETA_BLOCKER)
Beta Blocker

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_THIAZIDE THIAZIDE)
Thiazide

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_CALCIUM_ANTAGONIST CALCIUM_ANTAGONIST)
Calcium Antagonist

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_OTHER_HYP_DRUGS OTHER_HYP_DRUGS)
Other drug therapy for Hypertension

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_STATIN STATIN)
Statin

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_FIBRATE FIBRATE)
Fibrate

- No (:0:)
- Contraindicated / Not tolerated (:1:)
- Yes (:2:)

(Q_OTHER_LIPID_DRUGS OTHER_LIPID_DRUGS)
Other Lipid lowering drugs

No (:0:)
Contraindicated / Not tolerated (:1:)
Yes (:2:)

Investigation (1254) [INVESTIGATION]

(Q_GLUCOSE GLUCOSE)
Fasting glucose (for diabetes screening) mmol/L - Date: dd/mm/yyyy

(Q_CVD_HBA1C CVD_HBA1C)
HbA1c (for diabetes screening) % - Date: dd/mm/yyyy

(Q_LDL LDL)
LDL Cholesterol (fasting) mmol/L - Date: dd/mm/yyyy

(Q_TRI TRI)
Triglyceride (fasting) mmol/L - Date: dd/mm/yyyy

(Q_HDL HDL)
HDL Cholesterol mmol/L - Date: dd/mm/yyyy

Lifestyle Management (1255) [LIFESTYLE_MANAGEMENT]

(Q_SMK_QUIT SMK_QUIT)
Smoke Quit Advice given today? Yes - No

(Q_PHY_ACTIVE PHY_ACTIVE)
Physically active? Yes - No

(Q_GREEN_PRES GREEN_PRES)
Green Prescription given Yes - No

(Q_LAST_DIET_CHECK LAST_DIET_CHECK)
Date of last dietary assessment dd/mm/yyyy

(Q_REFERRAL_DIET_GIVEN REFERRAL_DIET_GIVEN)
Date referral for dietary advice dd/mm/yyyy

(Q_Diab_nurse_edu_provided Diab_nurse_edu_provided)
Nurse Education Provided Yes - No

(Q_DataReal_2 DataReal_2)
This data is the patient's real clinical information Yes - No

NEXT ...

RUN CVD MANAGEMENT Or PARK ONLY

'WHAT IF' / DEMONSTRATION CVD MANAGEMENT

PAGE: DIABETES MANAGEMENT (DIABETES_MANAGEMENT)

All underlined items are required.

Get Checked (2062) [DIABETIC_GETCHECKED_SH]

(Q_DIABETES_GETCHECKED DIABETES_GETCHECKED)
Is this a Get Checked annual review? Yes - No

Diabetes glycaemic control (1257) [DIABETES_GLYCAEMIC_CONTROL]

CAUTION: Please note that all medication-related questions in this section default to "No". Please review carefully before proceeding.

UPDATE DM MEDICATIONS FROM MEDTECH...

(Q_DIAB_HBA1C DIAB_HBA1C)
HbA1c % - Date: dd/mm/yyyy

(Q_DIAB_DIETONLY DIAB_DIETONLY)
Diet therapy only

No (:0:)
Yes (:1:)

(Q_DIAB_METFORMIN DIAB_METFORMIN)
Metformin

No (:0:)
Contraindicated / Not tolerated (:1:)
Yes (:3:)
On maximum tolerated dose (:2:)

(Q_DIAB_SULPHONYLUREA DIAB_SULPHONYLUREA)
Sulphonylurea

No (:0:)
Contraindicated / Not tolerated (:1:)
Yes (:3:)
On maximum tolerated dose (:2:)

(Q_DIAB_GLITAZONE DIAB_GLITAZONE)
Glitazone

No (:0:)
Contraindicated / Not tolerated (:1:)
Yes (:3:)
On maximum tolerated dose (:2:)

(Q_DIAB_ACARBOSE DIAB_ACARBOSE)
Acarbose

No (:0:)
Contraindicated / Not tolerated (:1:)
Yes (:3:)
On maximum tolerated dose (:2:)

(Q_DIAB_INSULIN DIAB_INSULIN)
Insulin

No (:0:)
Nocturnal only (:1:)
Once daily (:2:)
Twice daily (:3:)
Multiple injections/insulin pump (:4:)

(Q_DIAB_HYPO_ATTACKS DIAB_HYPO_ATTACKS)
Hypoglycaemic attacks

No (:0:)
Less than 1 per month (:1:)
Less than 1 per week (:2:)
More than 1 per week (:3:)

(Q_DIAB_LAST_DIET_ASSESS DIAB_LAST_DIET_ASSESS)
Date of last dietary assessment

dd/mm/yyyy

(Q_DIAB_DIET_REFERRAL DIAB_DIET_REFERRAL)
Date referral for dietary advice

dd/mm/yyyy

(Q_DIAB_EDU_REFERRAL DIAB_EDU_REFERRAL)
Date referral for diabetic education

dd/mm/yyyy

Renal (1258) [RENAL]

(Q_DIAB_ACR DIAB_ACR)
ACR

mg/mmol - Date: dd/mm/yyyy

(Q_SERUM_CREATININE SERUM_CREATININE)
Serum creatinine

??/l - Date: dd/mm/yyyy

(Q_DIAB_GFR DIAB_GFR)
Estimated GFR

ml/min/1.73 m2

Diabetic Feet (required for GetChecked) (1259) [DIABETIC_FEET_HEADER]

(Q_RUNDIAB_FEET RUNDIAB_FEET)
Do you want to complete the foot section?

No (:0:)
Yes (:1:)

(Q_DIAB_FEET_DATE_LAST_CHECK DIAB_FEET_DATE_LAST_CHECK)
Date of last foot examination

dd/mm/yyyy

(Q_DIAB_FEET_ULCER_HISTORY DIAB_FEET_ULCER_HISTORY)
History diabetic ulcer

Yes - No

(Q_DIAB_FEET_ULCER_CURRENT DIAB_FEET_ULCER_CURRENT)
Current diabetic ulcer

Yes - No

(Q_DIAB_FEET_HIGHRISK DIAB_FEET_HIGHRISK)
Other criteria for 'high-risk' foot

Yes - No

(Q_DIAB_FEET_PREV_LOWLIMB_AMP DIAB_FEET_PREV_LOWLIMB_AMP)
Previous diabetic lower limb amputation

Please select (::)
No (:0:)
Yes - Left (:1:)
Yes - Right (:2:)
Yes - Bilateral (:3:)

(Q_DIAB_FEET_SENSATION DIAB_FEET_SENSATION)
Foot - Sensation

Please select (::)
Not Examined (:0:)
Normal (:1:)
Abnormal (Left) (:2:)
Abnormal (Right) (:3:)
Abnormal (BOTH) (:4:)

(Q_DIAB_FEET_CIRCULATION DIAB_FEET_CIRCULATION)
Foot - Circulation

Please select (::)
Not Examined (:0:)
Normal (:1:)
Abnormal (Left) (:2:)
Abnormal (Right) (:3:)
Abnormal (BOTH) (:4:)

Diabetic Eyes (required for GetChecked) (1261) [DIABETIC_EYES_HEADER]

(Q_BLIND BLIND)
Blind in both eyes?

Yes - No

(Q_RUNDIAB_EYES RUNDIAB_EYES)
Do you want to complete the eye section?

No (:0:)
Yes (:1:)

(Q_DIAB_EYE_LASTRET DIAB_EYE_LASTRET)
Date of last retinal review

dd/mm/yyyy

(Q_DIAB_EYE_RETINOPATHY DIAB_EYE_RETINOPATHY)
Retinopathy worst eye

Please Select (::)
No retinopathy / no changes (:0:)
Non-proliferative (:1:)
Proliferative (:2:)
Macular oedema (:3:)
Not checked (:9:)

(Q_DIAB_VIS_ACUITY_LEFT DIAB_VIS_ACUITY_LEFT)
Corrected visual acuity (x/x)

(L) (R)

(Q_DIAB_RETINAL_REFERRAL DIAB_RETINAL_REFERRAL)
Eye referral today?

- Please Select (::)
- No (:0:)
- No - in screening programme (:1:)
- No - under care of Ophthalmologist (:2:)
- Yes - to retinal screening programme (:3:)
- Yes - to ophthalmologist (:4:)

(Q_DataReal_3 DataReal_3)
This data is the patient's real clinical information Yes - No

RUN DIABETES MANAGEMENT Or PARK ONLY

'WHAT IF' / DEMONSTRATION DIABETES MANAGEMENT

PAGE: Actions (Actions)

PAGE: Recommendations (Recommendations)

PAGE: Patient Information (Patient_Information)

PAGE: Risk Assessment Info (Risk_Assessment_Info)

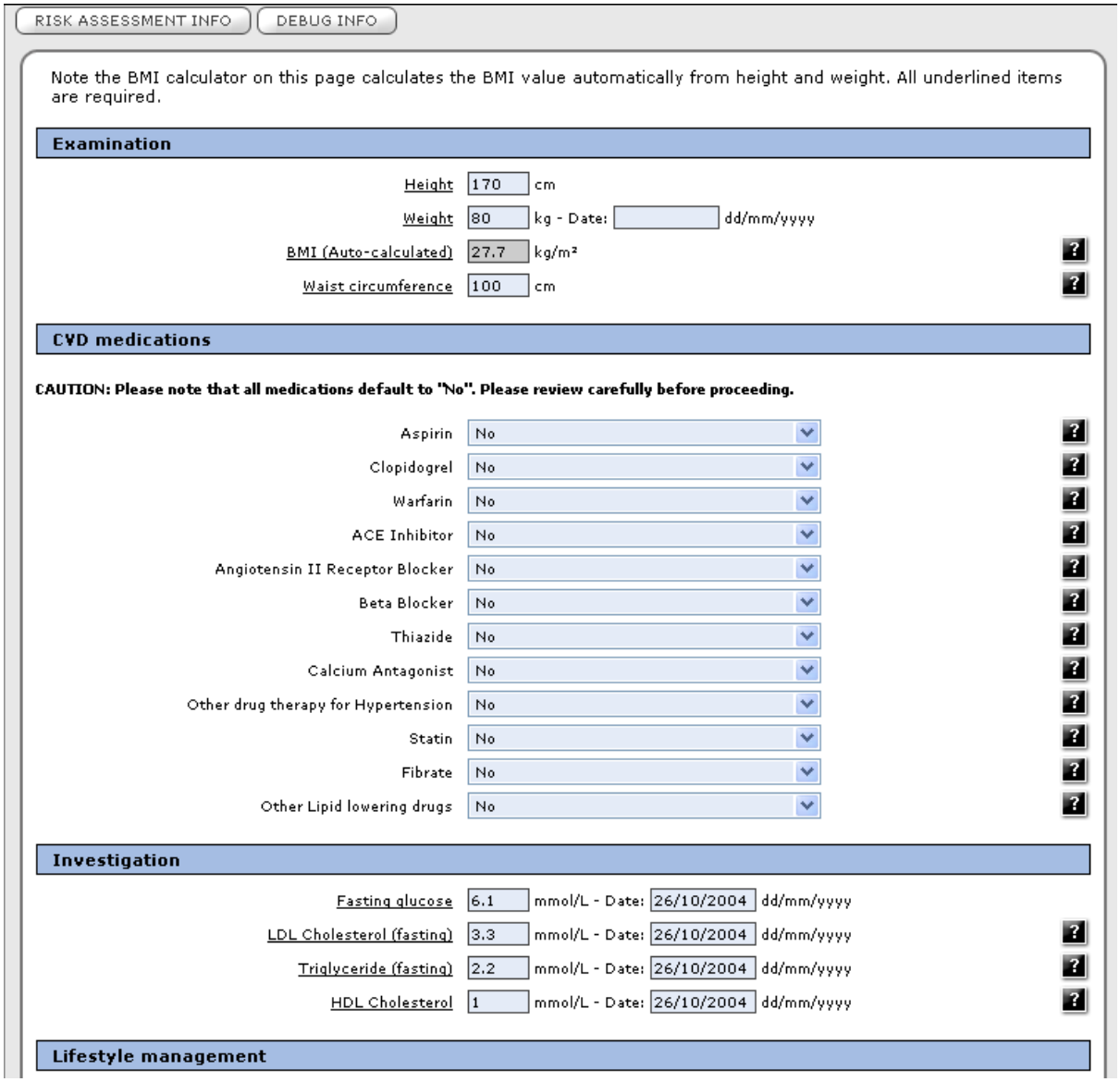
PAGE: Response Message (Response_Message)

PAGE: Debug Info (Debug_Info)

Well Dunedin PRIMARY HEALTH ORGANISATION PREDICT Medical ENIGMA

Logout

Appendix 4.2. Example of a completed PREDICT management template



RISK ASSESSMENT INFO | DEBUG INFO

Note the BMI calculator on this page calculates the BMI value automatically from height and weight. All underlined items are required.

Examination

Height	170 cm	
Weight	80 kg - Date: dd/mm/yyyy	
BMI (Auto-calculated)	27.7 kg/m²	?
Waist circumference	100 cm	?

CVD medications

CAUTION: Please note that all medications default to "No". Please review carefully before proceeding.

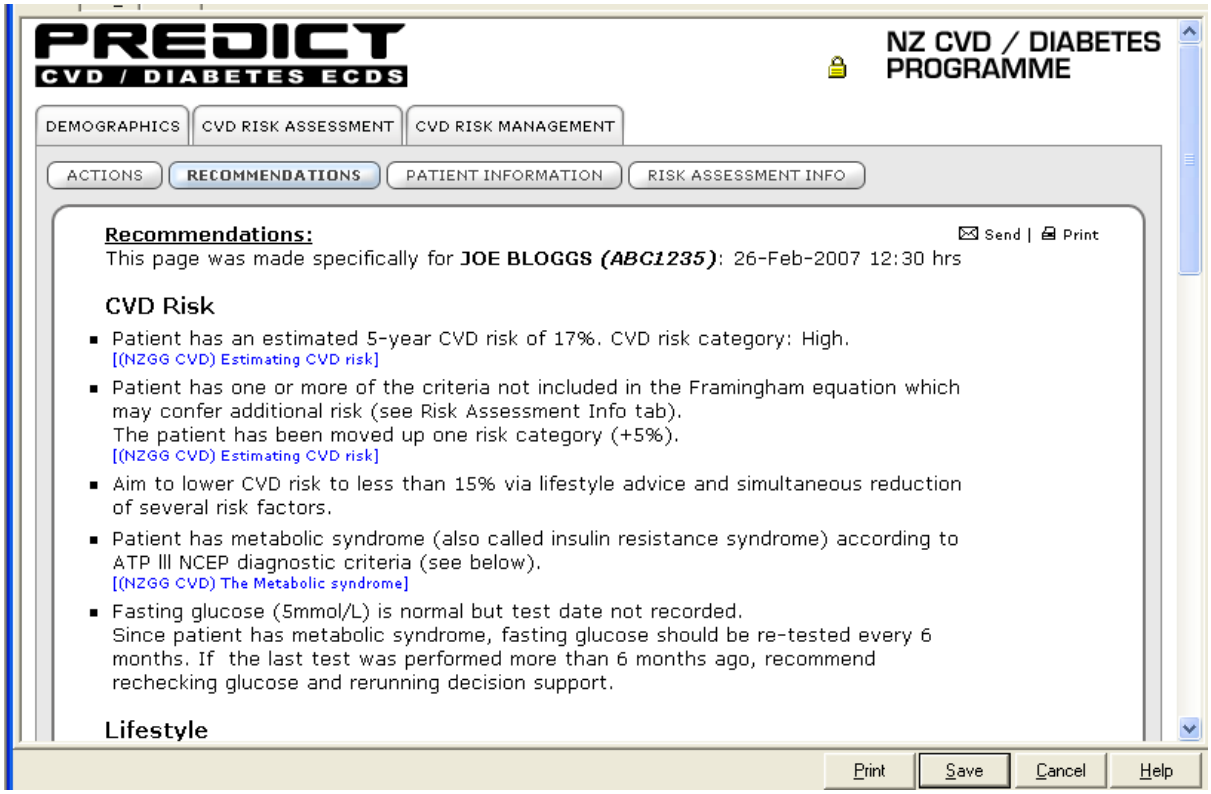
Aspirin	No	?
Clopidogrel	No	?
Warfarin	No	?
ACE Inhibitor	No	?
Angiotensin II Receptor Blocker	No	?
Beta Blocker	No	?
Thiazide	No	?
Calcium Antagonist	No	?
Other drug therapy for Hypertension	No	?
Statin	No	?
Fibrate	No	?
Other Lipid lowering drugs	No	?

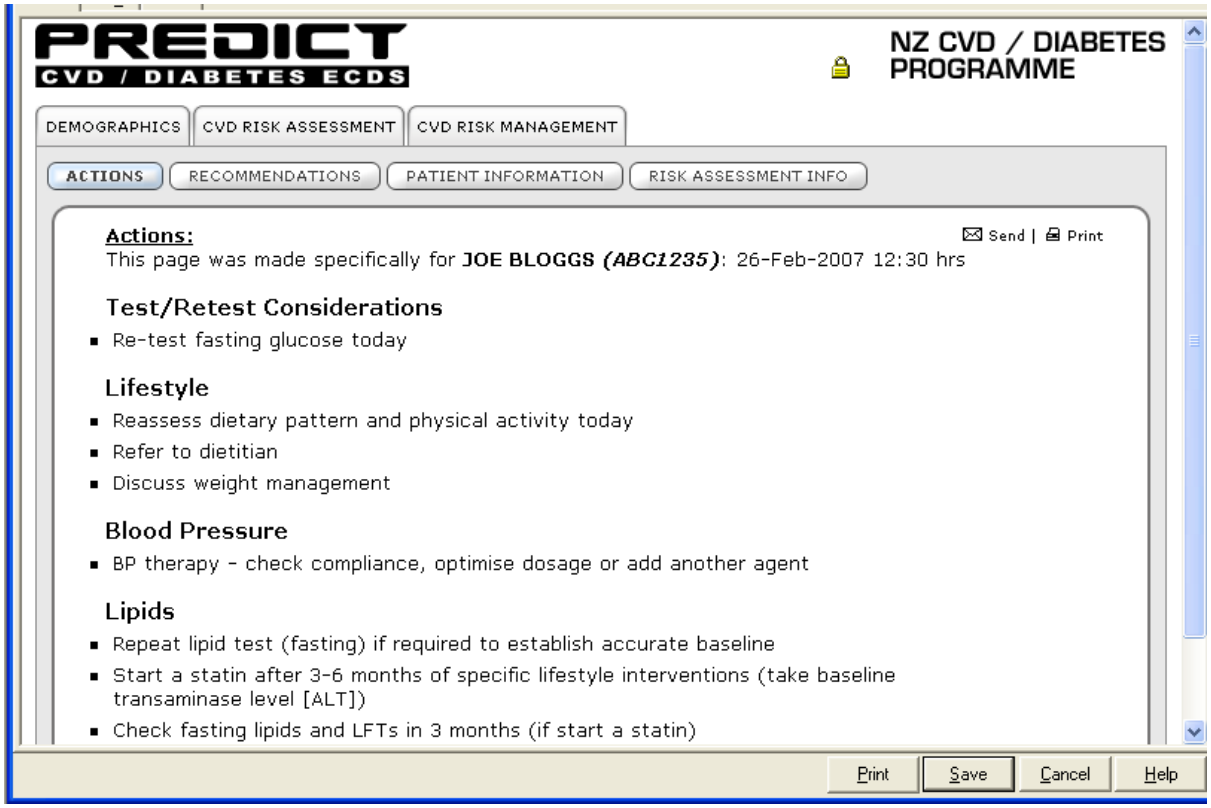
Investigation

Fasting glucose	6.1 mmol/L - Date: 26/10/2004 dd/mm/yyyy	
LDL Cholesterol (fasting)	3.3 mmol/L - Date: 26/10/2004 dd/mm/yyyy	?
Triglyceride (fasting)	2.2 mmol/L - Date: 26/10/2004 dd/mm/yyyy	?
HDL Cholesterol	1 mmol/L - Date: 26/10/2004 dd/mm/yyyy	?

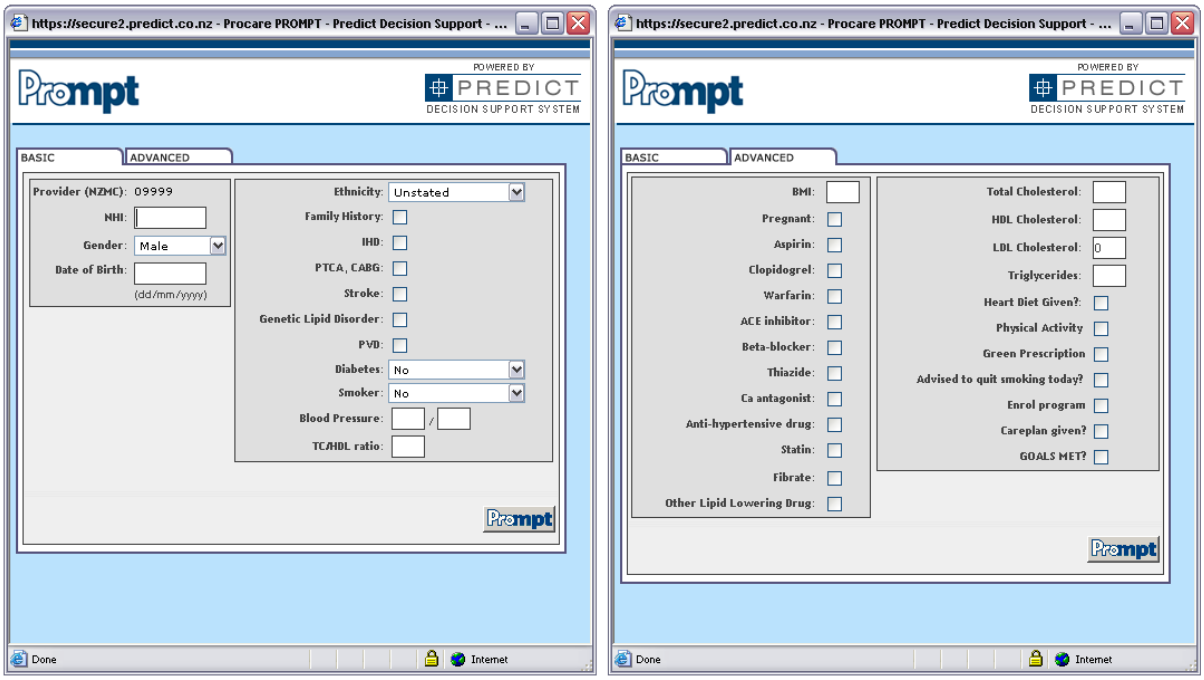
Lifestyle management

Appendix 4.3. Example of a response to a completed PREDICT template, with patient's absolute risk calculated





Appendix 4.4. PREDICT template, old version



Appendix 7.1. Adjusted hazard ratios for first cardiovascular event, women aged 30-74 years

	Adjusted hazards ratios (95% CI)	
	Women, complete case analysis	**Women, variables with missing values categorised***
Current age, y	1.039 (1.031, 1.047)	1.047 (1.039, 1.054)
Years since diagnosis of T2D, y	1.014 (1.004, 1.024)	N/A
<=1	N/A	1
2-5	N/A	1.057 (0.916, 1.218)
>5	N/A	1.253 (1.085, 1.446)
Unknown	N/A	0.947 (0.733, 1.223)
Ethnicity:		
European/other	1	
Māori	1.016 (0.850, 1.216)	1.098 (0.938, 1.285)
Pacific	0.761(0.642, 0.901)	0.827 (0.712, 0.961)
Indian	1.204 (1.003, 1.445)	1.218 (1.029, 1.442)
Chinese	0.702 (0.496, 0.995)	0.796 (0.587, 1.081)
NZ Deprivation quintile (per 1 quintile)	1.067 (1.017, 1.119)	1.071 (1.027, 1.117)
Smoking:		
Non-smoker	1	
Current smoker	1.630 (1.398, 1.901)	1.685 (1.470, 1.931)
Family history of premature CVD	0.991 (0.834, 1.177)	1.020 (0.881, 1.182)
History of atrial fibrillation	2.375 (1.731, 3.258)	2.291 (1.740, 3.017)
SBP, mmHg	1.011 (1.007, 1.014)	1.011 (1.008, 1.014)
TC:HDL	1.130 (1.087, 1.175)	1.121 (1.080, 1.164)
eGFR, mL/min/1.73 m^2	0.991 (0.988, 0.994)	N/A
>=90	N/A	1
60-89	N/A	1.094 (0.968, 1.237)
45-59	N/A	1.308 (1.085, 1.577)
30-44	N/A	1.682 (1.318, 2.146)
Unknown	N/A	0.796 (0.516, 1.229)
ACR, mg/g	1.210 (1.165, 1.248)	N/A
<3	N/A	1
3-30	N/A	1.348 (1.186, 1.532)
>30	N/A	2.440 (2.056, 2.895)
Unknown	N/A	1.127 (0.958, 1.324)
HbA_{1c}, mmol/mol	1.008 (1.006, 1.011)	1.008 (1.006, 1.011)
BMI, kg/m^2	1.009 (1.001, 1.017)	1.010 (1.003, 1.018)
Treatment status at index assessment:		
On oral hypoglycaemic medications	0.994 (0.866, 1.141)	0.983 (0.863, 1.119)
On insulin	1.300 (1.095, 1.543)	1.319 (1.118, 1.554)
On blood pressure lowering medications	1.193 (1.027, 1.386)	1.217 (1.067, 1.389)
On lipid lowering medications	0.862 (0.756, 0.984)	0.873 (0.772, 0.987)
On antithrombotic medications	1.036 (0.917, 1.171)	1.055 (0.941, 1.182)

*Categories defined as follows: ACR: <3, 3-30, >30, unknown; eGFR: >=90, 60-89, 45-59, 30-44, unknown; years since diagnosis: <=1, 2-5, >5, unknown.

Appendix 7.2. Adjusted hazard ratios for first cardiovascular event, men aged 30-74 years

	Adjusted hazards ratios (95% CI)	
	Men, complete case analysis	**Men, variables with missing values categorised***
Current age, y	1.042 (1.036, 1.049)	1.045 (1.040, 1.051)
Years since diagnosis of T2D, y	1.012 (1.004, 1.021)	N/A
<=1	N/A	1
2-5	N/A	1.092 (0.982, 1.215)
>5	N/A	1.282 (1.148, 1.432)
Unknown	N/A	1.067 (0.899, 1.266)
Ethnicity:		
European/other	1	1
Māori	1.049 (0.906, 1.214)	1.092 (0.966, 1.235)
Pacific	0.894 (0.779, 1.027)	0.935 (0.831, 1.052)
Indian	1.377 (1.199, 1.583)	1.322 (1.167, 1.496)
Chinese	0.728 (0.562, 0.942)	0.707 (0.560, 0.891)
NZ Deprivation quintile (per 1 quintile)	1.060 (1.022, 1.100)	1.063 (1.030, 1.097)
Smoking:		
Non-smoker	1	1
Current smoker	1.294 (1.152, 1.452)	1.394 (1.265, 1.536)
Family history of premature CVD	1.186 (1.025, 1.374)	1.060 (0.938, 1.198)
History of atrial fibrillation	1.660 (1.314, 2.097)	1.503 (1.208, 1.870)
SBP, mmHg	1.007 (1.004, 1.010)	1.008 (1.005, 1.010)
TC:HDL	1.074 (1.043, 1.106)	1.089 (1.063, 1.117)
eGFR, mL/min/1.73 m^2	0.995 (0.993, 0.998)	N/A
>=90	N/A	1
60-89	N/A	0.923 (0.843, 1.010)
45-59	N/A	1.107 (0.948, 1.293)
30-44	N/A	1.464 (1.158 1.849)
Unknown	N/A	0.548 (0.398, 0.755)
ACR, mg/g	1.168 (1.136, 1.201)	N/A
<3	N/A	1
3-30	N/A	1.365 (1.233, 1.512)
>30	N/A	1.951 (1.700, 2.239)
Unknown	N/A	1.298 (1.155, 1.458)
HbA_{1c}, mmol/mol	1.007 (1.005, 1.009)	1.007 (1.005, 1.009)
BMI, kg/m^2	1.010 (1.003, 1.018)	1.012 (1.005, 1.018)
Treatment status at index assessment:		
On oral hypoglycaemic medications	0.882 (0.791, 0.983)	0.869 (0.786, 0.960)
On insulin	1.276 (1.100, 1.481)	1.299 (1.129, 1.495)
On blood pressure lowering medications	1.130 (1.008, 1.267)	1.136 (1.028, 1.254)
On lipid lowering medications	0.977 (0.878, 1.088)	0.960 (0.872, 1.057)
On antithrombotic medications	0.999 (0.904, 1.103)	1.029 (0.940, 1.126)

*Categories defined as follows: ACR: <3, 3-30, >30, unknown; eGFR: >=90, 60-89, 45-59, 30-44, unknown; years since diagnosis: <=1, 2-5, >5, unknown.

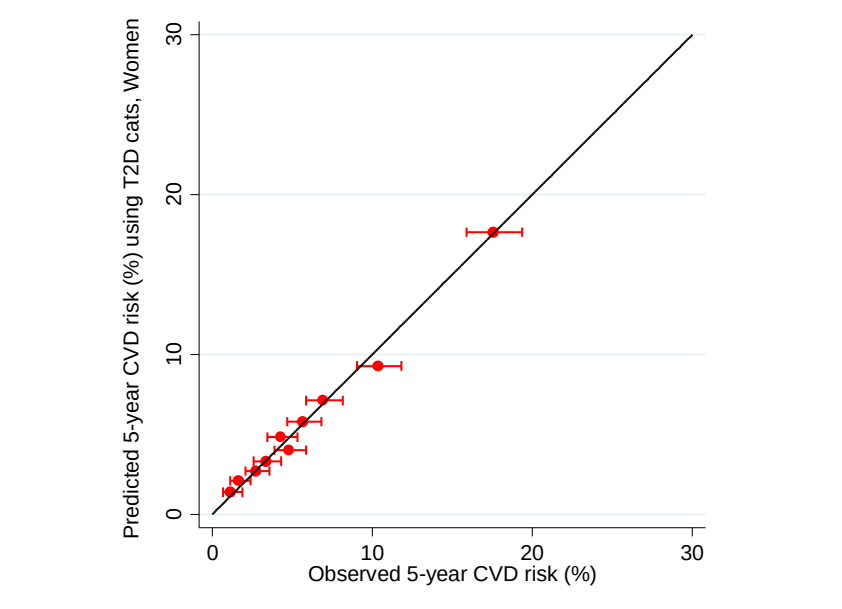
Appendix 7.3. Performance statistics for the T2D models with continuous and categorised eGFR, ACR, and years since T2D diagnosis, by sex

	PREDICT-1° T2D – continuous variables	**PREDICT-1° T2D – categorical variables**
Women		
R^2	30 (27, 33)*	28 (26, 31)
Harrell's C statistic	0.72 (0.70, 0.73)	0.71 (0.70, 0.72)
Royston's D statistic	1.346 (1.254, 1.438)*	1.292 (1.210, 1.374)
Calibration slope	1 (0.93, 1.07)	1 (0.94, 1.06)
Men		
R^2	22 (20, 24)*	20 (18, 22)
Harrell's C statistic	0.68 (0.67, 0.69)	0.67(0.66, 0.68)
Royston's D statistic	1.082 (1.009, 1.155)	1.033 (0.970, 1.096)
Calibration slope	1 (0.94, 1.06)	1 (0.94, 1.06)

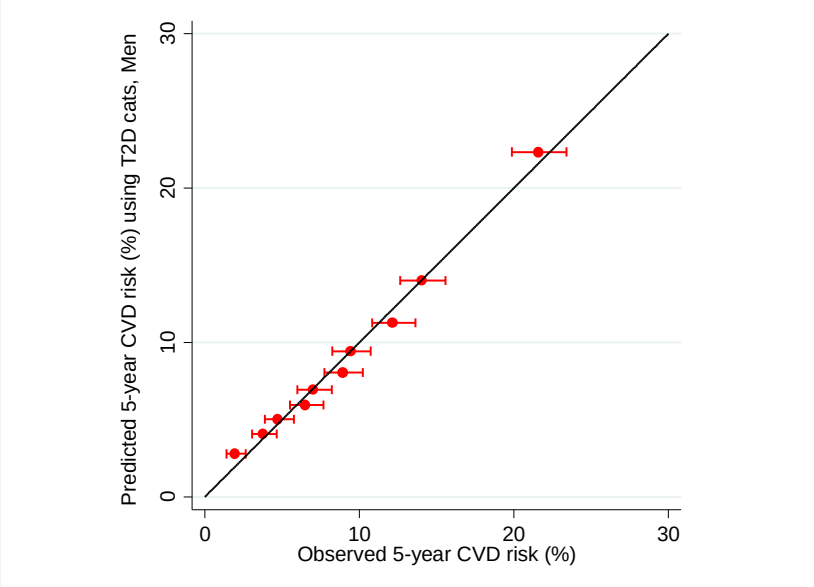
*95% confidence intervals were calculated using 5000 bootstrap replicates.

Appendix 7.4. Calibration plots for PREDICT-1º T2D score with eGFR, ACR, and years since T2D diagnoses categorised*.

a. Women



b. Men



Appendix 9.1.Net proportion of total person-time on additional treatment with blood pressure lowering, lipid lowering and antiplatelet/antithrombotic medications by decile of predicted risk in women and men in the PREDICT primary care cohort, aged 30-74 years

Decile	Women,% person-time treated	Men ,% person-time treated
1	5.5	3.9
2	8.1	4.9
3	9.2	6.1
4	10.2	8.7
5	12.8	10.1
6	13.6	13.5
7	13.9	16.5
8	14.7	19.7
9	14.4	20.9
10	8.4	17.0

In people who were taking preventive (i.e., blood pressure lowering, lipid lowering, or antithrombotic) drugs at baseline, the proportion of person-time they remained on these drugs during follow-up (P1) was computed. Also, for those not taking preventive drugs at baseline, we computed the proportion of person-time that they spent on any of these drugs during follow-up (P2). These proportions were obtained from linked national drug dispensing records. Participants' follow-up time was divided into 6-month periods, and if a specific drug was dispensed during a period, participants were assumed to be taking that drug for those 6 months. We considered the difference (i.e., P2 – P1) to be the net proportion of person-time spent in medication cross-over. If the difference is small, then the effect of cross-over medication effects should cancel out in the fitted models.

The net proportion of person-time spent in cross-over treatment, by deciles of predicted risk, is shown above. As expected, the proportion increased with increasing predicted risk and was on average 12%, with a maximum of about 20%. If a single additional medication was optimistically assumed to reduce risk by 25%, the maximum over-prediction of 5-year risk in any decile would only be 5% (i.e., 25% [relative risk reduction] × 20% [maximum net proportion of people on an additional treatment]). These are tentative estimates and, as far as we are aware, PREDICT is the first study to attempt to explicitly quantify this problem.

References

1. World Health Organization. The global burden of disease: 2004 update. Geneva; 2008.
2. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. 2013.
3. Moons K, Altman D, Reitsma J, Ioannidis J, Macaskill P, Steyerberg E, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162(1):W1-W73.
4. Royston P, Moons K, Altman D, Vergouve Y. Prognosis and prognostic research: developing a prognostic model. BMJ. 2009;338:b604.
5. Lee K, Pryor D, Harrell FJ, Califf R, Behar V, Floyd W. Predicting outcome in coronary disease. Statistical models versus expert clinicians. Am J Med. 1986;80:553-60.
6. Kannel W. Some lessons in Cardiovascular Epidemiology from Framingham. Am J Cardiol. 1976;37(2):269-82.
7. Kannel W, McGee D, Gordon T. A General Cardiovascular Risk Profile: The Framingham Study. Am J Cardiol. 1976;38(1):46-51.
8. Jackson R, Barham P, Maling T, MacMahon S, Bills J, Birch T, et al. Management of raised blood pressure in New Zealand: a discussion document. BMJ. 1993;307(6896):107-10.
9. Anderson K, Odell P, Wilson P, Kannel W. Cardiovascular disease risk profiles. Am Heart J. 1991;121(Number 1, Part 2):293– 8.
10. New Zealand Guideline Group. Assessment and management of cardiovascular risk (Guidance Library: Clinical Area: Cardiovascular) 2003. Available from: http://www.health.govt.nz/publication/assessment-and-management-cardiovascular-risk
11. Task Force. European guidelines on cardiovascular disease prevention in clinical practice: executive summary. Eur J Cardiovasc Prev Rehabil. 2007;14(2):E1-40.
12. Goff DJ, Lloyd-Jones D, Bennett G, Coady S, D'Agostino R, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S74-5.
13. Stone N, Robinson J, Lichtenstein A, Merz C, Blum C, Eckel R, et al. 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. J Am Coll Cardiol. 2014;63(25 Pt B):2889-934.
14. Damen J, Hooft L, Schuit E, Debray T, Collins G, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016;353:i2416.
15. Altman D, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453-73.
16. Benchimol E, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routilely-collected health Data (RECORD) Statement. PLoS Med. 2015;12(10).
17. Laupacis A, Sekar N, Stiell I. Clinical prediction rules: a review and suggested modifications of methodological standards. JAMA. 1997;277(6):488-94.

18. Mallett S, Royston P, Dutton S, Waters R, Altman D. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med. 2010;8(20).
19. Collins G, Mallett S, Omar O, Yu L. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol. 2013(66):268-77.
20. Collins G, Mallett S, Omar O, Yu L. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med. 2011;9(103).
21. Collins G, Reitsma J, Altman D, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement TEST. Br Med J. 2015;350(g7594).
22. Jackson R. Lifetime risk: does it help to decide who gets statins and when? Curr Opin Lipidol. 2014;25(4):247-53.
23. D'Agostino R, Vasan RS, Pencina M, Wolf P, Cobain M, Massaro J, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117:743-53.
24. Wilson P, D'Agostino R, Levy D, Belanger A, Silbershatz H, Kannel W. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97:1837– 47.
25. Chambless L, Folsomb A, Sharrettc A, Sorliec P, Couper D, Szklo M, et al. Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. J Clin Epidemiol. 2003;56:880-90.
26. Ridker P, Paynter N, Rifai N, Gaziano J, Cook N. C-Reactive Protein and Parental History Improve Global Cardiovascular Risk Prediction: The Reynolds Risk Score for Men. Circulation. 2008;118:2243-51.
27. Ridker P, Buring J, Rifai N, Cook N. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA. 2007;297(6):611-9.
28. Elley C, Robinson E, Keneealy T, Bramley D, Drury P. Derivation and Validation of a New Cardiovascular Risk Score for People With Type 2 Diabetes. The New Zealand Diabetes Cohort Study. Diabetes Care. 2010;33:1347–52.
29. Conroy R, Pyörälä K, Fitzgerald A, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003;24:987-1003.
30. Woodward M, Brindle P, Tunstall-Pedoe H. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). Heart. 2007;93:172–6.
31. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ. 2007;335:136.
32. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ. 2008;336:1475.
33. Stevens R, Kothari V, Adler A, Stratton I, Holman R. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). Clin Sci. 2011(101):671-9.
34. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the Prospective Cardiovascular Munster (PROCAM) study. Circulation. 2002;105:310 –5.
35. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. BMJ Clinical Research 2017(357:j2099).

36. Muntner P, Colantonio L, Cushman M, Goff DJ, Howard G, Howard V, et al. Validation of the atherosclerotic cardiovascular disease pooled cohort risk equations. JAMA. 2014;311(14):1406-15.
37. Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluationg assumptions and adequacy, and measuring and reducing error. Stat Med. 1996;15(4):361-87.
38. Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer; 2010.
39. Lagerweij G, Moons K, de Wit G, Koffijberg H. Interpretation of CVD risk predictions in clinical practice: mission impossible? PLoS ONE. 2018;14(1):e0209314.
40. Riley R, Hayden J, Steyerberg E, Moons K, Abrams K, Kyzas P, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2).
41. Cox D. Regression models and life-tables (with discussion). J R Stat Soc Ser B. 1972;34:187-220.
42. Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika. 1982;69:239-41.
43. Therneau T, Grambsch P, Fleming T. Martingale-Based residuals for Survival Models. Biometrika. 1990;77(1):147-60.
44. Cleveland W. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74:829-36.
45. Belsley D, Kuh E, Welsch R. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley & Sons; 1980.
46. Cook R, Peña D, Weisberg S. The likelihood displacement: a unifying principle for influence measures. Communications in Statistics - Theory and Methods. 1988;17(3):623-40.
47. Royston P, Altman D, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127–41.
48. Harrell Jr F. Problems caused by categorizing continuous variables 2004 [cited 2018 March 3]. Available from: http://biostat.mc.vanderbilt.edu/wiki/Main/CatContinuous
49. Lagakos S. Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. Stat Med. 1988;7(1-2):257-74.
50. Austin P, Brunner L. Inflation of the type I error rate when a continuous confounding variable is categorised in logistic regression analyses. Stat Med. 2004;23(7):1159-78.
51. Royston P, Sauerbrei W. Multivariable Model-Building. A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester: John Wiley & Sons; 2008.
52. Royston P, Altman D. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). Appl Stat. 1994;43(3):429-67.
53. Crowson C, Atkinson E, Therneau T. Assessing Calibration of Prognostic Risk Scores. Stat Methods Med Res. 2016;25(4):1692–706.
54. Steyerberg E, Vergouve Y. Towards better clinical prediction models: seven steps to development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-31.
55. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457-81.

56. van Houwelingen H. Validation, calibration, revision and combination of prognostic survival models. Stat Med. 2000;19(24):3401-15.
57. Cox D. Two further applications of a model for binary regression. Biometrika. 1958;45(3/4):562-5.
58. Miller M, Hui S, Tierney W. Validation technique for logistic regression models. Stat Med. 1991;10(8):1213-26.
59. Royston P, Altman D. External validation of a Cox prognostic model: principles and methods. BMC Medical Research Methodology. 2013;13(33).
60. Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982(143):29-36.
61. Harrell F, Califf R, Prior D, Lee K, Rosati R. Evaluating the yield of medical tests. JAMA. 1982;247:2543-6.
62. Gonen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika. 2005;92:965-70.
63. Pencina M, D'Agostino R, Song L. Quantifying discrimination of Framingham risk functions with different survival C statistics. Stat Med. 2012;31(15):1543-53.
64. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med. 2004;23(5):723-48.
65. Blom G. Statistical Estimates and Transformed Beta-Variables. New York: Wiley; 1958.
66. Choodari-Oskooei B, Royston P, Parmar M. A simulation study of predictive ability measures in a survival model: explained variation measures. Stat Med. 2011;31(23):2627-43.
67. Kent J, O'Quigley J. Measures of dependence for censored survival data. Biometrika. 1988;75(3):525--34.
68. Korn E, Simon R. Measures of explained variation for survival data. Stat Med. 1990;9:487--503.
69. O'Quigley J, Flandre P. Predictive capability of proportional hazards regression. Proceedings of the National Academy of Sciences of the United States of America. 1994.
70. Akazawa K. Measures of explained variation for a regression model used in survival analysis. J Med Syst. 1997;21(4):229-38.
71. O'Quigley J, Xu R. Handbook of Statistics in Clinical Oncology. New York: Marcel Dekker; 2001.
72. Royston P. Explained variation for survival models. Stata J.6(1):83-96.
73. Nagelkerke N. A note on a general definition of the coefficient of determination. Biometrika. 1991;78:691-2.
74. O'Quigley J, Xu R, Stare J. Explained randomness in proportional hazards models. Stat Med. 2005;24:479-89.
75. Xu R, O'Quigley J. A measure of dependence for proportional hazards models. J Nonparametr Stat. 1999;12:83-107.
76. Schemper M. The explained variation in proportional hazards regression. Biometrika. 1990;77(1):216-8.
77. Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. Biometrics. 2000;56(1):249-55.
78. Schemper M, Kaider A. A new approach to estimate correlation coefficient in the presence of censoring and proportional hazards. CSDA. 1997;23:467-76.
79. Harrell F. The phglm procedure. SUGI Supplemental Library Users Guide. 1986;5:437-66.
80. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Stat Med. 1999;18:2529-45.

81. Vickers A, Van Calster B, Steyerberg E. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352(i6).
82. Peirce C. The Numerical Measure of the Success of Predictions. Science. 1884;4(93):453-4.
83. Vickers A, Elkin E. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565–74.
84. Janssen K, Vergouwe Y, Kalkman C, Grobbee D, Moons K. A simple method to adjust clinical prediction models to local circumstances. Can J Anaesth.56(3):194-201.
85. Steyerberg E, Borsboom G, van Houwelingen H, Eijkemans M, Habbema J. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2014;23(16):2567-86.
86. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche P, Vandenbroucke J, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806-8.
87. Wells S, Riddell T, Kerr A, Pylypchuk R, Chelimo C, Marshall R, et al. Cohort Profile: The PREDICT Cardiovascular Disease Cohort in New Zealand Primary Care (PREDICT-CVD 19). Int J Epidemiol. 2015;6(1):22.
88. Cardiovascular Guideline Group. Management of Cardiovascular Risk - an Interim Consensus Statement for the Management of Dyslipidaemia Wellington: National Heart Foundation, Stroke Foundation, New Zealand Guidelines Group; 2002.
89. New Zealand Guideline Group. The Management of Type 2 Diabetes. Wellington; 2003.
90. Ministry of Health. Health targets: More heart and diabetes checks [Internet]. Wellington: Ministry of Health; 2014. Available from: http://www.health.govt.nz/new-zealand-health-system/health-targets/about-health-targets/health-targets-more-heart-and-diabetes-checks
91. Ministry of Health (2015) How is My PHO performing? 2014/15 Quarter 3 (January to March) results. Wellington: Ministry of Health; 2015.
92. National Health Index. Wellington, New Zealand: Ministry of Health [Internet]. Wellington, New Zealand: Ministry of Health. Available from: https://www.health.govt.nz/our-work/health-identity/national-health-index
93. The National Minimum Dataset (NMDS). Wellington, New Zealand: Ministry of Health [Internet]. Wellington, New Zealand: Ministry of Health. Available from: https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections/national-minimum-dataset-hospital-events
94. The Mortality Collection. Wellington, New Zealand: Ministry of Health [Internet]. Available from: https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections/mortality-collection
95. The Pharmaceutical Claims Data Mart (PHARMS DM). Wellington, New Zealand: Ministry of Health [Internet]. Available from: https://www.health.govt.nz/publication/pharmaceutical-claims-data-mart-data-dictionary
96. Heath Alliance on behalf of Nothern Regions DHBs [Internet]. Available from: http://www.careconnect.co.nz/testsafe/
97. New Zealand Guidelines Group. Management of type 2 diabetes: Evidence-based best practice guideline. New Zealand Guidelines Group (NZGG) 2003 [Internet]. Available from: www.nzgg.org.nz

98. Ministry of Health. Cardiovascular disease risk asessment. New Zealand primary care handbook 2012. Updated 2013. Wellington, New Zealand
99. National collections and surveys [Internet]. Wellinton, New Zealand,: Ministry of Health. Available from: https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections
100. Morrato E, Elias M, Gericke C. Using population-based routine data for evidence-based health policy decisions: lessons from three examples of setting and evaluating health policy in Australia, the UK and the USA. J Public Health. 2007;29(4):463-71.
101. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323–37.
102. Bohensky M, Jolley D, Sundararajan V, Evans S, Pilcher D, Scott I, et al. Data linkage: a powerful research tool with potential problems. BMC Health Services Research. 2010(10):346.
103. Poppe K, Doughty R, Harwood M, Barber P, Harrison J, Jackson R, et al. Identification, risk assessment, and management of patients with atrial fibrillation in a large primary care cohort. Int J Cardiol. 2018;254:119-24.
104. Pruitt Z, Pracht E. Upcoding emergency admissions for non-life-threatening injuries to children. Am J Manag Care. 2013;19(11):917–24.
105. McLintock K, Russell A, Alderson S, West R, House A, Westerman K. The effects of financial incentives for case finding for depression in patients with diabetes and coronary heart disease: interrupted time series analysis. BMJ Open. 2014;4(8):e005178.
106. Brunt C. CPT Fee Differentials and Upcoding of Physician Visits under Medicare Part B. Health Econ. 2011;20(7):831-41.
107. Nilson F, Bonander C, Andersson R. The effect of the transition from the ninth to the tenth revision of the International Classification of Diseases on external cause registration of injury morbidity in Sweden. Inj Prev. 2015;21(3):189–94.
108. Jagai J, Smith G, Schmid J, Wade T. Trends in gastroenteritis-associated mortality in the United States, 1985 inverted question mark2005: variations by ICD-9 and ICD-10 codes. BMC Gastroenterol. 2014;14(1):211.
109. Merry A, Boer J, Schouten L, Feskens E, Verschuren W, Gorgels A, et al. Validity of coronary heart diseases and heart failure based on hospital discharge and mortality data in the Netherlands using the cardiovascular registry Maastricht cohort study. Eur J Epidemiol. 2009;24(5):237–47
110. Kerr A, Lee M, Jiang Y, Grey C, Wells S, Williams M, et al. High level of capture of coronary intervention and associated Acute Coronary Syndromes in the All New Zealand Acute Coronary Syndrome-QI cardiac register and excellent agreement with National administrative datasets (ANZACS-QI 25). Forthcoming 2018.
111. Hanas R, John G, on behalf of the International HbA1c Consensus Committee. Consensus Statement on the Worldwide Standardization of the Hemoglobin A1c Measurement. Diabetes Med. 2010(27):737-8.
112. Collins G, Altman D. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ. 2012(344):344:e4181.
113. Wells S, Poppe K, Selak V, Kerr A, Pylypchuk R, Wu B, et al. Is general practice identification of prior cardiovascular disease at the time of CVD risk assessment accurate and does it matter? NZMJ. 2018;131 (1475):10-20.

114. Statistics New Zealand. Statistical standard for ethnicity [Internet]. Wellington: Statistics New Zealand; 2005. Available from: http://archive.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx
115. Ministry of Health. HISO 10001:2017 Ethnicity Data Protocols. Wellington: Ministry of Health [Internet]. Wellington: Ministry of Health; 2017.
116. Socioeconomic Deprivation Indexes: NZDep and NZiDep [Internet]. Wellington: University of Otago. Available from: http://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html#overview
117. Atkinson J, Salmond C, Crampton P. NZDep2013 Index of Deprivation [Internet]. Wellington: Department of Public Health, University of Otago; 2014. Available from: https://www.otago.ac.nz/wellington/otago069936.pdf
118. Levey A, Stevens L, Schmid C, Zhang Y, Castro A, Feldman H, et al. A New Equation to Estimate Glomerular Filtration Rate. Ann Intern Med. 2009;150:604-12.
119. Stevens P, Levin A, Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. Ann Intern Med. 2013;158(11):825-30.
120. van Houwelingen H. Validation, calibration, revision and combination of prognostic survival models. Stat Med. 2000;19(24):3401-15.
121. Riddell T, Wells S, Jackson R, Lee A, Crengle S, Bramley D, et al. Performance of Framingham cardiovascular risk scores by ethnic groups in New Zealand: PREDICT CVD-10. NZMJ. 2010;123(1309):50-61.
122. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorpLP2013.
123. Expert Panel on Detection Evaluation and Treatment of High Blood Cholesterol in Adults. Executive Summary of the Third Report og the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment og High Blood Cholestetol in Adults (Adult Treatment Panel III). JAMA. 2001(285):2486-97.
124. Dawber T, Kannel W, Lyell L. An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci. 1963(107):539-56.
125. Kannel W, Feinleib M, McNamara P, Garrison R, Castelli W. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol. 1979;110(3):281-90.
126. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989;129(4):687-702.
127. Fried L, Borhani N, Enright P, Furberg C, Gardin J, Kronmal R. Tha Cardiovascular Health study: design and rationale. Ann Epidemiol. 1991;1(3):263-76.
128. Friedman G, Cutter G, Donahue R, Hughes G, Hulley S, Jacobs DJ. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol. 1988;41(11):1105-16.
129. Walker A. Confounding by Indication. Epidemiology. 1996;7(4):335-6.
130. Broad J, Wells S, Marshall R, Jackson R. Zero end-digit preference in recorded blood pressure and its impact on classification of patients for pharmacologic management in primary care - PREDICT-CVD-6. Br J Gen Prac. 2007;57(544):897-903.
131. Harrell F. Regression modelling strategies: with applicationsto linear models, logistic regression, and survival analysis. New York: Springer; 2001.

132. Molinaro A, Simon R, Pfeiffer R. Prediction error estimation: a comparison of resampling methods. Bioinformatics. 2005;21(15):3301-7.
133. Schumacher M, Binder H, Gerds T. Asessment of sutvival prediction models based on microarray data. Bioinformatics. 2007;23(14):1768-74.
134. Steyerberg E, Harrell Jr F, Borsboom G, Eijkemans M, Vergouwe Y, Habbema J. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.
135. Damen J, Hooft L, Schuit E, Debray T, Collins G, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016;353(i2416).
136. Gijsberts C, Groenewegen K, Hoefer I, Eijkemans M, Asselbergs F, Anderson T, et al. Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events. PLoS ONE. 2015;10(7):e0132321.
137. Lloyd-Jones D, Nam B, D'Agostino RS, Levy D, Murabito J, Wang T, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA. 2004;291(18):2204-11.
138. Murabito J, Pencina MJ, Nam B, D'Agostino R, Wang T, Lloyd-Jones D, et al. Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. JAMA. 2005;294(24):3117-23.
139. Cook N, Ridker P. Calibration of the Pooled Cohort Equations for Atherosclerotic Cardiovascular Disease. An update. Ann Intern Med. 2016;165(11):786-94.
140. Vittinghoff E, McCulloch C. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165(6):710-8.
141. Mannsverk J, Wilsgaard T, Mathiesen E, Løchen M, Rasmussen K, Thelle D, et al. Trends in Modifiable Risk Factors Are Associated With Declining Incidence of Hospitalized and Nonhospitalized Acute Coronary Heart Disease in a Population. Circulation. 2016;133(1):74-81.
142. Woodward M, Zhang X, Barz F, Pan W, Ueshima H, Rodgers A, et al. The effects of diabetes on the risks of major cardiovascular diseases and death in the Asia-Pacific region. Diabetes Care. 2003;26(2):360-6.
143. Sarwar N, Gao P, Seshasai S, Gobin R, Kaptoge S, Di Angelantonio E, et al. Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375(9733):2215-22.
144. British Cardiac Society, British Hypertension Society, Diabetes UK, HEART UK, Primary Care Cardiovascular Society, Stroke Association. JBS 2: Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice. Heart. 2005;91(5):1-52.
145. American Diabetes Association. Standards of medical care in diabetese2011. Diabetes care. 2011;34(Suppl 1):11-61.
146. Canadian Diabetes Association. Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Can J Diabetes. 2008;32(1):1-201.
147. Davis W, Knuiman M, Davis T. An Australian cardiovascular risk equation for type 2 diabetes: the Fremantle Diabetes Study. Intern Med J. 2010;40:286-92.
148. Cederholm J, Eeg-Olofsson K, Eliasson B, Zethelius B, Nilsson P, Gudbjörnsdottir S. Swedish National Diabetes Register. Source Department of Public Health and Carin. Risk prediction of cardiovascular disease in type 2

diabetes: a risk equation from the Swedish National Diabetes Register. Diabetes Care. 2008(31):2038-43.
149. Folsom A, Chambless L, Duncan B, Gilbert A, Pankow J, Atherosclerosis Risk in Communities Study Investigators. Prediction of coronary heart disease in middle-aged adults with diabetes. Diabetes Care. 2003;26(10):2777-84.
150. Donnan P, Donnelly L, New J, Morris A. Derivation and validation of a prediction score for major coronary heart disease events in a U.K. type 2 diabetic population. Diabetes Care. 2006;29(6):1231-6.
151. Yang X, So W, Kong A, Ma R, Ko G, Ho C, et al. Development and validation of a total coronary heart disease risk score in type 2 diabetes mellitus. Am J Cardiol. 2008;101(5):596-601.
152. van Dieren S, Beulens J, Kengne A, Peelen L, Rutten G, Woodward M, et al. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. Heart. 2012(98):360-9.
153. Chamnan P, Simmons R, Sharp S, Griffin S, Wareham N. Cardiovascular risk assessment scores for people with diabetes: a systematic review. Diabetologia. 2009(52):2001–14.
154. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet. 2010;375:2073-81.
155. Vittinghoff E, McCulloh C. Relaxing the rule of ten events per variable in logistic and Cox regression. American Journal of Epidemiology. 2007(165):710-8.
156. Kannel W, McGee D. Diabetes and cardiovascular disease: the Framingham study. JAMA. 1979(241):2035– 8.
157. Hayes J. Non-ischeamic heart disease in diabetes melitus. In: Stout R, editor. Diabetes and Atherosclerosis. Dordrecht: Kluwer Academic Publishers; 1992. p. 255-65.
158. Naas A, Davidson N, Thompson C, Cummings F, Ogston S, Jung R, et al. QT and QTc dispersion are accurate predictors of cardiac death in newly diagnosed non-insulin dependent diabetes: cohort study. BMJ. 1998;316(745-746).
159. Fecher M, Elkeles R. Lipid modification and coronary heart disease in type 2 diabetes: different from the general population? Heart. 1999(81):10-1.
160. Manzato E, Zambon A, Lapolla A, Zambon S, Braghetto L, Crepaldi G, et al. Lipoprotein abnormalities in well-treated type II diabetic patients. Diabetes Care. 1993(16):469-75.
161. Meade T, Mellows S, Brozovic M, Miller G, Chakrabarti R, North W, et al. Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. Lancet. 1986;2(8506):533-7.
162. Mayne E. Haemostatic disorders in diabetes mellitus. In: Stout R, editor. Diabetes and Atherosclerosis. Dordrecht: Kluwer Academic Publishers; 1992. p. 219-35.
163. Yarnell J, Baker I, Sweetnam P, Bainton D, O'Brien J, Whitehead P, et al. Fibrinogen, viscosity, and white blood cell count are major risk factors for ischemic heart disease. The Caerphilly and Speedwell Collaborative Heart Disease Studies. Circulation. 1991(83):836-44.
164. Kannel W. Overview of hemostatic factors involved in atherosclerotic cardiovascular disease. Lipids. 2005;40(12):1215-20.
165. Kannel W, D'Agostino R, Wilson P, Belanger A, Gagnon D. Diabetes, fibrinogen, and risk of cardiovascular disease: The Framingham experience. Am Heart J. 1990;120(3):672-6.
166. Wayne E, Bridges J, Weaver J. Platelet adhesiveness, plasma fibrinogen and factor levels in diabetes mellitus. Diabetologia. 1970(6):436-40.

167. Chakabarti R, Meade T. Clotting factors, platelet function and fibrinolytic activity in diabetics and in a comparison group, WHO Multinational Group. Diabetologia. 1976(12):383-6.
168. The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk od vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375(9733):2215-22.
169. Thompson W. Effect modification and the limits of biological inference from epidemiologic data J Clin Epidemiol. 1991;44(3):221-32.
170. Greenland S. Basic Problems in Interaction Assessment. Environ Health Perspect. 1993;101(Suppl4):59-66.
171. Rothman K, Greenland S, Walker A. Concepts of interaction. Am J Epidemiol. 1980(112):467-70.
172. Rothman K. Synergy and antagonism in cause-effect relationships. Am J Epidemiol. 1974(99):385-8.
173. Rothman K. Modern epidemiology. Boston: Little, Brown & Co; 1986.
174. Guzder R, Gatling W, Mullee M, Mehta R, Byrne C. Prognostic value of the Framingham cardiovascular risk equation and the UKPDS risk engine for coronary heart disease in newly diagnosed type 2 diabetes: results from a United Kingdom study. Diabet Med. 2005(22):554-62.
175. Richard J. Stevens VK, Amanda I. Adler, Irene M. Stratton, Rury R. Holman. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). Clinical Science. 2001;101:671-9.
176. Gatling W, Guzder R, Turnbull J, Budd S, Mullee M. The Poole Diabetes Study: how many cases of Type 2 diabetes are diagnosed each year during normal health care in a defined community? . Diabetes Res Clin Prac. 2001(53):107-12.
177. Day N, Oaks S, Luben R. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br J Cancer. 1999;80(Suppl 1):95-103.
178. Simmons R, Sharp S, Boekholdt S. Evaluation of Framingham risk score in the European Prospective Investigation of Cancer - Norfolk cohort: does adding glycated hemoglobin improve the prediction of coronary heart disease events? Arch Intern Med. 2008;168(1209-1216).
179. Stephens J, Ambler G, Vallance P, Betteridge D, Humphries S, Hurel S. Cardiovascular risk and diabetes. Are the methods of risk prediction satisfactory? Eur J Cardiovasc Prev Rehabilitation. 2004;11(6):521-8.
180. Stevens R, Coleman R, Holman R. Framingham risk equations underestimate coronary heart disease risk in diabetes. Diabet Med. 2005(22):228.
181. Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk asessment in the primary prevention of cardiovascular disease: a systematic review. Heart. 2006;92(12):1752-9.
182. van der Leeuw J, van Dieren S, Beulens J, Boeing H, Spijkerman A, van der Graaf Y, et al. The validation of cardiovascular risk scores for patients with type 2 diabetes mellitus. Heart. 2014(101):222-9.
183. Liew S, Doust J, Glasziou P. Cardiovascular risk scores do not account for the effect of treatment: a review. Heart. 2011;97(9):689-97.
184. Ministry of Health. 2018. Cardiovascular Disease Risk Assessment and Management for Primary Care [Internet]. Wellington: Ministry of Health; 2018 [cited April 2018]. Available from: https://www.health.govt.nz/publication/cardiovascular-disease-risk-assessment-and-management-primary-care
185. Cook N. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M.J.

Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med. 2008;27(2):191-5.
186. Kooter A, Kostense P, Groenewold J, Thijs A, Sattar N, Smulders Y. Integrating information from novel risk factors with calculated risks: the critical impact of risk factor prevalence. Circulation. 2011;741(5):741-5.
187. Beam A, Kohane I. Big data and mashine learning in health care. JAMA. 2018;319(13):1317-8.
188. Wolfson J, Bandyopadhyay S, Elidrisi M, Vazquez-Benitez G, Vock D, Musgrove D, et al. A Naive Bayes machine learning approach to risk prediction using censored, time-to-event data. Stat Med 2015;34(21):2941–57.
189. Teramukai S, Okuda Y, Miyazaki S, Kawamori R, Shirayama M, Teramoto T. Dynamic prediction model and risk assessment chart for cardiovascular disease based on on-treatment blood pressure and baseline risk factors. Hypentens Res. 2016;39(2):113–8.
190. Hickey G, Grant S, Caiado C, Kendall S, Dunning J, Poullis M, et al. Dynamic prediction modeling approaches for cardiac surgery. Circ Cardiovasc Qual Outcomes. 2013;6(6):649-58.