International Journal of Epidemiology, 2017, 22–22j
doi: 10.1093/ije/dyv312
Advance Access Publication Date: 20 December 2015
Cohort Profile

Cohort Profile

Cohort Profile: The PREDICT Cardiovascular Disease Cohort in New Zealand Primary Care (PREDICT-CVD 19)

Sue Wells,[1]* Tania Riddell,[1] Andrew Kerr,[1,2] Romana Pylypchuk,[1] Carol Chelimo,[1] Roger Marshall,[1] Daniel J Exeter,[1] Suneela Mehta,[1] Jeff Harrison,[3] Cam Kyle,[4] Corina Grey,[1] Patricia Metcalf,[1] Jim Warren,[5] Tim Kenealy,[6] Paul L Drury,[4] Matire Harwood,[1] Dale Bramley,[7] Geeta Gala[8] and Rod Jackson[1]

[1]School of Population Health, University of Auckland, Auckland, New Zealand, [2]Middlemore Hospital, Cardiology Department, Auckland, New Zealand, [3]School of Pharmacy, University of Auckland, Auckland, New Zealand, [4]Endocrinology Services, Auckland District Health Board, Auckland, New Zealand, [5]Computer Sciences, University of Auckland, [6]School of Medicine, University of Auckland, Auckland, New Zealand, [7]Waitemata District Health Board, Auckland, New Zealand and [8]Northern Regional Alliance, Auckland, New Zealand

*Corresponding author. Section of Epidemiology and Biostatistics, School of Population Health, University of Auckland, P.O. Box 92019 Auckland Mail Centre, Auckland, New Zealand. E-mail: s.wells@auckland.ac.nz

Accepted 27 October 2015

Why was the cohort set up?

The PREDICT-CVD cohort was set up to facilitate research into aspects of cardiovascular disease (CVD) and to assist CVD risk assessment and management in routine general practice. In particular, there were concerns about how to best assess CVD risk and whether the Framingham risk model[1] best served New Zealand. The Framingham CVD score was developed from the Framingham Heart Study, a cohort study initiated in the early 1950s in New England, USA.[1] Clinicians had concerns about the accuracy of a risk score derived from a non-contemporaneous cohort and, in particular, the validity of this algorithm for high-risk groups such as Māori, Pacific and South Asian populations, or people with diabetes. Furthermore, the Framingham CVD risk score was developed for adults aged 35–74 years without a previous history of a CVD event (so could not be used to support secondary prevention risk management) and its applicability to an increasing population aged over 75 years was also unknown. The objectives were underpinned by principles of equity for Māori as New Zealand's indigenous peoples and equity according to health need. This included ethnic-specific estimation of CVD risk, more robust quantification of risk profile differences, closing gaps in evidence-based practice and reduction in vascular outcome inequities.

The PREDICT-CVD cohort was established in 2002 when a web-based CVD risk assessment and management decision support system (called 'PREDICT') was developed for primary care. Recognizing the difficulty with collecting complete prospective risk factor information in large cohorts, we took advantage of the widespread use of computers in general practice to set up a research cohort and simultaneously to implement national CVD and diabetes guidelines.[2–4] PREDICT is integrated with general practice electronic health records (EHRs) and incorporates data collection into clinical workflow. When clinicians use

PREDICT during a patient visit, risk scores and evidence-based treatment recommendations tailored to the patient's CVD and diabetes profile are computed and displayed. At the same time, a copy of the patient's CVD risk profile is securely stored both in the EHR and on a secure off-site server held by a private IT company (Enigma Solutions Ltd) on behalf of primary care providers.

Data linkage

Over 98% of New Zealanders have a unique health identifier (the National Health Index number or NHI) which identifies individuals in publicly funded health system databases.[5,6] With provider permission, patient risk factor profiles are anonymized by encrypting the NHI and are then transferred from the Enigma server to the University of Auckland. The ever-growing PREDICT cohort is then annually linked to health databases via similarly encrypted NHIs to routine national databases that include: medication dispensing; laboratory test claims; enrolments in primary health organizations; hospitalizations; and deaths.

Ethical approval

New Zealand ethics committees allow secondary re-use of health data without individual patient consent where data are not identifiable. Information about the PREDICT study is available at all general practice locations, and patients may opt out of having their de-identified data being included in the cohort. The PREDICT study was approved by the Northern Region Ethics Committee Y in 2003 (AKY/03/12/314) with subsequent annual approval by the National Multi-region Ethics Committee since 2007 (MEC07/19/EXP).

Who is in the cohort?

The cohort includes all people who have their CVD risk assessed by a general practitioner (GP) or practice nurse, entering patient data into PREDICT-CVD online forms. The PREDICT software was initially implemented in 2002 in general practices in Auckland and is now available in approximately 35–40% of New Zealand primary care practices. These are mainly in the Auckland and Northland regions, serving around 1.6 million people and representing around 35% of the New Zealand resident population.[7]

Since 2003, New Zealand CVD risk management guidelines have recommended that men aged over 45 years and women over 55 years have a regular CVD risk assessment, with the frequency of assessment and the intensity of risk management informed by their calculated 5-year CVD risk based on the Framingham risk model.[3] For specified sub-populations, the guidelines recommend that risk assessments start 10 years earlier than for the general population. These sub-populations include people of Māori, Pacific or South Asian ethnicity and other individuals with known CVD risk factors (such as smoking, diabetes and raised blood pressure or cholesterol levels). For people aged over 75 years, CVD risk assessment and management are also recommended, particularly for those without significant comorbidities and with reasonable life expectancy.[3] Whether a person visiting the primary care clinic is risk assessed or not, and therefore whether they enter the cohort, is at the discretion of the doctor or nurse. Most GPs have electronic reminders within their EHR that provide alerts for individual patient eligibility.The only entry exclusion to the online form is current pregnancy, and no decision support is given for those under 18 years of age.

All primary care patients risk assessed using PREDICT software between August 2002 and August 2012 are included in this cohort description. Figure 1 shows annual recruitment for all patients and by three sub-cohorts: those with a history of atherosclerotic CVD who may also have comorbid diabetes (the CVD sub-cohort); those with diabetes but no CVD (the Diabetes sub-cohort); and those with neither diabetes nor a history of CVD (the no CVD/no Diabetes sub-cohort).

CVD sub-cohort

The CVD sub-cohort was derived using data from three sources. First, in the PREDICT risk assessment process, data were used if the GP or nurse had classified patients with a known history of: angina or myocardial infarction; percutaneous coronary intervention or coronary artery bypass graft; ischaemic stroke or transient ischaemic attack; peripheral vascular disease or history of atherosclerotic vascular surgery. Second, linked national hospitalization data were used to identify patients who had had a publicly funded CVD-related hospital admission before the first (baseline) PREDICT assessment. [The International Classification of Diseases, version 10 Australian Modification (ICD-10 AM) codes used to define a CVD-related hospitalization or history of previous CVD are available on request to researchers interested in data sharing and collaboration.] Most (more than 95%) of CVD hospitalizations occur within New Zealand's state-funded public health service.[8] Third, the linked Pharmaceutical Collection (PHARMS), a national database of subsidized pharmaceutical dispensing, was used to identify patients who had three or more prescriptions before their baseline risk assessment, of the following anti-anginal medications: glyceryl trinitrate, isosorbide dinitrate, isosorbide mononitrate, nicorandil or perhexiline.[9] These anti-anginal medications (and all others included in these analyses) are government subsidized. Reliable identification

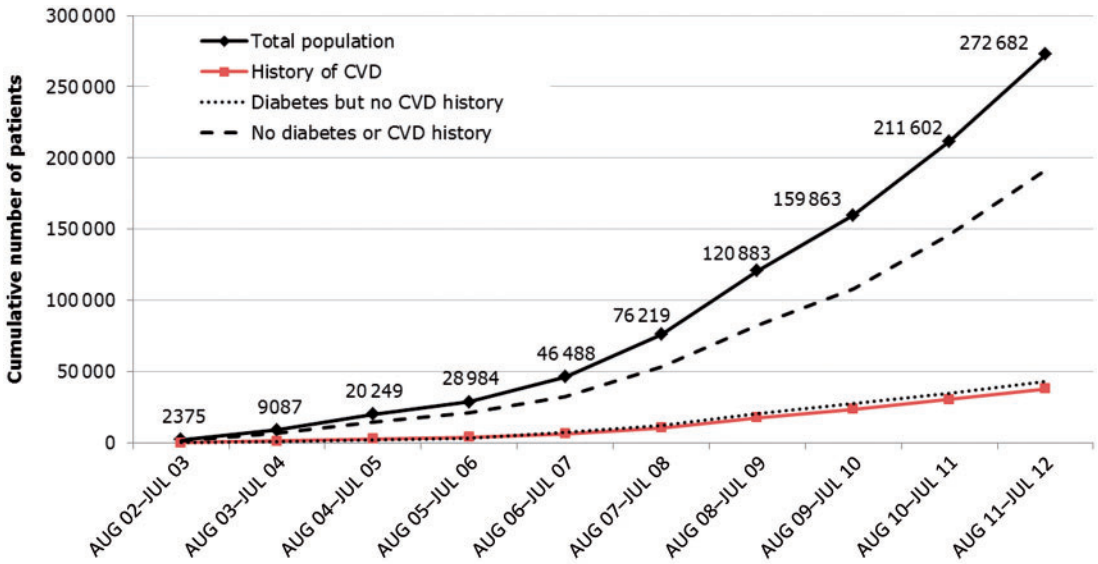

Figure 1. Number of patients recruited into the PREDICT cohort between August 2002 and July 2012.

of dispensing episodes by NHI number has increased over the past decade from 64% in 2004, to 92% in 2006 and over 96% from 2010 onwards (S. Ross, Ministry of Health personal communication, 2014).

Diabetes sub-cohort

A history of diabetes was also derived using data from the same three sources, that is: (i) the PREDICT database if the primary care practitioner classified patients as type 1, type 2 or diabetes type unknown; (ii) the national hospitalization database if patients had been discharged with ICD-9 or 10 AM codes: ICD-9 250 (Diabetes), ICD-10 AM E10 to E14 (Diabetes mellitus); or (iii) the PHARMS database if patients had been dispensed one or more prescriptions of oral hypoglycaemic agents or insulin.

Demographic characteristics of the total cohort ($n = 272\,682$) stratified by age groups are presented in Table 1. Just over 90% were between 35 and 74 years of age. Ethnicity was defined according to a national prioritization protocol[10] in the following order: Māori, Pacific, Indian, Other Asian and finally NZ European combined with other ethnicities (NZEO). Whereas the majority were in the combined NZEO category (58.5%), there were sizeable groups of Māori, Pacific and Indian patients at 35 996 (13.2%), 37 167 (13.6%) and 20 644 (7.6%), respectively. Socioeconomic status was assessed using the New Zealand Deprivation Index Score (NZDep), which is a measure assigned to a patient's area of residence. NZDep is based on nine variables from the Census, reflecting eight dimensions of relative deprivation of census tracts.[11] For these analyses, quintiles (1 to 5, from least to most deprived) of the nationwide distribution of NZDep score were obtained for each of the census tracts. Generally, more deprived areas are over-represented; nearly half (48.9%) of the PREDICT participants lived in the two most deprived quintile areas.

In terms of the cohort's representativeness of the general population, the socio-demographic distribution of the cohort is strongly influenced by New Zealand CVD guidelines recommendations for screening (as noted in the beginning of this section) and national funding priorities. Over the past decade, the Ministry of Health has provided additional funding to primary care to screen high-risk disadvantaged groups such as Maori, Pacific and those living in NZDep Quintile 5 as well as separate funding for annual reviews for people with diabetes (that includes a CVD risk assessment).

The 55–74 years age group is largely representative of the age, sex and ethnicity distribution of the Auckland and Northland regions. Women aged below 55 years are under-represented, and Maori, Pacific and Indian patients, those living in the most deprived areas and people with diabetes are over-represented in our cohort.

What has been measured?

When clinicians open a PREDICT form within a patient's EHR, the software automatically fills in the form with relevant clinical and demographic data from the medical record. This can then be checked and any missing or incorrect fields can be updated by the clinician. The software will not calculate a patient's 5-year CVD risk unless all compulsory risk

Table 1. Characteristics of patients in the PREDICT cohort, August 2002 to July 2012, by age group

	Total, *n*	Age group, *n*							
Variable	272 682	< 25	25–34	35–44	45–54	55–64	65–74	75–84	85+
	n (%)	689	4197	32963	79 483	81 989	50 979	18 348	4034
Sex									
Female	121 760 (44.7)	45.3	37.3	23.6	39.6	51.4	51.1	54.2	61.4
Male	150 922 (55.4)	54.7	62.7	76.4	60.4	48.6	48.9	45.8	38.6
Ethnicity									
European/Other	159 389 (58.5)	47.9	42.6	31.6	48.4	65.0	72.0	80.6	89.7
NZ Māori	35 966 (13.2)	23.5	20.3	21.3	17.0	10.8	8.6	5.9	2.5
Pacific	37 167 (13.6)	17.7	19.1	26.6	16.9	10.4	8.4	5.8	3.4
Indian	20 644 (7.6)	5.7	12.6	15.2	9.4	5.9	4.3	2.6	1.4
Other Asian	19 516 (7.2)	5.2	5.3	5.4	8.3	7.9	6.7	5.2	3.1
[a]NZDep 2001 (quintiles)									
1 (least deprived)	48 764 (17.9)	13.8	11.1	12.7	17.1	20.2	19.9	16.8	17.3
2	44 269 (16.2)	14.5	13.9	13.7	16.1	17.1	16.7	16.5	17.4
3	46 063 (16.9)	13.4	15.6	14.2	15.5	16.9	18.9	21.6	21.8
4	56 877 (20.9)	22.1	25.6	21.7	19.7	20.3	21.1	24.2	25.8
5 (most deprived)	76 235 (28.0)	36.0	33.4	37.6	31.5	25.3	23.4	20.8	17.7
History of:									
Total CVD[b]	38 098 (14.0)	4.9	2.3	3.3	6.2	12.3	24.0	41.1	52.8
Coronary heart disease	26 849 (70.5)	8.8	44.8	61.1	69.4	71.0	70.4	72.2	72.0
Stroke, TIA	10 518 (27.6)	14.7	24.0	24.3	23.2	24.2	28.0	31.8	38.8
Peripheral vascular disease	6382 (16.8)	76.5	29.2	17.7	13.7	14.2	16.9	20.0	21.2
PCI or CABG	11 875 (31.2)	0.0	28.1	26.5	31.0	32.1	32.9	31.0	20.8
Heart failure	12 833 (4.7)	0.2	0.9	1.1	1.7	3.4	7.5	17.4	31.5
Diabetes mellitus	56 944 (20.9)	57.6	35.2	18.0	16.6	19.5	24.6	33.3	32.7

History of CVD and Heart Failure not mutually exclusive: 4358 patients out of 6238 with a history of Heart Failure also have a history of CVD.

TIA = transient ischaemic attack.

[a]Missing data values for 474 patients (0.2%).

[b]Total CVD = coronary heart disease, stroke/transient ischaemic attack (TIA), peripheral vascular disease, percoronary intervention (PCI) or coronary artery bypass grafting (CABG).

assessment fields are completed. Once a completed form is submitted by the clinician, a risk score is computed and returned interactively to the clinician. This score and the patient risk profile are stored both in the EHR and on the centrally hosted secure webserver. After de-identification, the patient profile is able to contribute to the cohort.

The following variables are required fields for calculating a patient's 5-year CVD risk: date of birth, sex, ethnicity, previous diagnosis of CVD (as defined in the previous section), diabetes (type 1, type 2 or type unknown), atrial fibrillation, a self-reported family history of premature ischaemic CVD, smoking status, systolic and diastolic blood pressure (mean of two measures) and total cholesterol to high-density lipoprotein (HDL)-cholesterol ratio (one measure). Family history of premature CVD was defined as a self-reported familial history of ischaemic heart disease or ischaemic stroke occurring in a father or brother before 55 years of age, or a mother or sister before 65 years of age.

Other lipid fractions, body mass index (BMI) and dispensed cardiovascular medications (classified into blood pressure (BP)-lowering and lipid-lowering medications, antiplatelet and anticoagulant agents) may also be filled in but are not compulsory for CVD risk assessment. These variables are routinely entered if clinicians require individualized guideline-based recommendations for patient management. If not provided on the PREDICT form, lipid profiles can be augmented by linkage to 'TestSafe', a regional laboratory repository for nearly all the patients/primary care providers in this study. This database was developed in 2005 to hold all test results (Biochemical, Haematological, Microbiological and Histological) undertaken for patients in hospital and in the community. Dispensed cardiovascular medications along with NZDep scores can also be obtained from the linked national databases.

A history of congestive heart failure is not a required field on the PREDICT form but is an important comorbidity. Data on this variable were derived from two sources: linked hospitalization data using ICD-9 or 10 AM codes (I110, 130, 132, I500-501, 509) and linked pharmaceutical data on dispensed oral loop diuretics. The latter has been used by other cohort studies as a proxy for congestive heart failure.[12,13]

Baseline medical history and CVD risk factors are reported in Table 2 for the CVD sub-cohort, Table 3 for the Diabetes sub-cohort and Table 4 for the no CVD/no Diabetes sub-cohort. The CVD sub-cohort also had a high prevalence of other comorbid conditions; 36.0% had diabetes, 16.4% had atrial fibrillation, 19.6% had heart failure and 29.8% were categorized as being obese. The majority were on aspirin (64.7%) and/or a statin (68.2%) and/or blood pressure lowering drugs (79.5%).

The Diabetes sub-cohort had a lower prevalence of atrial fibrillation (3.4%) and heart failure (5.5%) but a higher proportion were obese (47.2%). The majority (64%) were on a blood pressure lowering drug. The no CVD/no Diabetes sub-cohort ($n = 191\,343$) had the lowest prevalence of atrial fibrillation (2.1%), heart failure (1.6%) and obesity (23.2%) although a quarter were on blood pressure lowering drugs. Smoking prevalence was similar across all three sub-cohorts, at about 13–14%.

How often have they been followed up?

PREDICT was designed as a clinical tool and a baseline risk factor cohort with follow-up determined by routine clinical practice. Repeated measures of risk factors have been undertaken and recorded when deemed relevant by the patient and/or practitioner. Approximately 40% of patients have had repeat assessments to date. However, the entire cohort has been 'electronically followed' every 2–3 years through encrypted NHI linkage to routine national databases.

CVD outcome events were classified as ischaemic cardiovascular events if a hospital discharge included ICD-10 AM codes for a diagnosis of acute coronary syndrome,

Table 2. History of CVD sub-cohort: baseline medical history and CVD risk factors, August 2002 to July 2012

Variable	Total	Age group, *n*							
	n 38 098	< 25	25–34	35–44	45–54	55–64	65–74	75–84	85+
	n (%)	34	96	1095	4888	10 065	12 246	7546	2128
Medical history:									
Diabetes mellitus	13 703 (36.0)	76.5	39.6	31.9	32.5	36.1	36.0	38.8	34.8
Coronary heart disease	26 849 (70.5)	8.8	44.8	61.1	69.4	71.0	70.4	72.2	72.0
Stroke/TIA[a]	10 518 (27.6)	14.7	24.0	24.3	23.2	24.2	28.0	31.8	38.8
Peripheral vascular disease	6382 (16.8)	76.5	29.2	17.7	13.7	14.2	16.9	20.0	21.2
[a]PCI/CABG	11 875 (31.2)	0.0	28.1	26.5	31.0	32.1	32.9	31.0	20.8
Family history of CVD	5968 (15.7)	11.8	16.7	22.1	22.1	18.4	14.6	11.0	7.6
[b]Atrial fibrillation (AF)	6233 (16.4)	0.0	5.2	7.4	7.5	11.1	16.2	25.9	34.6
Heart failure	7467 (19.6)	2.9	9.4	10.5	11.5	14.4	18.6	28.9	41.4
Risk factor:									
Smoking									
Yes (current)	4892 (12.8)	38.2	22.9	31.4	24.7	17.4	9.9	4.1	1.8
Past (former)	10 273 (27.0)	2.9	19.8	19.0	23.6	25.6	29.2	29.1	25.4
No (never)	22 933 (60.2)	58.8	57.3	49.6	51.8	57.0	61.0	66.7	72.8
[c]Body mass index (BMI)									
Underweight/normal (< 25)	5919 (15.5)	23.5	7.3	8.0	9.8	11.1	14.9	22.1	34.5
Overweight (25–29.9)	10 202 (26.8)	29.4	18.8	17.4	22.0	24.5	28.6	31.0	28.6
Obesity (30+)	11 335 (29.8)	41.2	47.9	44.8	39.7	35.6	28.4	20.4	12.0
		Mean (SD)							
Systolic blood pressure (mm Hg)	133 (18)	122 (15)	127 (15)	128 (18)	130 (18)	132 (18)	134 (18)	135 (18)	134 (19)
Diastolic blood pressure (mm Hg)	77 (11)	78 (14)	82 (12)	82 (12)	81 (11)	79 (10)	76 (10)	74 (10)	73 (10)
[d]Total cholesterol: HDL ratio	3.7 (1.2)	4.3 (1.4)	4.9 (1.8)	4.6 (1.5)	4.2 (1.4)	3.9 (1.2)	3.6 (1.1)	3.5 (1.1)	3.4 (1.3)
Medications at baseline									
Aspirin	24 662 (64.7)	2.9	33.3	43.8	56.4	63.2	67.8	69.8	68.8
Clopidogrel	1848 (4.9)	0.0	2.1	7.1	7.0	4.9	4.6	3.8	4.0
Warfarin	3712 (9.7)	0.0	12.5	8.4	6.6	7.3	9.7	14.2	14.2
BP-lowering drugs	30 291 (79.5)	14.7	44.8	57.1	68.2	76.7	82.4	87.3	89.0
Statin	25 981 (68.2)	5.9	32.3	49.4	63.1	70.4	71.4	70.3	55.9
Other lipid-lowering drugs	2430 (6.4)	0.0	1.0	4.5	6.1	7.4	7.1	5.4	3.2

[a]Transient ischaemic attack (TIA), percoronary intervention (PCI), coronary artery bypass grafting (CABG).
[b]AF not assessed for 3507 patients (9.2%) in initial PREDICT online form 2003–05.
[c]BMI missing data values for 10 642 patients (27.9%)
[d]Total cholesterol: HDL ratio missing data values for 5 patients.

Table 3. Diabetes sub-cohort: baseline medical history and CVD risk factors, August 2002 to July 2012

Variable	Total, $n = 43241$ n (%)	Age group, n < 25 371	25-34 1438	35–44 5585	45–54 11 568)	55–64 12 383	65–74 8141	75–84 3177	85+ 578
Medical history:									
Family history CVD	4288 (9.9)	7.0	10.7	12.1	11.1	9.9	8.2	7.3	6.1
[a]Atrial fibrillation (AF)	1447 (3.4)	0.5	0.4	0.9	1.5	3.0	5.5	10.1	15.4
Heart failure	2355(5.5)	0.0	1.0	1.9	3.1	5.0	8.3	13.8	25.1
Risk factor:									
[b]Smoking									
Yes (current)	6045 (14.0)	20.2	24.5	21.2	18.2	12.7	7.6	4.0	2.4
Past (former)	7820 (18.1)	8.4	11.5	14.0	15.4	19.2	22.2	23.9	18.7
No (never)	29 375 (67.9)	71.4	64.0	64.8	66.4	68.1	70.2	72.1	78.9
[c]Body mass index (BMI)									
Underweight/normal (< 25)	5575 (12.9)	34.2	14.3	9.3	9.6	11.9	15.4	22.1	32.5
Overweight (25–29.9)	10 681 (24.7)	23.5	18.7	19.7	22.6	24.8	27.9	33.6	35.8
Obesity (30+)	20 393 (47.2)	36.4	57.5	56.2	52.3	47.0	40.8	31.1	19.7
		Mean (SD)							
Systolic blood pressure (mm Hg)	132 (18)	120 (15)	124 (16)	128 (16)	131 (17)	134 (17)	136 (17)	137 (18)	137 (20)
Diastolic blood pressure (mm Hg)	80 (11)	75 (11)	80 (11)	82 (11)	82 (11)	80 (10)	77 (10)	75 (10)	74 (11)
[d]Total cholesterol: HDL ratio	4.1 (1.3)	4.0 (1.5)	4.6 (1.5)	4.6 (1.5)	4.3 (1.3)	4.0 (1.2)	3.8 (1.1)	3.6 (1.1)	3.5 (1.1)
Medications at baseline									
Aspirin	17 543 (40.6)	2.4	11.8	24.9	36.6	46.1	50.1	52.3	50.5
Clopidogrel	69 (0.2)	0.0	0.0	0.1	0.1	0.2	0.2	0.2	0.2
Warfarin	822 (1.9)	0.0	0.3	0.6	0.8	1.6	3.4	5.6	6.4
BP-lowering drugs	27 655 (64.0)	15.4	33.8	46.9	57.8	68.9	75.7	82.5	84.4
Statin	23 498 (54.3)	10.0	27.7	42.6	52.9	60.6	61.7	56.4	42.6
Other lipid-lowering drugs	1549 (3.6)	0.0	1.6	2.6	3.4	4.0	4.1	4.6	2.9

[a]AF not assessed for 3037 patients (7.0%) in initial PREDICT online form 2003–05.
[b]Smoking missing data values for 1 patient.
[c]BMI missing data values for 6592 patients (15.2%).
[d]Total cholesterol: HDL ratio missing data values for 10 patients.

ischaemic or haemorrhagic stroke, peripheral arterial disease, coronary or peripheral arterial procedures or congestive heart failure. Deaths were classified as ischaemic CVD if the underlying cause of death was from the same range of codes as for CVD hospitalization (above) or from sudden death ascribed to acute myocardial infarction. For deaths registered during 2011 to 2012, causes of death were available as text only because there is typically a 2-year delay in the publication of ICD coding due to the time taken to finalize the cause of death in cases referred to coroners. CVD deaths in 2011–12 were therefore classified by the authors in duplicate, with discrepancies agreed by consensus. The full list of ICD-10 AM codes used for fatal and non-fatal events is available on request. Other disease outcomes (e.g. chronic kidney disease) have been considered, but at present the main focus is on vascular outcomes.

Table 5 presents the mean and median follow-up times for the three sub-cohorts and the CVD events that occurred among these patient groups during follow-up. As it is an 'open' cohort, with ongoing additions to it and deletions from it due to death, the follow-up time ranges from 1 day to 10 years. Due to the rapid recruitment over 2010–12, mean and median follow-up times are relatively short. The CVD sub-cohort had a median follow-up time of 1.9 years (72 386 person-years) and mean follow-up of 2.4 years. Those in the Diabetes sub-cohort and the no CVD/no Diabetes sub-cohort had a median follow-up time of 2.7 years (116 751 person-years) and 2.4 years (459 223 person-years), respectively, and mean follow-up time was 2.9 years in both sub-cohorts.

What has been found? Key findings and publications

The main aim of developing this cohort was to validate the Framingham CVD score and generate new CVD risk equations, but the unique approach to data collection via a clinical tool has allowed us to gain a better understanding of

Table 4. No Diabetes/No CVD sub-cohort: baseline medicalhistory and CVD risk factors, August 2002 to July 2012

Variable	Total	Age group, *n*							
	n = 191 343	< 25	25–34	35–44	45–54	55–64	65–74	75–84	85+
	n (%)	284	2663	26 283	63 027	59 541	30 592	7625	1328
Medical history:									
Family history CVD	23 258 (12.2)	16.2	18.9	13.7	12.9	12.4	9.8	7.1	4.3
[a]Atrial fibrillation (AF)	4036 (2.1)	0.0	0.3	0.5	1.0	1.9	4.1	8.8	16.0
Heart failure	3011 (1.6)	0.0	0.6	0.5	0.7	1.2	2.8	7.6	18.4
Risk factor:									
[b]Smoking									
Yes (current)	27 191 (14.2)	26.8	22.6	21.5	18.0	11.6	7.4	4.1	1.4
Past (former)	28 875 (15.1)	6.3	10.4	10.3	13.4	16.7	19.3	18.3	16.2
No (never)	135 275 (70.7)	66.9	67.1	68.2	68.6	71.7	73.3	77.6	82.4
[c]Body mass Index (BMI)									
Underweight/normal (< 25)	31 265 (16.3)	17.3	10.2	11.4	14.7	17.8	19.0	23.4	32.7
Overweight (25–29.9)	46 133 (24.1)	16.2	18.9	23.2	24.4	24.2	24.9	24.0	20.0
Obesity (30+)	44 393 (23.2)	26.4	32.2	33.7	26.0	20.1	16.8	12.7	8.0
		Mean (SD)							
[d]Systolic blood pressure (mm Hg)	130 (17)	123 (16)	125 (16)	126 (16)	128 (17)	132 (17)	136 (17)	138 (18)	139 (19)
[e]Diastolic blood pressure (mm Hg)	80 (11)	76 (11)	80 (11)	81 (11)	81 (11)	80 (10)	79 (10)	77 (10)	76 (10)
[f]Total cholesterol: HDL ratio	4.1 (1.3)	4.4 (1.6)	4.9 (1.6)	4.6 (1.4)	4.2 (1.3)	3.9 (1.2)	3.8 (1.1)	3.6 (1.1)	3.4 (1.0)
Medications at baseline									
Aspirin	17 012 (8.9)	1.1	0.8	2.1	4.5	9.6	17.8	26.9	31.3
Clopidogrel	81 (0.04)	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.2
Warfarin	2132 (1.1)	0.0	0.4	0.3	0.5	1.0	2.2	5.1	6.5
BP-lowering drugs	48 697 (25.5)	2.8	7.3	9.5	16.9	28.5	42.8	58.2	65.7
Statin	26 015 (13.6)	1.4	3.6	6.0	9.2	16.1	22.7	24.2	15.7
Other lipid-lowering drugs	2155 (1.1)	0.4	0.2	0.5	0.7	1.2	2.1	2.7	1.5

[a]AF not assessed for 20 208 patients (10.6%) in initial PREDICT online form 2003–05.
[b]Smoking missing data values for 2 patients.
[c]BMI missing data values for 69 552 patients (36.4%).
[d]Systolic blood pressure missing data values for 1 patient.
[e]Diastolic blood pressure missing data values for 4 patients.
[f]Total cholesterol: HDL ratio missing data values for 14 patients.

the acceptability and impact of computerized decision support in primary care, data reliability and variations in risk factor profiles between ethnicities.

This is the 19th paper arising from the cohort—hence the title PREDICT-CVD 19. We have published on the development and adoption and impact of the PREDICT computerized decision support system on CVD risk assessment practice.[14–17] Using PREDICT software has led to improvements in documentation and classification of risk, risk factors and medical history in general practice.[18] Audits of data captured by PREDICT were also found to be strongly consistent with data held in the primary care EHRs.[18]

Good agreement was found between ethnicity coding in GP records and patient self-identified ethnicity.[19] However, agreement between GP ethnicity coding and ethnicity coding data held centrally on other national databases was found to be not as good, indicating the need for caution in the interpretation of ethnic-specific findings as different categorizations of ethnicity data from routine (and other) databases can lead to different ethnic-specific estimates of epidemiological effects.[20]

When the PREDICT software was first implemented, CVD risk assessments increased from 3% to 11% over 1 year.[16] By mid 2012, CVD risk assessment had increased to almost 50% of the eligible total New Zealand adult population, according to national guideline age, sex and ethnicity criteria. This was due to widespread implementation of several other CVD decision support tools and the introduction of government health targets accompanied by modest financial incentives.[21] Currently the primary health organizations contributing to the PREDICT cohort study have risk assessed between 79% and 88% of their enrolled eligible patients.[22]

Comparative CVD risk factor levels of Māori,[23] Pacific[24,25] and Indian[26] participants in the PREDICT-CVD cohort have observed markedly different risk factor

Table 5. Follow-up time and occurrence of CVD events among patients in the PREDICT cohort, August 2002 to July 2012

	Total	Age group, *n*							
Variable	*n* (%)	< 25	25–34	35–44	45–54	55–64	65–74	75–84	85+
	Patients with a history of CVD								
	38 098	**34**	**96**	**1095**	**4888**	**10 065**	**12 246**	**7546**	**2128**
Follow-up time (years)									
Median	1.9	1.6	3.2	2.9	2.5	2.2	1.9	1.6	1.3
Mean	2.4	2.1	3.4	3.2	2.9	2.7	2.4	2.1	1.6
CVD event									
No	29 065 (76.3)	100.0	84.4	83.7	82.5	80.3	76.5	68.9	63.9
Yes	9033 (23.7)	0.0	15.6	16.4	17.5	19.7	23.5	31.1	36.1
	Patients with diabetes but no CVD history								
	43 241	**371**	**1438**	**5585**	**11 568**	**12 383**	**8141**	**3177**	**578**
Follow-up time (years)									
Median	2.7	1.6	2.1	2.7	3.0	2.9	2.6	2.3	1.6
Mean	2.9	1.9	2.4	2.9	3.1	3.0	2.8	2.5	1.9
CVD event									
No	40 320 (93.2)	99.7	99.0	97.0	95.1	93.2	90.5	85.9	80.1
Yes	2921 (6.8)	0.3	1.0	3.0	4.9	6.8	9.5	14.1	19.9
	Patients without diabetes or a CVD history								
	19 1343	**284**	**2663**	**26 283**	**63 027**	**59 541**	**30 592**	**7625**	**1328**
Follow-up time (years)									
Median	2.4	3.6	3.6	3.2	2.5	2.2	2.1	2.2	1.5
Mean	2.9	4.1	4.1	3.4	3.0	2.8	2.6	2.6	2.0
CVD event									
No	18 5527 (97.0)	99.7	99.5	98.7	98.2	97.1	95.2	88.8	79.4
Yes	5816 (3.0)	0.4	0.5	1.3	1.9	2.9	4.8	11.2	20.6

clustering compared with European levels. Other studies on medication management and maintenance[27,28] in the PREDICT cohort have demonstrated that initiation of treatment is strongly associated with predicted CVD risk, but varies considerably by age.[28] Of the participants with previous CVD, overall 81% were dispensed blood pressure medications and 73% lipid-lowering medications, and 67% were receiving both.[27] However, although there were minimal differences in the likelihood of dispensing by sex, ethnicity or deprivation, patients aged 35–44 years were 30–40% less likely to receive these medications than those aged 65–75 years.[27] In addition, most patients already prescribed CVD medications at the time of their baseline CVD risk assessment remained on this prescribed treatment during follow-up (up to 3 years), irrespective of their estimated baseline risk.[28] Quality improvement studies are under way, investigating primary care laboratory monitoring behaviour and the extent of under-classification of CVD history in primary care records.

We have also investigated the performance of Framingham risk estimation algorithms (and other CVD risk algorithms) for those with type 2 diabetes[29] or a history of CVD,[30] and for high-risk ethnicity groups.[31] For primary prevention populations, the Framingham Heart Study CVD algorithm overestimated risk for the New Zealand European population but underestimated risk for the combined high-risk ethnic populations (Māori, Pacific and Indian).[31] We are also investigating the independent effects of a range of other risk predictors. In one study, a two standard deviation difference in serum urate (0.45 versus 0.27 mmol/l) was associated with a hazard ratio (HR) of 1.56 (95% CI: 1.32 to 1.84) for incident CVD events.[32]

What are the main strengths and weaknesses?

Entry to the cohort is at the discretion of the patient's doctor or nurse. Therefore, some recruitment bias is likely as risk assessment was initially prioritized in high-risk patients. However, the cohort is becoming increasingly representative of the source population as coverage of the PREDICT system grows. This paper presents data on the cohort recruitment up to mid 2012 when it included about 50% of guideline-eligible patients in the practices using PREDICT software. Data up to 2014 (not shown) indicate that this is now between 79% and 88% of eligible patients.[22]

A major strength is the integration of PREDICT software with patient EHRs, which allows robust mapping of clinical variables directly from the EHR into the online form. The PREDICT form has a number of compulsory fields required to calculate risk. This has facilitated nearly complete (99%) risk factor data collection for key variables. Built-in range and validity checks at the point of data entry have reduced input errors. Whereas these are strengths of the cohort, the need to limit respondent burden / clinical workload has constrained the number of additional variables that could be measured. As a result 15–36% of BMI values are missing, depending on the sub-cohort because height and weight were not compulsory variables for CVD risk assessment. Many other variables have been associated with CVD prognosis (e.g. urinary albumin to creatinine ratio, estimated glomerular filtration rate, serum uric acid and inflammatory markers). We plan to further augment PREDICT variables by linkage to the regional laboratory repository.

Encrypted NHI linkage to national health datasets enables almost complete ascertainment of CVD events, as more than 95% of patients with an acute CVD event in New Zealand are managed by public health services. However, participants who have CVD outcome events outside New Zealand will be missed unless these events are subsequently documented in primary or secondary care records. Participants who emigrate are also lost to follow-up. Outcomes are determined using ICD-coded hospitalizations and deaths reported to the New Zealand Ministry of Health and are therefore subject to weaknesses of health data reported for administrative purposes.[33]

A strength of the study is that the cohort has been derived directly from routine practice, the setting where new risk algorithms will be implemented. New CVD risk prediction algorithms for each of the three sub-cohorts are being prepared for publication, and we also plan to develop separate risk prediction algorithms for Māori, Pacific and Indian peoples. The national prioritization protocol enables us to generate a single ethnicity classification across multiple databases. For example, if a patient self-identifies as Māori in any of the linked databases, they will be classified as Māori. Unfortunately a weakness of the national ethnicity coding system is that it only allows accurate identification of Indian patients and not other South Asian ethnicities at high CVD risk (e.g. Pakistani, Bangladeshi, Sri Lankan). This means that the latter high-risk ethnicities are aggregated within the NZEO group.

Can I get hold of the data? Where can I find out more?

Expressions of interest for international collaborative research are welcomed. Although the cohort and system of data linkage have been developed for CVD research, the cohort has the potential to enable research on any other health outcomes that are recorded by the New Zealand health services. Proposals would be expected to involve researchers from the University of Auckland PREDICT research steering group and would be subject to scrutiny by Māori, Pacific and South Asian governance groups to ensure congruence with equity research goals. Applications will only be granted and data provided after agreement from our contributing providers and the Ministry of Health and after ethical approval by the New Zealand Mult-region Ethics Committee. For further enquiries, please contact the corresponding author, Susan Wells [s.wells@auckland.ac.nz].

Profile in a nutshell

- The PREDICT-CVD is an open primary care cohort study set up to facilitate cardiovascular disease research and to assist CVD risk assessment and management in routine general practice.
- PREDICT-CVD participants are automatically recruited when primary care practitioners in New Zealand use PREDICT, a web-based computerized decision support system that provides a CVD risk assessment and individualized CVD management advice based on national guidelines.
- By July 2012, the PREDICT cohort included: 272 682 people aged over 18 years; 38 098 with a history of CVD (the CVD sub-cohort); 43 241 with diabetes but no history of CVD (Diabetes sub-cohort); and 191 343 with no history of either CVD or diabetes (no CVD/no Diabetes sub-cohort).
- Mean follow-up to 2012 varied in these three sub-cohorts between 2.4 and 2.9 years (72 386 to 459 223 person-years) with approximately half of the participants recruited between 2010 and 2012. During follow-up: 17 770 people had an incident CVD event; 9033 (51%) were identified in the CVD sub-cohort; 2921 (16%) events in the Diabetes cohort; and 5816 (33%) events in the no CVD/no Diabetes sub-cohort.
- The cohort has near complete data for traditional CVD risk factors, ethnicity and socioeconomic characteristics and medication dispensing. Over 98% of New Zealanders have a unique personal health identifier (the NHI number) that records all publicly funded health service interactions. Robust outcome ascertainment is achieved by anonymized linkage to national and regional health databases.
- Expressions of interest for international collaborative research are welcomed.

Funding

The PREDICT CVD study has been supported by funding from the Health Research Council of New Zealand via two project grants (HRC 03/183 and 08/121) and under a programme grant (HRC VIEW 11/800 programme). S.W. has a Stevenson Fellowship in Health Innovation and Quality Improvement.

Acknowledgements

In memory of Tadd Clayton, our steadfast data manager whose diligent efforts will long be appreciated.

The authors would like to thank the primary health care organizations, affiliated general practitioners, practice nurses and patients for their contributions to this study.

The development of the PREDICT cohort is the result of a collaboration between epidemiologists and other clinical researchers at the University of Auckland, IT specialists at Enigma Solutions Ltd (a private IT company who developed and maintain the PREDICT software and webserver), primary health care organizations (and their member GPs), non-governmental organizations (New Zealand Guidelines Group, National Heart Foundation, Diabetes New Zealand, Diabetes Auckland), several district health boards and the Ministry of Health. The PREDICT software platform is owned by Enigma Publishing (PREDICT is a trademark of Enigma Solutions Ltd).

Conflict of interest: Outside this study, S.W. has received a research grant from Roche Diagnostics Ltd and TK is an investigator on an investigator-led trial partially sponsored by Nestec. All other authors declare no conflicts of interest.

References

1. Anderson KM, Odell PM, Wilson PWF *et al*. Cardiovascular disease risk profiles. *Am Heart J* 1991;**121:**293–98.
2. 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.
3. New Zealand Guideline Group. *The Assessment and Management of Cardiovascular Risk*. Wellington: Zealand Guidelines Group, 2003.
4. New Zealand Guideline Group. *The Management of Type 2 Diabetes*. Wellington: New Zealand Guideline Group, 2003.
5. Ministry of Health. *National Health Index Data Dictionary* (version 5.3). Wellington: Ministry of Health, 2009.
6. Ministry of Health. *National Health Index*. 2015. http://www.health.govt.nz/our-work/health-identity/national-health-index (5 February 2015, date last accessed).
7. Statistics New Zealand. *Subnational Population Estimates Tables*. Wellington: Statistics New Zealand, 2014.
8. Chan WC, Wright C, Tobias M, Mann S, Jackson R. Explaining trends in coronary heart disease hospitalizations in New Zealand: trend for admissions and incidence can be in opposite directions. *Heart* 2008;**94:**1589–93.
9. Ministry of Health. *Pharmaceutical Collection*. 2015. http://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections/pharmaceutical-collection (5 February 2015, date last accessed).
10. Ministry of Health. *Ethnicity Data Protocols for the Health and Disability Sector*. Wellington: Ministry of Health, 2004.
11. Salmond C, Crampton P, Atkinson J. *NZDep 2006 Index of Deprivation User's Manual*. 2002. Wellington: Department of Public Health, Wellington School of Medicine and Health Sciences, 2007.
12. Demant MN, Gislason GH, Kober L, Vaag A, Torp-Pedersen C, Andersson C. Association of heart failure severity with risk of diabetes: a Danish nationwide cohort study. *Diabetologia* 2014;**57:**1595–600.
13. Andersson C, Norgaard ML, Hansen PR *et al*. Heart failure severity, as determined by loop diuretic dosages, predicts the risk of developing diabetes after myocardial infarction: a nationwide cohort study. *Eur J Heart Fail* 2010;**12:**1333–38.
14. Warren J, Gaikwad R, Mabotuwana T *et al*. The challenge of evaluating electronic decision support in the community. *Health Care and Informatics Review Online* 2008;**10**. www.hinz.org.nz/?page=HCIROJournal
15. Wells S, Bycroft J, Lee A-W *et al*. Patterns of adoption and use of a web-based decision support system in primary care. *Health Care and Informatics Review Online* 2007;**3**. www.hinz.org.nz/?page=HCIROJournal
16. Wells S, Furness S, Rafter N *et al*. Integrated electronic decision support increases cardiovascular disease risk assessment four fold in routine primary care practice. *Eur J Cardiovasc Prev Rehaband* 2008;**15:**173–78.
17. Wells S, Jackson R. Online management of cardiovascular risk in New Zealand with PREDICT – Getting evidence to the 'moment of care'. *Health Care and Informatics Review Online* 2005;**March:**1–6. www.hinz.org.nz/?page=HCIROJournal
18. Riddell T, Kenealy T, Wells S, Jackson R, Broad J. Audit of health data captured routinely in primary healthcare for the Clinical Decision Support System PREDICT (PREDICT CVD-4). *Health Care and Informatics Review Online* 2008;**March.** www.hinz.org.nz/?page=HCIROJournal
19. Riddell T, Lindsay G, Kenealy T *et al*. The accuracy of ethnicity data in primary care and its impact on cardiovascular risk assessment and management – PREDICT CVD-8. *N Z Med J* 2008;**121:**40–48.
20. Marshall RJ, Zhang Z, Broad JB, Wells S. Agreement between ethnicity recorded in two New Zealand health databases: effects of discordance on cardiovascular outcome measures (PREDICT CVD3). *Aust N Z J Public Health* 2007;**31:**211–16.
21. Ministry of Health. *Health Targets: More Heart and Diabetes Checks*. 2014. http://www.health.govt.nz/new-zealand-health-system/health-targets/about-health-targets/health-targets-more-heart-and-diabetes-checks (1 December 2014, date last accessed).
22. Ministry of Health. *How is My PHO Performing?* 2014/15, Quarter 3 (January to March) results. Wellington: Ministry of Health, 2015.
23. Riddell T, Jackson RT, Wells S, Broad J, Bannink L. Assessing Maori/non-Maori differences in cardiovascular disease risk and risk management in routine primary care practice using web-based clinical decision support: (PREDICT CVD-2).[see comment]. *N Z Med J* 2007;**120:**U2445.
24. Grey C, Wells S, Riddell T *et al*. A comparative analysis of the cardiovascular disease risk factor profiles of Pacific peoples and Europeans living in New Zealand assessed in routine primary care: PREDICT CVD-11. *N Z Med J* 2010;**123:**62–75.

25. Grey C, Wells S, Riddell T *et al*. A comparative analysis of cardiovascular disease risk profiles of five Pacific ethnic groups assessed in New Zealand primary care practice: PREDICT CVD-13. *N Z Med J* 2010;**123:**41–52.
26. Perumal L, Wells S, Ameratunga S *et al*. Cardiovascular disease risk factors and estimated risk profiles of Indian participants in a New Zealand primary care cohort (PREDICT CVD-14). *Aust N Z J Public Health* 2011. online doi:10.1111/j.1753-6405.2011.00808.x.
27. Mehta S, Wells S, Riddell T *et al*. Under-utilisation of preventive medication in patients with cardiovascular disease is greatest in younger age groups (PREDICT-CVD 15). *J Primary Health Care* 2011;**3:**93–101.
28. Mehta S, Wells S, Grey C *et al*. Initiation and maintenance of cardiovascular medications following cardiovascular risk assessment in a large primary care cohort: PREDICT CVD-16. *Eur J Prev Cardiol* 2014;**21:**192–202.
29. Metcalf PA, Wells S, Scragg RKR, Jackson R. Comparison of three different methods of assessing cardiovascular disease risk in New Zealanders with Type 2 diabetes mellitus. *N Z Med J* 2008;**121:**49–57.
30. Kerr AJ, Broad J, Wells S, Riddell T, Jackson R. Should the first priority in cardiovascular risk management be those with prior cardiovascular disease? *Heart* 2009;**95:**125–29.
31. Riddell T, Wells S, Jackson R *et al*. Performance of Framingham cardiovascular risk scores by ethnic groups in New Zealand: PREDICT CVD-10. *N Z Med J* 2010;**123:**50–61.
32. Thornley S, Marshall RJ, Jackson R *et al*. Is serum urate causally associated with incident cardiovascular disease? *Rheumatology* 2013;**52:**135–42.
33. Smith G, Noble M, Anttilla C *et al*. *The Value of Linked Administrative Records for Longitudinal Analysis, Report to the ESRC National Longitudinal Strategy Committee*. Oxford, UK: Social Disadvantage Research Centre, University of Oxford, 2004.