Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study

Romana Pylypchuk, Sue Wells, Andrew Kerr, Katrina Poppe, Tania Riddell, Matire Harwood, Dan Exeter, Suneela Mehta, Corina Grey, Billy P Wu, Patricia Metcalf, Jim Warren, Jeff Harrison, Roger Marshall, Rod Jackson

Summary

Background Most cardiovascular disease risk prediction equations in use today were derived from cohorts established last century and with participants at higher risk but less socioeconomically and ethnically diverse than patients they are now applied to. We recruited a nationally representative cohort in New Zealand to develop equations relevant to patients in contemporary primary care and compared the performance of these new equations to equations that are recommended in the USA.

Methods The PREDICT study automatically recruits participants in routine primary care when general practitioners in New Zealand use PREDICT software to assess their patients' risk profiles for cardiovascular disease, which are prospectively linked to national ICD-coded hospitalisation and mortality databases. The study population included male and female patients in primary care who had no prior cardiovascular disease, renal disease, or congestive heart failure. New equations predicting total cardiovascular disease risk were developed using Cox regression models, which included clinical predictors plus an area-based deprivation index and self-identified ethnicity. Calibration and discrimination performance of the equations were assessed and compared with 2013 American College of Cardiology/American Heart Association Pooled Cohort Equations (PCEs). The additional predictors included in new PREDICT equations were also appended to the PCEs to determine whether they were independent predictors in the equations from the USA.

Findings Outcome events were derived for 401 752 people aged 30–74 years at the time of their first PREDICT risk assessment between Aug 27, 2002, and Oct 12, 2015, representing about 90% of the eligible population. The mean follow-up was 4·2 years, and a third of participants were followed for 5 years or more. 15 386 (4%) people had cardiovascular disease events (1507 [10%] were fatal, and 8549 [56%] met the PCEs definition of hard atherosclerotic cardiovascular disease) during 1 685 521 person-years follow-up. The median 5-year risk of total cardiovascular disease events predicted by the new equations was 2·3% in women and 3·2% in men. Multivariable adjusted risk increased by about 10% per quintile of socioeconomic deprivation. Māori, Pacific, and Indian patients were at 13–48% higher risk of cardiovascular disease than Europeans, and Chinese or other Asians were at 25–33% lower risk of cardiovascular disease than Europeans. The PCEs overestimated of hard atherosclerotic cardiovascular disease by about 40% in men and by 60% in women, and the additional predictors in the new equations were also independent predictors in the PCEs. The new equations were significantly better than PCEs on all performance metrics.

Interpretation We constructed a large prospective cohort study representing typical patients in primary care in New Zealand who were recommended for cardiovascular disease risk assessment. Most patients are now at low risk of cardiovascular disease, which explains why the PCEs based mainly on old cohorts substantially overestimate risk. Although the PCEs and many other equations will need to be recalibrated to mitigate overtreatment of the healthy majority, they also need new predictors that include measures of socioeconomic deprivation and multiple ethnicities to identify vulnerable high-risk subpopulations that might otherwise be undertreated.

Funding Health Research Council of New Zealand, Heart Foundation of New Zealand, and Healthier Lives National Science Challenge.

Published **Online** May 4, 2018 http://dx.doi.org/10.1016/S0140-6736(18)30664-0

See **Online/Comment** http://dx.doi.org/10.1016/S0140-6736(18)30842-0

School of Population Health (R Pylypchuk MSc, S Wells PhD, A Kerr MD, K Poppe PhD, T Riddell MBChB, M Harwood PhD, D Exeter PhD, S Mehta MBChB, C Grey MBChB, B P Wu MPH, P Metcalf PhD, R Marshall PhD, Prof R Jackson PhD) **and School of Pharmacy** (J Harrison PhD)**, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand; Cardiology Department, Middlemore Hospital, Auckland, New Zealand** (A Kerr)**; and Department of Computer Science, Faculty of Science, University of Auckland, Auckland, New Zealand** (Prof J Warren PhD)

Correspondence to:
Prof Rod Jackson, Section of Epidemiology and Biostatistics, School of Population Health, Faculty of Medical and Health Sciences, University of Auckland, PO Box 92019, Auckland Mail Centre, Auckland, New Zealand
rt.jackson@auckland.ac.nz

Introduction

More than 40 years ago, Framingham Heart Study investigators developed multivariable cardiovascular disease risk prediction equations that identified high-risk patients much more accurately than traditional classifications based on blood pressure or blood cholesterol concentrations alone.[1] As the benefits of interventions that reduce the risk of cardiovascular disease are proportional to pretreatment risk,[2,3] treating patients who are assessed as high-risk with multivariable prediction equations is also more effective than treating patients with high levels of single risk factors. Most existing guidelines on cardiovascular disease risk factor management therefore recommend using risk prediction equations to inform

Research in context

Evidence before this study

In a 2016 systematic review of cardiovascular disease risk prediction models, 363 equations were identified, mainly from Europe and North America. The models had substantial variation in predictor and outcome definitions, and most models included only age, sex, smoking, diabetes, blood pressure, and blood lipids as predictors. More than 70 definitions of cardiovascular disease outcomes were reported, and the authors concluded that most prediction models are insufficiently reported to allow external validation by others, let alone be implemented. Moreover, models were largely derived in cohorts established last century, when cardiovascular disease event rates were more than double current rates and included participants who were less socioeconomically and ethnically diverse and less likely to be on preventive medications than the patients the models are applied to at present. Only the UK QRISK risk prediction equations are regularly updated in contemporary representative cohorts and include a comprehensive range of predictors, including deprivation measures, but they are complex and difficult to implement or validate outside UK general practice.

Added value of this study

We developed simple equations for predicting the 5-year risk of ICD-coded fatal cardiovascular disease and non-fatal cardiovascular disease hospitalisations that were designed to facilitate external validation and implementation. They were derived in a contemporary cohort of 401 752 New Zealanders aged 30–74 years without prior cardiovascular disease, congestive heart failure, or significant renal disease in the primary care setting where most risk assessments of cardiovascular disease are done. Aside from QRISK, we are unaware of any similar contemporary cohorts, yet such cohorts are necessary for developing accurate risk prediction equations. Median 5-year risk of cardiovascular disease was only 2·3% in women and 3·2% in men, highlighting the low risk in this typical high-income country population. This explains why the recommended 2013 American College of Cardiology/American Heart Association Pooled Cohort Equations (PCEs) were poorly calibrated in the PREDICT cohort, overestimating hard atherosclerotic cardiovascular disease events by up to 60%, although incidentally estimating total ischaemic cardiovascular disease hospitalisations and deaths reasonably well. Adding measures of socioeconomic status, ethnicity, and several other variables routinely available in clinical care to the PCEs would identify patient groups with predicted risk from about 25% lower to 65% higher than equations based on standard risk predictors. Moreover, the poor performance of the PCEs could not be explained by increasing use of preventive medications.

Implications of all the available evidence

Unless risk of cardiovascular disease is clearly defined and estimated using equations derived or recalibrated in contemporary populations that represent the patients they are applied to, substantial underestimation or overestimation of risk, and therefore substantial undertreatment or overtreatment, is likely. Furthermore, in the era of precision medicine, recalibrating old equations will be insufficient, and new predictors (including measures of socioeconomic deprivation and multiple ethnicities) that could be made routinely available in medical records should be included to avoid undertreatment of high-risk subpopulations. With increasing computerisation of medical practice, many countries or health-care organisations could replicate the PREDICT approach by linking primary care records to hospitalisations and deaths.

treatment decisions.[4–9] Although more than 360 cardiovascular disease risk equations have been published since the pioneering Framingham research,[10] most are based on cohort studies established last century. Participants in these older studies, including those used to derive the 2013 American College of Cardiology/American Heart Association Pooled Cohort Equations (PCEs)[7] that are recommended at present, are very different to the patient populations the equations are now applied to, and their applicability is uncertain.

In the 1990s, New Zealand developed the world's first national cardiovascular disease risk factor management guidelines based on multivariable predicted risk[11] and recommended using 1991 Framingham Heart Study prediction equations[12] to inform treatment decisions. At the time, no local cohort studies were available to validate the Framingham equations. In 2002, we developed a computerised decision support system that helped general practitioners implement the national guidelines while simultaneously generating a cohort study to investigate whether a 20th century Framingham equation was applicable to an ethnically and socioeconomically diverse New Zealand population in the 21st century. Here we describe the derivation and validation of new equations based on the Framingham equations that also include measures of deprivation, ethnicity, and other predictors of increased risk. For comparison, we externally validated the PCEs[7] that have replaced Framingham equations and are integral to current cholesterol and blood pressure management guidelines in the USA.[8,9]

Methods

Study design and participants

PREDICT is an ongoing, prospectively designed, open cohort study in New Zealand that automatically recruits participants when primary health-care practitioners complete standardised cardiovascular disease risk assessments using PREDICT decision support software.[13] When opened, the software attempts to auto-populate PREDICT risk factor templates from patient records. Clinicians must fill in any missing fields before a cardiovascular disease risk can be calculated and

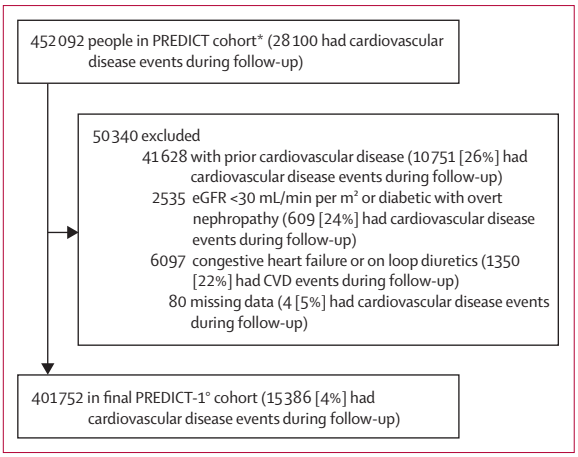

Figure 1: **PREDICT cohort enrolment, exclusions, and incidence of cardiovascular disease events during follow-up**
*Excludes 448 people with inconsistent sociodemographic variables across data sources and 7822 people in ethnic groups with fewer than 1000 participants (mainly Middle Eastern, Latin American, and African).

	Women	Men
Participants (percentage of total cohort)	175 699 (44%)	226 053 (56%)
Incident total cardiovascular disease events (percentage of sex-specific cohort)	5650 (3%)	9736 (4%)
Total person-years observed	743 640	941 881
Crude incidence of total cardiovascular disease events per 1000 per year (95% CI)	7·6 (7·4–7·8)	10·3 (10·1–10·5)
Mean follow-up time, years (SD)*	4·2 (2·7)	4·2 (2·7)
People with follow-up ≥5 years	58 493 (33%)	72 417 (32%)
Mean age, years (SD)	56 (8·9)	51·8 (9·9)
Self-identified ethnicity		
European	96 032 (55%)	128 503 (57%)
Māori	23 853 (14%)	27 573 (12%)
Pacific	22 537 (13%)	28 073 (12%)
Indian	14 188 (8%)	20 232 (9%)
Chinese or other Asian	19 089 (11%)	21 672 (10%)
NZDep quintile		
1 (least deprived)	38 523 (22%)	50 379 (22%)
2	34 230 (20%)	44 609 (20%)
3	31 808 (18%)	40 684 (18%)
4	32 626 (19%)	41 553 (18%)
5 (most deprived)	38 512 (22%)	48 828 (22%)
Smoking		
Never smoker	129 158 (74%)	149 139 (66%)
Ex-smoker	24 838 (14%)	39 856 (18%)
Current smoker	21 703 (12%)	37 058 (16%)
Family history of premature cardiovascular disease	22 996 (13%)	24 495 (11%)
Atrial fibrillation	1777 (1%)	3680 (2%)
Diabetes	27 377 (16%)	30 942 (14%)
Mean SBP, mm Hg (SD)	129 (17·7)	129 (16·2)
Mean TC/HDL (SD)	3·7 (1·1)	4·4 (1·3)
Medications at index assessment†		
Blood pressure-lowering medication	45 973 (26%)	43 253 (19%)
Lipid-lowering medication	27 540 (16%)	33 372 (15%)
Antithrombotic medication	17 831 (10%)	21 723 (10%)

Data are n (%) unless indicated otherwise. NZDep=New Zealand Index of Socioeconomic Deprivation. SBP=systolic blood pressure. TC/HDL=total cholesterol to HDL cholesterol ratio. *Follow-up time ranged from 1 day to 13·3 years in both men and women. †33% of women and 27% of men were treated with one or more class of drugs at index assessment.

Table 1: **Description of the PREDICT-1° cohort in women and men**

recruitment completed. Participant risk factor profiles captured by the software are regularly linked to national databases documenting drug dispensing and ICD-coded hospitalisations and deaths related to cardiovascular disease. The PREDICT study was approved by the Northern Region Ethics Committee Y in 2003 (AKY/03/12/314), with annual approval by the National Multi Region Ethics Committee since 2007 (MEC07/19/EXP).

The study included primary care patients who had cardiovascular disease risk assessments at primary health organisations that use PREDICT software. Approximately 95% of New Zealanders are enrolled in primary health organisations,[14] which provide most primary health care nationally. About a third of the country's population is served by clinics that use PREDICT software, mainly in the Auckland and Northland regions of New Zealand. These two regions include large urban and substantial rural populations, and New Zealand's diverse socioeconomic and ethnic groups are well represented in the study population. National guidelines recommend formal cardiovascular disease risk assessments every 5 years for men aged 45–74 years and for women aged 55–74 years, and assessments are recommended 10 years earlier for Māori, Pacific, and Indian subcontinent peoples and for people with known cardiovascular disease risk factors.[4] About 90% of all New Zealanders meeting these eligibility criteria were risk assessed between 2010 and 2015 as part of a nationally coordinated and funded programme.[15]

People with prior cardiovascular disease, renal disease, and congestive heart failure were excluded. These exclusions were based on a combination of diagnoses by general practitioners, hospital discharge records, and dispensing of anti-anginal drugs and loop diuretics (appendix p 2). Self-identified ethnicity is documented on all routine health records in New Zealand using a standard national classification system. Ethnic groups with fewer than 1000 participants were excluded. No participants had missing data on the mandatory variables required for the cardiovascular disease risk assessment using PREDICT software. The few participants with missing data on the New Zealand Index of Socioeconomic Deprivation (NZDep) were excluded.

Participants' risk factor profiles, measured at their index assessments, were linked to national health databases using encrypted national health identifiers. More than 99% of New Zealanders have a unique national health identifier that is attached to almost all interactions with publicly funded or subsidised health services and most

See **Online** for appendix

	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)
Indian	1·13 (1·00–1·27)	1·34 (1·24–1·45)
Chinese or other Asian	0·75 (0·66–0·85)	0·67 (0·61–0·74)
NZDep 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 cardiovascular disease	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 mm Hg*	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		
Taking blood pressure lowering medication	1·40 (1·31–1·50)	1·34 (1·27–1·42)
Taking lipid lowering medication	0·94 (0·88–1·01)	0·95 (0·90–1·00)
Taking antithrombotic medication	1·12 (1·04–1·21)	1·10 (1·03–1·17)
Interactions		
Age × diabetes	0·978 (0·972–0·984)	0·980 (0·977–0·984)
Age × SBP per 10 mm Hg	0·996 (0·994–0·997)	0·996 (0·995–0·997)
Taking blood pressure lowering medication × SBP per 10 mm Hg	0·958 (0·931–0·985)	0·948 (0·926–0·971)

Hazard ratios are adjusted for all other variables included in the model. NZDep=New Zealand Index of Socioeconomic Deprivation. SPB=systolic blood pressure. TC/HDL=total cholesterol to HDL cholesterol ratio. *The hazard ratios for SBP are per 10 mm Hg but were modelled per 1 mm Hg for absolute risk calculations.

Table 2: **Adjusted hazard ratios for total CVD in the PREDICT-1° equations**

private hospital services.[16] National health databases include all public hospitalisations, deaths, and subsidised drugs dispensed by community pharmacies. All common cardiovascular disease preventive drugs are publicly subsidised.

Outcomes

The primary PREDICT outcome was prespecified using the total cardiovascular disease outcome in 1991 Framingham equations,[12] defined by ICD-10-AM codes as a hospitalisation or death from: ischaemic heart disease (including angina); ischaemic or haemorrhagic cerebrovascular events (including transient ischaemic attacks); or peripheral vascular disease, congestive heart failure, or other ischaemic cardiovascular disease deaths (appendix p 3). An event was defined as fatal if the person died of cardiovascular disease without being admitted to hospital or died within 28 days of their first cardiovascular disease-related hospital admission.

Statistical analysis

The variables included in the new PREDICT models were all prespecified (appendix p 4). These included the variables required for calculating cardiovascular disease risk with the modified 1991 Framingham equations used in PREDICT software (ie, sex, age, self-identified ethnicity, family history of premature cardiovascular disease, smoking status, diabetes status, systolic blood pressure, and the ratio of total cholesterol to high density lipoprotein cholesterol concentrations [TC/HDL]). Additionally, NZDep,[17] atrial fibrillation confirmed by electrocardiograph (ECG), and use of blood pressure-lowering, lipid-lowering, and antithrombotic drugs in the 6 months before the index assessment were included.

Cox proportional hazards modelling[18] was used to develop new prediction equations for time to a first hospital admission or death related to cardiovascular disease, using all pre-specified variables (appendix). Time on study was the time from index assessment to the first of the following: hospital admission or death related to cardiovascular disease, death from other causes, or end of follow-up. Sex-specific analyses were undertaken. Reference groups for categorical variables are highlighted in the appendix. NZDep was initially modelled as a five-level categorical variable but was treated as a continuous variable in the final equations because risk increased monotonically with increasing deprivation. Model diagnostics included testing the proportionality assumption with the global Schoenfeld test[19] and plotting log(–log[survival]) versus log(time). Checks were also made for influential observations using delta beta (DFBETA) plots.[20] Linearity was assessed by visual inspection of LOWESS smoothed plots of Martingale residuals versus continuous covariates.[21] Non-linearity of continuous variables and first-order interactions between continuous and categorical variables were assessed using fractional polynomials.[22] Interaction terms were included if they met a strict predetermined threshold statistical significance of $p<0·001$ and were clinically plausible and if the plotted data suggested effect modification.

We used Stata 13.0 software for all analyses.[23]

Performance and internal validation of new PREDICT absolute risk prediction equations

Separate models were built for men and women. Continuous predictors in the models were centred at their mean values. The 5-year baseline survival probabilities of each model were obtained by the smoothed kernel estimator feature of the Stata stcox command that was used to fit the models.[23]

Calibration performance was assessed graphically by categorising participants into deciles of predicted 5-year cardiovascular disease risk and plotting mean 5-year predicted risk against observed 5-year risk. A diagonal line with slope of 1 represents perfect calibration. Observed 5-year risk was obtained by the Kaplan-Meier method,[24] and the slopes of regression lines comparing deciles of predicted versus observed 5-year risk were

calculated. Standard statistical metrics of model and discrimination performance (R^2, Harrell's C statistic, and Royston's D statistic)[25–27] were calculated.

The whole cohort was used to develop new equations, as recommended by Steyerberg,[28] and a split sample internal validation was done as a sensitivity analysis. Following recommendations from the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Initiative,[29] the cohort was split into two geographically defined subcohorts rather than randomly. The calibration and discrimination performance of equations developed in the derivation subcohort was assessed in the validation subcohort and compared with the performance of models developed in the whole cohort; baseline survival functions and hazard ratios were also compared.

External validation of the 2013 American College of Cardiology/American Heart Association PCEs

The same calibration, model, and discrimination performance measures described above for assessing the new PREDICT equations were also used in the external validation of the PCEs in the PREDICT cohort. We used the PCEs' 5-year Whites-only equations[30] to predict the PCEs' hard atherosclerotic cardiovascular disease outcome (ie, non-fatal myocardial infarction, death from coronary heart disease, and fatal and non-fatal stroke; appendix).[7] Calibration plots were drawn using both the original PCE models and models recalibrated to the PREDICT cohort. To recalibrate the PCE models, we updated the baseline survival values estimated by fitting Cox models with the prognostic index from the PCE model (offset term) in the PREDICT dataset.[31] Then, to determine whether the additional variables available in PREDICT were also independent predictors, over and above the PCE predictors, we built Cox models with the sex-specific prognostic indices[32] from the PCEs plus the additional variables available in PREDICT (ie, ethnicity, NZDep, family history of premature cardiovascular disease, personal history of atrial fibrillation, lipid-lowering drug treatment, antithrombotic drug treatment).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. RP, KP, and RJ had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The study population included 452 092 men and women aged 30–74 years at the time of their first PREDICT risk assessment (index assessment) between Aug 27, 2002, and Oct 12, 2015. More than half of participants were recruited after Dec 31, 2010 (figure 1). We excluded 50 260 people with prior cardiovascular disease, impaired renal function, or heart failure and 80 people with missing risk factor data. The remaining 401 752 people constituted the PREDICT-1° cohort used in these analyses. The cohort included about 90% of people eligible for cardiovascular disease risk assessments[4] in primary care practices using PREDICT software. 15 386 (4%) participants had their first major cardiovascular disease event during 1 685 521 person-years of follow-up. Mean follow-up was 4·2 years, and a third of participants were followed for 5 years or more. Participant characteristics are described in table 1.

Outcome events were derived exclusively from national mortality and public hospitalisation databases between Aug 27, 2002, and Dec 31, 2015. Non-fatal myocardial infarction was the most common outcome (4984 [34%] events), and 8237 (54%) of the total 15 386 cardiovascular disease events were coronary-related outcomes (appendix p 6). 4053 (26%) events were strokes and transient ischaemic attacks, 1908 (12%) events were congestive heart failure, and 852 (6%) events were peripheral vascular disease. Only

	Women	Men
Age	0·0756412	0·0675532
Māori	0·3910183	0·2899054
Pacific	0·2010224	0·1774195
Indian	0·1183427	0·2902049
Chinese or other Asian	–0·28551	–0·3975687
NZDep quintile	0·1080795	0·0794903
Ex-smoker	0·087476	0·0753246
Smoker	0·6226384	0·5058041
Family history of cardiovascular disease	0·0445534	0·1326587
Atrial fibrillation	0·8927126	0·5880131
Diabetes	0·5447632	0·5597023
SBP per 1 mm Hg	0·0136606	0·0163778
TC/HDL	0·1226753	0·1283758
OBPLM*	0·339925	0·2947634
OLLM	–0·0593798	–0·0537314
OATM	0·1172496	0·0934141
Age × diabetes	–0·0222549	–0·020235
Age × SBP	–0·0004425	–0·0004184
OBPLM × SBP	–0·004313	–0·0053077
Means for centering		
Age	56·13665	51·79953
NZDep quintle	2·990826	2·972793
SBP	129·0173	129·1095
TC/HDL	3·726268	4·38906
Baseline survival function (at 5 years)	0·983169213058	0·974755526232

NZDep=New Zealand Index of Socioeconomic Deprivation. SBP=systolic blood pressure. TC/HDL=total cholesterol to HDL cholesterol ratio. OBPLM=on blood pressure-lowering medications. OLLM=on lipid-lowering medications. OATM=on antithrombotic medications. *Denotes centred variables.

Table 3: **Beta coefficients in the PREDICT-1° equations**

***Panel*: Example calculation of 5-year risk of total cardiovascular disease**

Patient description

The patient is a European woman, aged 55 years, with diabetes. She is an ex-smoker, has no family history of cardiovascular disease or atrial fibrillation, and is rated as NZDep quintile 3. Her systolic blood pressure (SBP) is 135 mm Hg, and her ratio of total cholesterol to HDL cholesterol (TC/HDL) is 5 units. She is taking blood pressure-lowering medications (OBPLM) but not lipid-lowering medications or antithrombotic medications.

Beta coefficient × variable

- Age: 0·0756412 × c.Age*=–0·08597757
- NZDep quintile: 0·1080795 × c.NZDep*=0·00099152
- Ex-smoker: 0·087476 × 1 (Ex-smoker)=0·087476
- Diabetes: 0·5447632 × 1 (Diabetes)=0·5447632
- SBP: 0·0136606 × c.SBP*=0·08172727
- TC/HDL: 0·1226753 × c.TC/HDL*=0·15625546
- OBPLM : 0·339925 × 1 (OBPLM)=0·339925
- Age × diabetes: –0·0222549 × c.Age × 1(diab)=0·02529603
- Age × SBP: –0·0004425 × c.Age × c.SBP=0·0030091
- OBPLM × SBP: –0·004313 × 1(OBPLM) × c.SBP=–0·02580339

Sum coefficients × variables=1·1276626

Centred variables used in calculations of beta coefficient × variable (marked with asterisks above)

- c.Age: 55–56·13665=–1·13665
- c.NZDep: 3–2·990826=0·009174
- c.SBP: 135–129·0173=5·9827
- c.TC/HDL: 5–3·726268=1·273732

5-year risk of cardiovascular disease

- $(1\text{-baseline surv}^{\exp(\text{sum of coefficients} \times \text{variables})}) \times 100 = (1 - 0{\cdot}983169213058^{\exp(1{\cdot}1276626)}) \times 100 = 5{\cdot}11\%$

1507 (10%) events were fatal, and 556 (37%) fatal events were in people who had never been admitted to hospital with cardiovascular disease. The remainder of fatal events were deaths within 28 days of a hospital admission because of cardiovascular disease. 8549 (56%) PREDICT-defined total cardiovascular disease events met the PCEs definition of hard atherosclerotic cardiovascular disease.[7]

In the new PREDICT-1° equations, all continuous variables were fitted as linear terms after assessment using the fractional polynomials procedure,[22] and Martingale residuals plots[21] provided no compelling support for fitting non-linear terms. Adjusted hazard ratios for total cardiovascular disease in the PREDICT-1° equations were calculated for women and men (table 2). Each additional year of age was associated with an increased estimated 5-year cardiovascular disease risk of 7–8% in relative terms. Māori, Pacific, and Indian peoples were all at increased risk compared with Europeans, whereas Chinese and other Asian peoples were at lower risk than Europeans. Risk increased in women and men per quintile of the socioeconomic deprivation index, and family history of premature cardiovascular disease was a statistically significant predictor in men only. Smoking, diabetes, atrial fibrillation, increased systolic blood pressure, and increased TC/HDL were all statistically significant predictors, as were use of blood pressure-lowering and antithrombotic medications at the index assessment (but not use of lipid-lowering medications). Interactions between diabetes and age, between systolic blood pressure and age, and between use of blood pressure-lowering drugs and systolic blood pressure were statistically significant in both sexes.

Regression coefficients, means of centred variables, and baseline survival functions for the sex-specific 5-year cardiovascular disease risk PREDICT-1° equations are presented in table 3, and an example risk calculation is shown in the panel. The mean estimated 5-year risk of total cardiovascular disease 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% (1·8–6·0%) in men.

Predicted versus observed 5-year risk plots for total cardiovascular disease using the PREDICT-1° equations showed excellent calibration across all risk deciles in both sexes (figure 2A–B). The slopes of regression lines comparing predicted and observed total cardiovascular disease risk in deciles were 0·98 (95% CI 0·93–1·02) for women and 0·98 (0·98–1·01) for men. Under-prediction or overprediction did not exceed 0·5% in any predicted risk decile. By contrast, the original PCEs significantly overpredicted observed 5-year risk of hard atherosclerotic cardiovascular disease in the top seven deciles of predicted risk in both men and women (figure 2C–D). The slopes of regression lines comparing deciles of predicted and observed 5-year risk of hard atherosclerotic cardiovascular disease were 1·79 (95% CI 1·58–2·0) for women and 1·56 (1·38–1·75) for men, and on average the PCEs overestimated risk by 62% in women and by 41% in men. After recalibration, the PCEs still overestimated risk of hard atherosclerotic cardiovascular disease but only in the top three deciles for women and top two deciles for men (figure 2E–F). The slopes of regression lines were 1·35 (95% CI 1·19–1·52) for women and 1·21 (1·06–1·36) for men.

Model and discrimination metrics indicated that the PREDICT-1° equations performed better in predicting total cardiovascular disease events than the PCEs performed in predicting hard atherosclerotic cardiovascular disease events, and the differences were statistically significant for all comparisons (table 4). Hazard ratios in PREDICT-1° models developed in the derivation subcohort sensitivity analyses were similar to the models developed in the full cohort, and their model and discrimination performance, when tested in the validation subcohort, were also similar (appendix p 7).

The adjusted hazard ratios for the additional variables available in PREDICT, when added to the PCEs models, are shown in table 5. Ethnicity, socioeconomic deprivation, family history of premature cardiovascular disease, atrial fibrillation, and lipid-lowering or antithrombotic medications were all statistically significant predictors of cardiovascular disease risk in either men or women or both.

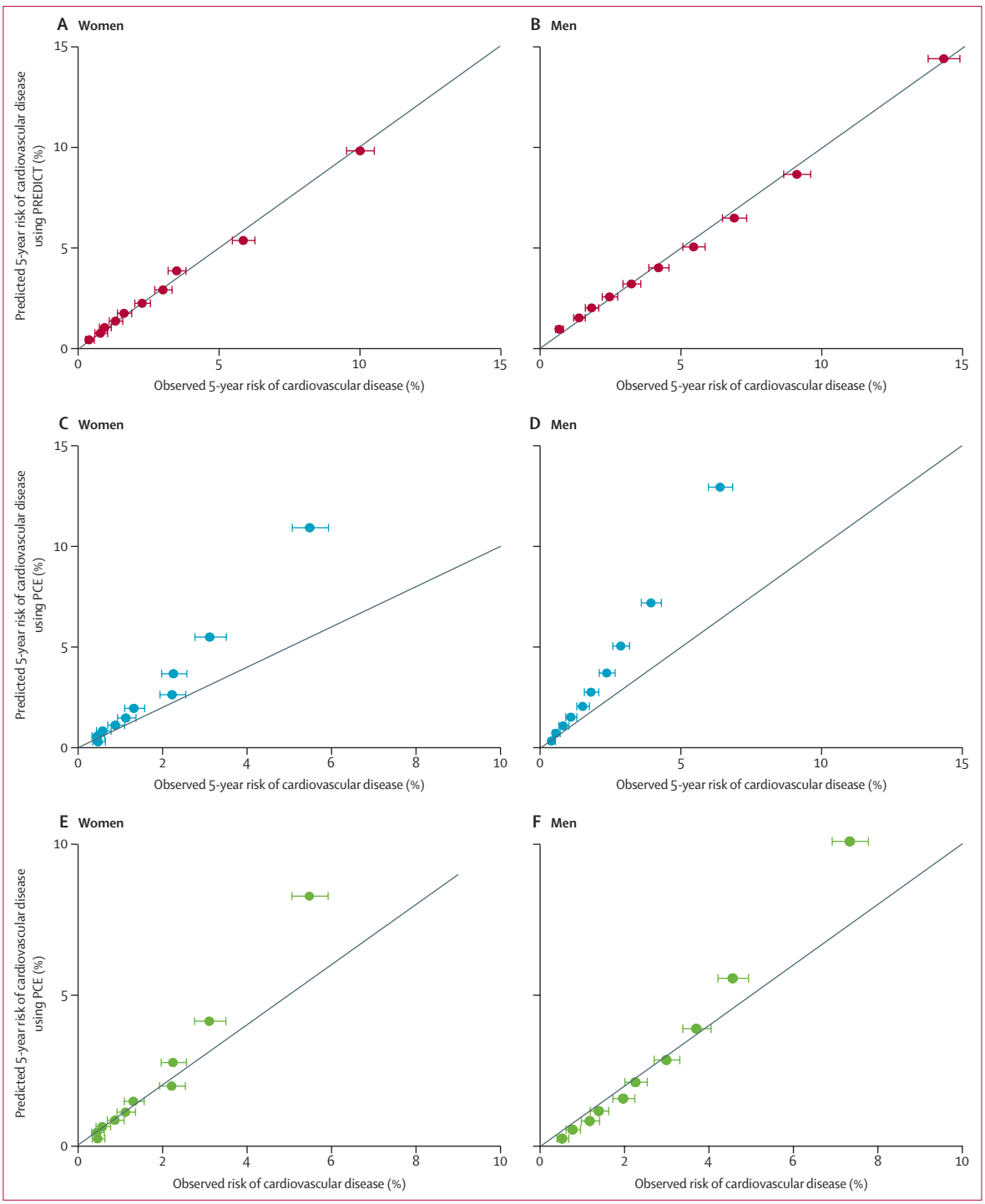

Figure 2: **Calibration plots for predicted versus observed 5-year risk of cardiovascular disease**
PREDICT-1° equations (total cardiovascular disease outcome) in (A) women and (B) men. Original 2013 American College of Cardiology/American Heart Association Pooled Cohort Equations (hard atherosclerotic cardiovascular disease outcome) in (C) women and (D) men. Recalibrated Pooled Cohort Equations (hard atherosclerotic cardiovascular disease outcome) in (E) women and (F) men.

Discussion

PREDICT is a large prospective cohort study representing patients in primary care who are recommended for cardiovascular disease risk assessment in New Zealand, a country with relatively similar cardiovascular disease event rates to many high-income nations, including the USA.[33]

	PREDICT-1° equations	Pooled Cohort Equations
Women		
R^2 (95% CI)	30 (29–31)	26 (24–28)
Harrell's C statistic (95% CI)	0·73 (0·72–0·73)	0·71 (0·70–0·72)
Royston's D statistic (95% CI)	1·334 (1·291–1·377)	1·225 (1·162–1·288)
Men		
R^2 (95% CI)	29 (28–30)	24 (23–26)
Harrell's C statistic (95% CI)	0·73 (0·72–0·73)	0·71 (0·70–0·72)
Royston's D statistic (95% CI)	1·318 (1·285–1·351)	1·157 (1·112–1·202)

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

Table 4: **Standard performance metrics for PREDICT-1° equations (estimating 5-year risk of total cardiovascular disease) and the 2013 American College of Cardiology/American Heart Association Pooled Cohort Equations (estimating 5-year risk of hard atherosclerotic cardiovascular disease) applied to the whole PREDICT-1° cohort**

	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)
Indian	1·30 (1·11–1·53)	1·65 (1·49–1·83)
Chinese or other Asian	0·82 (0·68–0·98)	0·76 (0·67–0·86)
NZDep 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 cardiovascular disease	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)
Taking lipid-lowering medications	0·86 (0·78–0·95)	0·82 (0·76–0·88)
Taking antithrombotic medications	0·95 (0·86–1·06)	0·89 (0·82–0·96)

NZDep=New Zealand Index of Socioeconomic Deprivation. *Hazard ratios are adjusted for all other variables included in the model.

Table 5: **Adjusted hazard ratios for hard atherosclerotic cardiovascular disease of new PREDICT-1° equation variables added to the Pooled Cohorts Equations**

All 401752 participants had cardiovascular disease risk assessments completed by general practitioners or their practice nurses. More than half of the participants were assessed after 2010, and no data on standard risk predictors were missing. This is the most appropriate type of study population in which to develop or validate cardiovascular disease risk prediction equations, yet similar cohorts are rare.

As a consequence of the major decrease in rates of cardiovascular disease events internationally in the past few decades[33] and the substantial changes in preventive treatments,[34] most published cardiovascular disease risk prediction equations are now likely to be out-of-date because they are based largely on older cohorts[10] such as the 2013 American College of Cardiology/American Heart Association PCEs.[7] Median predicted 5-year cardiovascular disease risk using new PREDICT equations was only 2·3% in women and 3·2% in men, and so for the PCEs[7] to markedly overestimate cardiovascular disease risk is not surprising. Moreover, although recalibration improved the PCEs performance, we also found that adding routinely available measures of socioeconomic deprivation, self-identified ethnicity, and several other easily measured predictors identified groups of patients whose risk would otherwise be appreciably underestimated or overestimated. For example, Māori, Pacific, and Indian patients with high deprivation scores had predicted cardiovascular disease risks that were twice as high as those of European or Chinese patients with low deprivation scores.

A funded national cardiovascular disease risk assessment programme introduced during PREDICT recruitment led to about 90% of all eligible primary care patients in New Zealand completing electronic cardiovascular disease risk assessments,[15] and more than a third of these assessments were done in regions using PREDICT software. By using valid comprehensive national health identifiers linked to national databases, PREDICT also captured all public hospital admissions and deaths related to cardiovascular disease occurring during follow-up. Private hospital admissions were not included but represent less than 2% of all hospital admissions related to cardiovascular disease. Most private hospital admissions are for non-acute procedures.[35] As national guidelines only provide explicit risk assessment recommendations for people who are younger than 75 years, we did not include older people in the cohort. Therefore, the new equations will be less accurate if applied to elderly people. We developed equations predicting 5-year risk, as recommended by New Zealand cardiovascular disease risk management guidelines,[4] rather than the more common 10-year risk, because most trials of cardiovascular disease risk reduction have about 5 years' follow-up.[2,3]

We followed TRIPOD recommendations for developing new prediction equations.[29] To reduce overfitting, all potential predictors and outcome definitions were prespecified. To assess the degree of overoptimism of resubstitution validation, sensitivity analyses were done by splitting the cohort into derivation and validation subcohorts and replicating the equation development and model performance procedures (appendix). Whether this type of validation is necessary in very large studies is increasingly questioned,[28,29] and unsurprisingly, equation coefficients, baseline survival functions, and performance metrics were similar irrespective of whether the whole cohort or derivation cohort was used to develop equations.

New preventive drug treatment initiated during follow-up has been proposed as a reason for why equations derived from older studies, with largely untreated

participants, now overpredict risk in contemporary, commonly treated cohorts.[34,36] In people who were taking preventive (ie, blood pressure lowering, lipid lowering, and antithrombotic) medication at baseline, we computed the proportion of person-time they remained on these medications during follow-up (P1). 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 consider the difference (ie, 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, are shown in the appendix (p 11). 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%,[36] the maximum underprediction of 5-year risk in any decile would only be 5% (ie, 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.

Nevertheless, the high level of preventive medication use in contemporary primary care populations (about a third of the PREDICT cohort) 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[7,37] include baseline blood-pressure-lowering treatment; however, for completeness, we also included use of lipid-lowering and antithrombotic medication. Because preventive treatment is seldom optimal, patients who remain at high predicted risk despite treatment (often monotherapy) will be candidates for additional interventions and should therefore be included in risk prediction cohorts.

In a review of 15 external validation studies of the 2013 PCEs, observed risk was almost always overestimated.[36] However, participants in these studies (including several randomised trials) were largely volunteers, and the authors acknowledged a possible healthy-volunteer bias. As PREDICT participants were automatically recruited in routine practice and represented patients in contemporary primary care, our findings provide the most definitive evidence that the PCEs overestimate risk. The 2017 American High Blood Pressure Guidelines[9] recommend that people with a systolic blood pressure of 130–139 mm Hg or diastolic blood pressure of 80–89 mm Hg and 10-year PCEs-predicted hard atherosclerotic cardiovascular disease risk of 10% or more should be offered blood pressure-lowering medication. In preliminary analyses with PREDICT participants meeting these treatment criteria (not shown), the original PCEs classified 30–50% more people as treatment-eligible than the recalibrated PCEs did.

In a recent systematic review of cardiovascular disease risk prediction models,[10] 363 models were identified, mostly developed in Europe and North America. The authors reported substantial variation in outcome definitions and recommended use of more uniform definitions, preferably ICD-coded events, as we have done. Although the accuracy of ICD coding for specific diagnoses can be unreliable, our broader definition of cardiovascular disease is likely to be more reliable, and high sensitivities and positive predictive values have been reported for ICD-coded cardiovascular disease events in national datasets.[38] The PREDICT total cardiovascular disease outcome was based on a Framingham Study ischaemic cardiovascular disease outcome definition[12] and included hospitalisations and deaths from angina, transient ischaemic attacks, congestive heart failure, and peripheral vascular disease as well as the so-called hard atherosclerotic cardiovascular disease outcomes predicted by the PCEs,[7] which only accounted for 56% of the PREDICT cardiovascular disease outcomes. By contrast, the UK QRISK3[37] cardiovascular disease outcomes excluded congestive heart failure and peripheral vascular disease but included diagnoses by general practitioners as well as hospitalisations and deaths from myocardial infarction, angina, stroke, and transient ischaemic attack (23% of angina and 55% of transient ischaemic attack outcomes in QRISK3 came only from general practitioner records). Coincidentally, the PCEs predicted total cardiovascular disease events (but not hard atherosclerotic cardiovascular disease events) reasonably well (not shown). An international consensus is clearly needed on outcome definitions for cardiovascular disease risk prediction equations.

Most published equations include limited numbers of predictors (typically age, sex, smoking, diabetes, blood pressure, and blood lipids[10]), yet other relatively easily measured variables are independently associated with cardiovascular disease risk. The UK QRISK3 equations,[37] which include 22 variables, are the most comprehensive equations, but they are complex, difficult to access, and not easily implementable outside of UK general practice. Separate equations have been developed in the USA for black and white people,[7] but the many other ethnic groups in the USA are not represented. However, there is a middle ground, and we present equations with more variables than the PCEs but fewer than QRISK3 that are fully specified to facilitate external validation and implementation.

We believe measures that reflect health inequity, such as socioeconomic deprivation and, if relevant, self-identified ethnicity, are important to add. Their associations with cardiovascular disease risk have been

documented elsewhere[37] and will help identify high-risk patient groups who might otherwise be undertreated. A discussion on why these measures predict cardiovascular risk is beyond the scope of this Article. Self-reported ethnicity is relatively easy to measure using a standardised question, and area-based deprivation scores have been developed from census or administrative data in many countries, although they are not commonly available to clinicians.[39] Area-based deprivation scores are included in the English QRISK[37] and Scottish ASSIGN[40] cardiovascular disease risk equations, although they require UK postcodes. ASSIGN developed so-called visitors' equations where users enter one of three values associated with the least, middle, and most deprived quintiles of the Scottish population.[41] Similarly, users of the PREDICT equation who do not have access to nationally developed deprivation indices can enter a number from 1 to 5 (lowest to highest deprivation quintile) that they consider best characterises a patient's socioeconomic status. If other predictors can be added at no or very low cost and are usually available in medical records (eg, current cardiovascular disease preventive treatment and diagnosed atrial fibrillation), then it is also appropriate to consider their inclusion. However, equation developers first need to balance the cost and safety of measuring an additional predictor against equation performance improvement. Unfortunately, the standard performance metrics (eg, C and D statistics) are global measures that are relatively insensitive to the addition of new variables, which might have clinically relevant predictive effects for subpopulations, as reflected by predictors' hazard ratios and prevalence.[42,43] Therefore, the decision to include additional predictors is ultimately a value judgment balancing potential clinical benefits against costs or harms of measurement.

In the developing era of precision medicine, old and poorly calibrated cardiovascular disease risk prediction equations like the 2013 American College of Cardiology/American Heart Association PCEs[7] need updating. Although recalibration is likely to reduce overtreatment of the healthy majority, additional predictors, including measures of equity, are needed to avoid undertreatment of vulnerable, high-risk patient groups. With rapidly evolving computerisation of medical records and the ability to link primary care, hospitalisations, and mortality data, new prediction equations can increasingly be tailored to specific populations.

Contributors

RJ, SW, AK, and TR conceptualised and designed the study. RJ, SW, AK, TR, and JH were involved in the data collection process. RP analysed the data with input from RJ, RM, KP, SM, SW, BPW, and AK. All authors were involved in data interpretation. RJ and RP drafted the manuscript and all authors revised the manuscript. All authors approved the final submitted version and agreed to be accountable for the report.

Declaration of interests

SW reports an unrelated grant from Roche Diagnostics, New Zealand. All other authors declare no competing interests.

Acknowledgments

We thank the staff and patients in the primary health-care organisations using PREDICT software who contributed to the study. We thank the Ministry of Health, Pharmac and Health Alliance for providing access to national and regional health databases. We thank Enigma Solutions Ltd for developing and implementing the PREDICT software in primary care patient management systems, for preparing the data for analyses, and for providing the encrypted national health identifiers required for anonymised data linkage. The study was funded by the Health Research Council of New Zealand, the Heart Foundation of New Zealand, and Healthier Lives National Science Challenge.

References

1 Kannel WB. Some lessons in cardiovascular epidemiology from Framingham. *Am J Cardiol* 1976; **37:** 269–82.
2 Mihaylova B, Emberson J, Blackwell L, et al. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet* 2012; **380:** 581–90.
3 Sundström J, Jackson R, Neal B, for the BPLTTC. Blood pressure-lowering treatment based on cardiovascular risk: a meta-analysis of individual patient data. *Lancet* 2014; **384:** 591–98.
4 New Zealand Guideline Group. Assessment and management of cardiovascular risk. 2003. http://www.health.govt.nz/publication/assessment-and-management-cardiovascular-risk (accessed Jan 24, 2018).
5 Task Force. European guidelines on cardiovascular disease prevention in clinical practice: executive summary. *Eur J Cardiovasc Prev Rehab* 2007; **14** (suppl 2)**:** E1–40.
6 Rabar S, Harker M, O'Flynn N, Wierzbicki AS. Lipid modification and cardiovascular risk assessment for the primary and secondary prevention of cardiovascular disease: summary of updated NICE guidance. *BMJ* 2014; **349:** g4356.
7 Goff DC Jr, Lloyd-Jones DM, Bennett G, 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)**:** S49–73.
8 Stone N, Robinson J, Lichtenstein A, 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. *Circulation* 2014; **129** (25 suppl 2)**:** S1–45.
9 Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults. *Hypertension* 2017; published online Nov 13. DOI:10.1161/HYP.0000000000000065.
10 Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016; **353:** i2416.
11 Jackson R, Barham P, Maling T, et al. The management of raised blood pressure in New Zealand. *BMJ* 1993; **307:** 107–10.
12 Anderson KV, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991; **121:** 293–98.
13 Wells S, Riddell T, Kerr A, et al. Cohort profile: the PREDICT Cardiovascular Disease Cohort in New Zealand primary care (PREDICT-CVD 19). *Int J Epidemiol* 2015; **46:** 22.
14 New Zealand Ministry of Health. Primary care data and stats. http://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/primary-care-data-and-stats (accessed Jan 24, 2018).
15 New Zealand Ministry of Health. More heart and diabetes checks. http://www.health.govt.nz/publication/more-heart-and-diabetes-checks-evaluation (accessed Jan 24, 2018).
16 New Zealand Ministry of Health. National Health Index. http://www.health.govt.nz/our-work/health-identity/national-health-index (accessed Jan 24, 2018).
17 University of Otago. Socioeconomic deprivation indexes: NZDep and NZiDep. http://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html (accessed Jan 24, 2018).
18 Cox D. Regression models and life-tables. *J R Stat Soc Series B* 1972; **34:** 187–220.

19 Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982; **69:** 239–41.

20 Belsley D, Kuh E, Welsch R. Regression diagnostics: identifying influential data and sources of collinearity. New York, NY: Wiley, 1980.

21 Therneau T, Grambsch P, Fleming T. Martingale-based residuals for survival models. *Biometrika* 1990; **77:** 147–60.

22 Royston P, Sauerbrei W. Multivariable model-building. A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. New York, NY: John Wiley & Sons, 2008.

23 StataCorp. Stata Statistical Software. Release 13. College Station, TX: StataCorp LP, 2013.

24 Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; **53:** 457–81.

25 Royston P. Explained variation for survival models. *Stata J* 2006; **6:** 83–96.

26 Harrell F, Califf R, Prior D, Lee K, Rosati R. Evaluating the yield of medical tests. *JAMA* 1982; **247:** 2543–46.

27 Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004; **23:** 723–48.

28 Steyerberg EW. Clinical prediction models. A practical approach to development, validation and updating. London: Springer, 2008.

29 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162:** W1–73.

30 Muntner P, Colantonio LD, Cushman M, et al. Validation of the atherosclerotic cardiovascular disease Pooled Cohort risk equations. *JAMA* 2014; **311:** 1406–15.

31 Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. *Stat Methods Med Res* 2016; **25:** 1692–706.

32 Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; **13:** 33.

33 Institute for Health Metrics and Evaluation, University of Washington. Causes of death visualization. https://vizhub.healthdata.org/cod/ (accessed Jan 24, 2018).

34 Liew SM, Doust J, Glasziou P. Cardiovascular risk scores do not account for the effect of treatment: a review. *Heart* 2011; **97:** 689–97.

35 Grey C, Jackson R, Schmidt M, et al. One in four major ischaemic heart disease events are fatal and 60% are pre-hospital deaths: a national data-linkage study. *Eur Heart J* 2017; **38:** 172–80

36 Cook NR, Ridker PM. Calibration of the pooled cohort equations for atherosclerotic cardiovascular disease: an update. *Ann Intern Med* 2016; **165:** 786–94.

37 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* 2017; **357:** j2099.

38 Merry AH, Boer JM, Schouten LJ, 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:** 237–47.

39 Phillips RL, Liaw W, Crampton P, et al. How other countries use deprivation indices—and why the United States desperately needs one. *Health Affairs (Project Hope)* 2016; **35:** 1991–98.

40 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:** 172–76.

41 ASSIGN score. Estimating risk for ASSIGN visitors. http://assign-score.com/estimate-the-risk/visitors/ (accessed Jan 24, 2018).

42 Cook NR. Comments on 'Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond' by MJ Pencina, et al. *Stat Med* 2008; **27:** 191–95.

43 Kooter AJ, Kostense PJ, Groenewold J, Thijs A, Sattar N, Smulders YM. Integrating information from novel risk factors with calculated risks: the critical impact of risk factor prevalence. *Circulation* 2011; **124:** 741–45.