THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Pylypchuk R, Wells S, Kerr A, et al. Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. *Lancet* 2018; published online May 4. http://dx.doi.org/10.1016/S0140-6736(18)30664-0.

APPENDIX

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

R Pylypchuk[1], S Wells[1], AJ Kerr[1,2], KK Poppe[1], T Riddell[1], M Harwood[1], D Exeter[1], S Mehta[1], C Grey[1], BP Wu[1], P Metcalf[1], J Warren[3], J Harrison[4], R Marshall[1], R Jackson[1].

[1]School of Population Health, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

[2]Cardiology Department, Middlemore Hospital, Auckland, New Zealand

[3]Department of Computer Science, Faculty of Science, University of Auckland, Auckland, New Zealand

[4]School of Pharmacy, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand.

Corresponding author

Rod Jackson

Professor of Epidemiology

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

Phone: +64 21301147

Appendix 1. Definition of prior CVD, significant renal disease and congestive heart failure

Prior CVD was defined as a history of angina, hospitalisation for ischaemic heart disease, transient ischaemic attacks, cerebrovascular disease, or peripheral vascular disease, either reported by the patient's general practitioner on the PREDICT templates or recorded in linked national hospitalisation data. Patients dispensed anti-anginal drugs prior to the index assessment recorded in the linked national drug dispensing database, were also classified as having prior CVD. Significant renal disease was defined as 'overt nephropathy among patients with diabetes' reported by the patient's general practitioner on the PREDICT templates, or estimated glomerular filtration rate (eGFR) ≤ 30 mL/min/1·73m^2 (irrespective of diabetes status) reported in the linked laboratory dataset. Congestive heart failure was defined as any prior hospitalisation for congestive heart failure recorded in linked national hospitalisation data or at least three dispensings of loop diuretics in the five years prior to the index assessment and/or dispensing of metolazone in the last 6 months prior to index assessment, recorded in the linked national drug dispensing database.

Appendix 2. International Classification of Disease-10-Australian Modification (ICD-10-AM) codes for the PREDICT-1° total CVD events outcome (and the PCEs hard atherosclerotic CVD outcome*), from hospital discharge and mortality records

Outcome type	ICD-10-AM codes
Myocardial infarction	**I210, I211 - I214, I219 - I222, I228, I229**
Unstable angina	I200
Other coronary heart disease	**I201, I208, I209, I230 - I236, I238, I240, I248, I249, I253 - I256, I460, I469**
Ischaemic stroke	**I630 - I636, I638, I639, I64**
Haemorrhagic stroke	**I600 - I616, I618, I619**
Transient ischaemic attack	G450 - G453, G458 - G468
Peripheral vascular disease	E1050 - E1052, E1150 - E1152, E1451, E1452, I7021 - I7024, I7100 - I7103, I711, I713, I715, I718, I739 - I745, I748, I749
Congestive heart failure	I110, I130, I132, I50, I500, I501, I509
Other ischaemic CVD-related deaths	E1059, E1159, E1459, I250, I2510 - I2513, I252, I258, I259, I461, I650 - I653, I658 - I664, I668 - I670, I672, I690, I691, I693, I694, I698, I700, I701, I7020, I708, I709, I714, Z951, Z955, Z958, Z959.

* The bolded codes were used to define hard atherosclerotic CVD for external validation of the Pooled Cohort Equations (PCEs). Myocardial infarction and Stroke codes defined fatal and nonfatal events but 'Other coronary heart disease' codes were only used to define fatal events in the hard atherosclerotic CVD outcome.

Appendix 3: Risk predictors included in sex-specific PREDICT-1° models and their methods of measurement

- Age in years (continuous). This variable was derived from the participant's index assessment date and their date of birth. Date of birth is a component of the National Health Index (NHI) dataset on all New Zealanders and was automatically linked to the PREDICT dataset.
- Self-identified prioritised ethnicity (New Zealand Māori, Pacific, Indian, Chinese/other Asian, **European**). For people who self-identify with more than one of the listed ethnicities, they are classified to only one ethnicity in the prioritised order above. The ethnicity variable is a component of the NHI dataset on all New Zealanders and was automatically linked to the PREDICT dataset.
- New Zealand Index of Socioeconomic Deprivation (NZDep: an area-based socio-economic deprivation score derived from national census data. [**1 = least deprived** and 5 = most deprived; included as a continuous variable in the models]).* This variable is a component of the NHI dataset on all New Zealanders and was automatically linked to the PREDICT dataset.
- Family history of premature CVD in a first-degree male relative before the age of 55 years or a first-degree female relative before the age of 65 years (**no**, yes). This variable was recorded by the health professional completing the electronic form, based on the above definition which was stated on the form.
- Smoking status (three categories: **never smoker**, ex-smoker, current smoker). This variable was recorded by the health professional completing the electronic form. The categories provided were: no-never; no – quit over 12 months ago; no – quit less than 12 months ago; yes – up to 10/day; yes – 11-19/day; yes – 20+/day. These were combined into three categories based on preliminary risk models.
- Diabetes (**no**, yes [combined type I, type II, type unknown]). This variable was recorded by the health professional completing the electronic form. The categories provided were: none; Type 1; Type 2 (including Type 2 on insulin); Type unknown; current gestational diabetes. Participants with current gestational diabetes were excluded from these analyses. In addition, if participants had been previously hospitalised with diabetes or were taking diabetic medications prior to the index assessment, they were classified as having diabetes
- History of atrial fibrillation (**no**, yes). This variable (ECG confirmed AF) was recorded by the health professional completing the electronic form. In addition, if participants had been hospitalised with atrial fibrillation they were classified as having atrial fibrillation.
- Systolic blood pressure (SBP) in mmHg (mean of two measures, continuous). These were the two most recently recorded sitting blood pressure levels at the time of the index assessment, measured by either a general practitioner or practice nurse.
- Total cholesterol to High Density Lipoprotein cholesterol (TC/HDL) ratio (one fasting or non-fasting measure, continuous). These are the most recently recorded TC and HDL levels at the time of the index assessment. They are measured in community laboratories and are automatically downloaded into patient records.
- Blood pressure lowering medication dispensed during the six months prior to the index risk assessment (**no**, yes). This variable was extracted from the National drug dispensing database.
- Lipid-lowering medication dispensed during the six months prior to the index risk assessment (**no**, yes). This variable was extracted from the National drug dispensing database.
- Antithrombotic medication dispensed during the six months prior to the index risk assessment (**no**, yes). This variable was extracted from the National drug dispensing database.

Reference categories included in the models are underlined and in bold

* A new NZDep Index is developed following each New Zealand census and PREDICT participants were allocated a NZDep score derived from the NZDep Index based on the census closest to their index assessment (i.e. NZDep 2001, 2006 or 2013). NZDep2013 combines nine variables from the 2013 census reflecting eight dimensions of deprivation and provides a deprivation score for each meshblock in New Zealand. Meshblocks are geographic units containing a median of 81 people in 2013. The dimensions of deprivation included are listed below in order of decreasing weight in the index:

Communication: People aged < 65 years with no access to the Internet at home

Income: People aged 18 – 64 years receiving a means tested benefit

Income: People living in equivalised (i.e. controlled for household composition) households with income below an income threshold

Employment: People aged 18 – 64 years unemployed

Qualifications: People aged 18 – 64 years without any qualifications

Owned home: People not living in own home

Support: People aged < 65 years living in a single parent family

Living space: People living in equivalised households below a bedroom occupancy threshold

Transport: People with no access to a car

From: http://www.otago.ac.nz/wellington/otago069936.pdf

Appendix 4. Number and type of first CVD events in the PREDICT-1° cohort

Outcome type	Non-Fatal events, n	Fatal events, n[a]	Proportion of all CVD events, %
Myocardial infarction	4,984	188	33·6
Unstable angina	2,275	11	14·9
Other coronary heart disease	343	436	5·1
Ischaemic stroke	2,124	156	14·8
Haemorrhagic stroke	445	205	4·2
Transient ischemic attack	1,123	0	7·3
Peripheral vascular disease	790	62	5·5
Congestive heart failure	1,795	113	12·4
Other Ischaemic CVD-related deaths	n/a	336	2·2
Total CVD events (N = 15,386)[b]	13,879	1,507	100

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

Appendix 5. Derivation and Validation cohort sensitivity analyses

We divided the cohort into two geographically defined sub-cohorts based on the District Health Board (DHB) area in which participants lived. Auckland and Counties Manukau DHBs formed the derivation sub-cohort and Waitemata and Northland DHB formed the validation sub-cohort. There were 235,141 people in the Derivation cohort (43% women) and 166,611 in the Validation cohort (44% women). Approximately 50% of the Derivation cohort was European compared to about 65% of the Validation cohort. The Derivation cohort also had fewer Māori participants, but the proportions of Pacific, Indian, and Chinese participants were consistently higher than in the Validation cohort (Appendix 5a). Almost a quarter of the Derivation cohort were in the highest deprivation quintile, compared with only 17% of the Validation cohort. Proportions of participants with diabetes were slightly higher in the Derivation than in the Validation cohort; 18% and 13% respectively in women, 15% and 12% respectively in men. Mean SBP was 2mmHg lower in the Derivation cohort in both sexes, but mean TC/HDL values were identical and the proportions on drug treatment were similar in the two cohorts.

The hazard ratios were similar in Cox regression models fitted in the whole and Derivation cohorts (Appendix 5b) with all 95% confidence intervals overlapping.

The calibration performance of the risk score developed in the Derivation cohort and applied to the Validation cohort was also similar to the performance observed in the whole cohort analyses, except for the highest decile of predicted risk where the observed risk was lower than predicted by 2-4% (Appendix 5c).

The model performance statistics for the Derivation cohort models applied in the Validation cohort were similar to those for the full PREDICT-1° equations and applied in the full cohort (Appendix 5d) with all 95% confidence intervals overlapping.

5a. Description of the **Derivation and Validation cohorts**

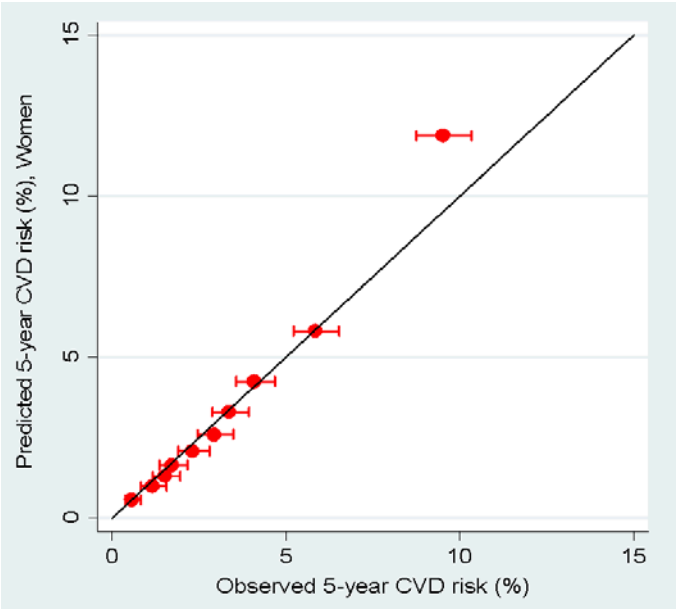
	Women (Derivation)	Women (Validation)	Men (Derivation)	Men (Validation)
Participants; n (% of total cohort)	102,242	73,457	132,899	93,154
Incident CVD events; n (% of sex-specific cohort)[a]	3,108	2,542	5,422	4,314
Total person-years observed	424,288	319,352	549,774	392,107
Crude incidence of CVD (per 1000 per year)	7·3 (7·07, 7·6) [b]	8·0 (7·7, 8·3)	9·9 (9·6, 10·1)	11·0 (10·7, 11·3)
Follow-up time in years, mean (SD)[c]	4·1 (2·7)	4·3 (2·7)	4·1 (2·7)	4·2 (2·7)
People with follow up ≥ 5 years	32,094 (31.4%)	26,399 (35.9%)	40,946 (30.3%)	31,471 (33.8%)
Age in years; mean (SD)	56 (9·0)	57 (8·7)	51 (10·0)	53 (9·7)
Ethnicity:				
European	48,857 (47·8)	47,175 (64·2)	66,338 (49·9)	62,165 (66·7)
Māori	11,445 (11·2)	12,408 (16·9)	13,592 (10·2)	13,981 (15·0)
Pacific	17,767 (17·4)	4,770 (6·5)	22,296 (16·8)	5,777 (6·2)
Indian	11,428 (11·2)	2,760 (3·8)	16,197 (12·2)	4,035 (4·3)
Chinese/other Asian	12,745 (12·5)	6,344 (8·6)	14,476 (10·9)	7,196 (7·7)
NZ Deprivation quintile:				
1 (least deprived)	23,871 (23·4)	14,652 (20·0)	30,547 (23·0)	19,832 (21·3)
2	19,132 (18·7)	15,098 (20·6)	24,995 (18·8)	19,614 (21·1)
3	15,998 (15·7)	15,810 (21·5)	20,982 (15·8)	19,702 (21·2)
4	17,435 (17·1)	15,191 (20·7)	22,949 (17·3)	18,604 (20·0)
5 (most deprived)	25,806 (25·2)	12,706 (17·3)	33,426 (25·2)	15,402 (16·5)
Smoking:				
Never smoker	78,489 (76·8)	50,669 (69·0)	90,758 (68·3)	58,381 (62·7)
Ex-smoker	11,957 (11·7)	12,881 (17·5)	20,383 (15·3)	19,473 (20·9)
Current smoker	11,796 (11·5)	9,907 (13·5)	21,758 (16·4)	15,300 (16·4)
Family history of premature CVD	11,455 (11·2)	11,541 (15·7)	12,826 (9·7)	11,669 (12·5)
Atrial fibrillation	927 (0·9)	850 (1·16)	1,916 (1·44)	1,764 (1·9)
Diabetes	17,970 (17·6)	9,407 (12·8)	19,929 (15·0)	11,013 (11·8)
SBP mmHg; mean (SD)	128 (17·6)	130 (17·8)	128 (16·2)	130 (16·2)
Mean TC/HDL; mean (SD)	3·7 (1·1)	3·7 (1·1)	4·4 (1·2)	4·4 (1·3)
Medications at index assessment:				
On blood pressure lowering medications	26,372 (25·8)	19,601 (26·7)	24,926 (18·8)	18,327 (19·7)
On lipid lowering medications	16,928 (16·6)	10,612 (14·5)	20,544 (15·5)	12,828 (13·8)
On antithrombotic medications	10,699 (10·5)	7,132 (9·7)	12,987 (9·8)	8,736 (9·4)

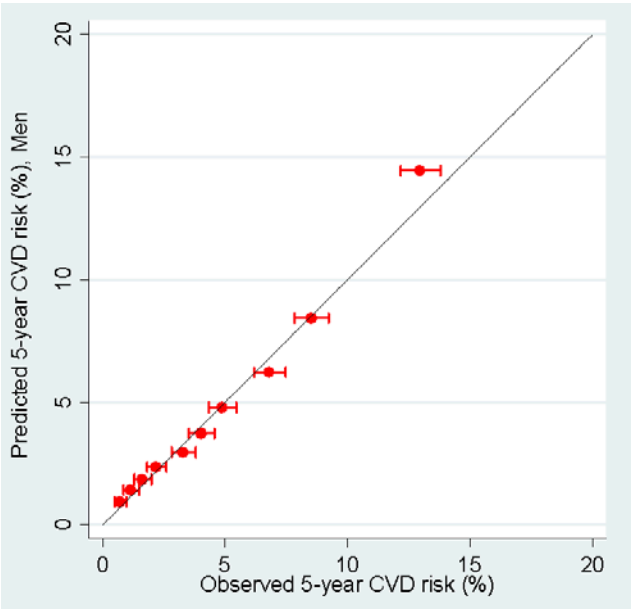
[a] Values are n (% of sex-specific cohort) unless otherwise stated.
[b] 95% Confidence Intervals
[c] The follow-up time ranged from one day to 13·3 years, in both men and women.

5b. Adjusted hazard ratios in risk models derived from the Derivation cohort and the full PREDICT-1° cohort by gender

	Women (Derivation cohort)	Women (full PREDICT-1° cohort)	Men (Derivation cohort)	Men (full PREDICT-1° cohort)
Numbers of participants / first CVD events	102242/3108	175699 / 5650	132899/5422	226053 / 9736
Predictors				
Age (per year)	1·08(1·07, 1·09)	1·08 (1·07, 1·08)	1·07 (1·07, 1·07)	1·07 (1·07, 1·07)
Ethnicity:				
European	1	1	1	1
NZ Māori	1·47 (1·31, 1·64)	1·48 (1·37, 1·60)	1·33 (1·22, 1·46)	1·34 (1·26, 1·42)
Pacific	1·15 (1·03, 1·28)	1·22 (1·12, 1·33)	1·15 (1·06, 1·25)	1·19 (1·12, 1·27)
Indian	1·10 (0·96, 1·27)	1·13 (1·00, 1·27)	1·33 (1·22, 1·46)	1·34 (1·24, 1·45)
Chinese/other Asian	0·71 (0·60, 0·83)	0·75 (0·66, 0·85)	0·71 (0·63, 0·80)	0·67 (0·61, 0·74)
NZ Deprivation quintile (per 1 quintile)	1·15 (1·11, 1·18)	1·11 (1·09, 1·14)	1·10 (1·07, 1·12)	1·08 (1·07, 1·10)
Smoking:				
Non-smoker	1	1	1	1
Ex-smoker	1·13 (1·00, 1·27)	1·09 (1·01, 1·18)	1·07 (0·99, 1·16)	1·08 (1·02, 1·14)
Smoker	1·93 (1·75, 2·13)	1·86 (1·73, 2·00)	1·64 (1·53, 1·75)	1·66 (1·57, 1·75)
Family history premature CVD	1·03 (0·92, 1·14)	1·05 (0·97, 1·12)	1·11(1·02, 1·20)	1·14 (1·08, 1·21)
Atrial fibrillation	2·28(1·86, 2·79)	2·44 (2·12, 2·81)	2·02 (1·75, 2·32)	1·80 (1·62, 2·00)
Diabetes	1·64 (1·50, 1·79)	1·72 (1·61, 1·85)	1·79 (1·67, 1·91)	1·75 (1·66, 1·85)
Systolic blood pressure (per 10 mmHg)	1·14 (1·11, 1·17)	1·15 (1·12, 1·17)	1·19 (1·16, 1·21)	1·18 (1·16, 1·20)
TC/HDL ratio (per 1 unit)	1·15 (1·12, 1·19)	1·13 (1·11, 1·15)	1·14 (1·12, 1·16)	1·14 (1·12, 1·15)
Medications at index assessment				
Not on blood pressure lowering, anti-clotting or lipid lowering medications	1	1	1	1
On blood pressure lowering medications	1·47 (1·35, 1·61)	1·40 (1·31, 1·50)	1·37 (1·27, 1·47)	1·34 (1·27, 1·42)
On antithrombotic medications	1·18(1·07, 1·30)	1·12 (1·04, 1·21)	1·07 (1·00, 1·16)	1·10 (1·03, 1·17)
On lipid lowering medications	0·91 (0·83, 1·00)	0·94 (0·88, 1·01)	0·96 (0·89, 1·03)	0·95 (0·90, 1·00)
Interactions				
Age x diabetes	0·977 (0·968, 0·985)	0·978 (0·972, 0·984)	0·980 (0·974, 0·985)	0·980 (0·976, 0·984)
Age x SBP (per 10 mmHg)	0·996 (0·993, 0·998)	0·996 (0·994, 0·997)	0·996 (0·995, 0·998)	0·996 (0·995, 0·997)
OBPLM x SBP (per 10 mmHg)	0·963 (0·927, 0·999)	0·958 (0·931, 0·985)	0·935 (0·906, 0·964)	0·948 (0·926, 0·971)

5c. Calibration plots: predicted 5-year CVD risk (%) compared to observed 5-year CVD risk (%) using the Derivation cohort models and applied to the Validation cohort, in women (left) & men (right).





5d. Performance statistics for the Derivation cohort models applied in the Validation cohort, (and comparison with the full PREDICT-1° equations applied in the full cohort) by gender

	Derivation cohort model	Full PREDICT-1° cohort
Women		
R^2	27 (25, 29)	30 (29, 31)
Harrell's C statistic	0·72 (0·71, 0·73)	0·73 (0·72, 0·73)
Royston's D statistic	1·26 (1·19, 1·32)	1·334 (1·29, 1·38)
Men		
R^2	28 (26, 29)	29 (28, 30)
Harrell's C statistic	0·72 (0·71, 0·72)	0·73 (0·72, 0·73)
Royston's D statistic	1·26 (1·21, 1·31)	1·32 (1·29, 1·35)

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

Appendix 6. Net proportion of total person-time on additional treatment with blood pressure lowering, lipid lowering and antiplatelet/antithrombotic medications by decile of predicted risk in women and men

Decile	Women,% person-time treated	Men ,% person-time treated
1	5·5	3·9
2	8·1	4·9
3	9·2	6·1
4	10·2	8·7
5	12·8	10·1
6	13·6	13·5
7	13·9	16·5
8	14·7	19·7
9	14·4	20·9
10	8·4	17·0