Australian National University

CRE Theme 3: Improving primary level chronic care

Validation and recalibration of the Framingham cardiovascular disease risk models in an Australian Indigenous cohort

Does the current Framingham risk calculator accurately estimate true CVD risk for Indigenous Australians?

Xinyang Hua, Robyn McDermott, Thomas Lung, Mark Wenitong, An Tran-Duy, Ming Li, Philip Clarke

University of South Australia

James Cook University Australia

ACKNOWLEDGEMENTS

This research is a project of the Australian Primary Health Care Research Institute, which is supported by a grant from the Australian Government Department of Health. The information and opinions contained in it do not necessarily reflect the views or policy of the Australian Primary Health Care Research Institute or the Australian Government Department of Health.

Hua X, McDermott R, Lung T, Wenitong M, Tran-Duy A, Li M, Clarke P. (2017). Does the current Framingham risk calculator accurately estimate true CVD risk for Indigenous Australians? Report prepared for the Australian Primary Health Care Research Institute (APHCRI), Canberra, ACT: Australia.

Professor Robyn McDermott

James Cook University

Cairns Qld 4870 Australia

T +61 7 4232 1604

E robyn.mcdermott@jcu.edu.au

http://www.ccdp.jcu.edu.au

CONTENTS

Contents

Background

Over a quarter of Indigenous Australian adults have some form of cardiovascular disease (CVD) [1]. In Australia, clinical guidelines for primary prevention of CVD recommend the 1991 Framingham CVD model [2] as an important component of guidelines used to identify those at high risk of developing CVD over a 5-year period [3]. The development of CVD risk prediction models using Framingham study data dates back to the late 1960s [4, 5]. The most widely used model was developed by Anderson and colleagues to predict the risk of developing CVD and its component diseases (coronary heart disease (CHD), myocardial infarction, and stroke) for people aged 30-75 years [2]. In 2008 an updated sex-specific Framingham model was published to predict 10-year CVD risks [6].

The Framingham CVD models have been validated and recalibrated in various countries and ethnicities [7-9], as well as in the Australian general population [10, 11]. However, it has not been validated or recalibrated for an Australian Indigenous population. One previous study (2005) compared the predicted and observed CHD rates in an Australian Aboriginal remote community sample and found that the 1991 Framingham model substantially underestimated CHD rates across all age groups and both sexes [12]. In the current Australian CVD management guidelines, people with the following characteristics are automatically put into the high risk category,

- diabetes and aged over 60 years
- diabetes with microalbuminuria
- eGFR<45mL/min/1.73m2
- familial hypercholesterolemia
- high blood pressure and serum total cholesterol >7.5mmol/L, and
- any Aboriginal or Torres Strait Islander over the age of 74 years [3].

However, the developers of these guidelines acknowledge that there is little empirical evidence supporting this classification system (a combination of level D weak evidence plus a consensus-based recommendation) [3]. Further, for Aboriginal and Torres Strait Islander adults aged between 35 and 74 years who are not in this clinically determined high risk category, the guidelines recommend the use of the 1991 Framingham CVD model to estimate 5-year absolute CVD risks while acknowledging that it might result in an underestimation of these risks [3].

In this study, we validated both the 1991 and 2008 Framingham CVD models using a cohort of Aboriginal and Torres Strait Islander adults drawn from remote Indigenous communities in Far North Queensland. Recalibration was also conducted to help generate more accurate CVD risk predictions for this population. Finally, we developed a CVD risk chart that could help improve the assessment and management of CVD in the Australian Indigenous population, particularly those in remote regions of Australia.

Methods

PARTICIPANTS

The source population for the present study was obtained from the Well Person's Health Check (WPHC), which was conducted between 1998 and 2000 and consisted of 3,508 people in 26 remote Indigenous communities in Far North Queensland [13]. The study was approved by the Cairns Institutional Health Ethics Committee with support from Apunipima Cape York Health Council (HREC/141QCH/121-936). Participation in the WPHC study was open to all people residing in these communities and utilised a broad range of recruitment strategies including printed media and local radio, as well as through health services and community groups [13]. Information collected in the WPHC survey can be found in the Appendices.

Baseline data of the participants were linked to hospitalisation and death records in the Queensland Hospital Admitted Patient Data Collection dataset from the initial screening date to the end of 2014, using linkage software applying deterministic and probabilistic methodologies, as well as manual clerical reviews where required. For our study, we included 1,684 (98.8%) people aged between 30 and 74 years who have a unique link to their hospitalisation and death records. We excluded people with previous CVD events (n=101) or whose baseline characteristics were missing (n=135).

Baseline risk factors including systolic blood pressure, total cholesterol level, high-density lipoprotein (HDL) cholesterol level, fasting glucose level and smoking status were collected from the WPHC screening data. Details of the methods used to collect these indicators have been published elsewhere [13]. People with diabetes were identified if they self-reported (confirmed through medical record check) or had a baseline fasting glucose >7.8 mmol/L.

One hundred and forty (8.3%) people had one or more baseline risk factors missing (six with missing systolic blood pressure value; 49 with missing total cholesterol and 134 with missing HDL cholesterol), and were excluded from the main analysis. Sensitivity analysis was conducted where we replaced the missing values with estimated values from multiple imputation by chained equations on five occasions and used standard statistical rules to produce reported results [14]. ECG-LVH (required for the 1991 Framingham model) and hypertension treatment (required for the 2008 Framingham model) were not collected during WPHC screening. We presumed these values to be the average of the Framingham population and used hypothetical values in the sensitivity analysis.

OUTCOMES

CVD events were identified using the International Classification of Disease (ICD) diagnosis and procedure codes (versions 9 and10; see Appendix 1) for the following outcomes in hospitalisation and death records,

- CHD (including myocardial infarction, angina pectoris and coronary insufficiency)
- CHD death
- Stroke
- Congestive heart failure, and
- Peripheral vascular disease as defined by Anderson et al. in the Framingham study [2].

The start date of follow-up was the screening date of the WPCH and the censor date was the date of first CVD event or death, whichever came first; or otherwise the date of last known admission. If neither admission nor death occurred during follow-up, the censor date was the 1st December 2014, which was the last known admission date of the study.

STATISTICAL ANALYSIS

We used three models to predict 5-year and 10-year CVD risks in our study cohort. The first two were the original 1991 and 2008 Framingham models [2, 6]. As only the 10-year baseline survival rate was reported in the 2008 Framingham model, this model predicted 10-year CVD risk only. The third model was a recalibrated 2008 Framingham equation, in which both the baseline 5-year and 10-year survival rates and mean values of the risk factors were estimated using the WPHC sample and coefficients on risk factors were obtained from the original 2008 Framingham model.

In detail, the 2008 Framingham model (Cox equation): $P = 1 - S_0(t)^{exp(\sum \beta iXi - \sum \beta iMi)}$, where *βi* represents the regression coefficients, *Xi* represents an individual's risk factors, *Mi* represents the means of the risk factors of the Framingham cohort, $S_0(t)$ represents the Framingham baseline CVD rate at year ten. To recalibrate this Framingham equation, we replaced the Framingham means of the risk factors (*Mi*) with the means in our own cohort, while the Framingham baseline CVD rate $S_0(t)$ was replaced with the cohort's baseline 5-year or 10-year CVD rate. The coefficients were kept the same as in the Framingham model (Table A, Appendix 2).

Age- and sex-specific predicted 5-year (original 1991 and recalibrated 2008 models) and 10-year (all three models) CVD risks for people aged between 30 and 74 years were calculated and compared with the observed CVD probabilities (estimated using the Kaplan-Meier method). The 95% confidence intervals of the differences between the predicted and observed probabilities were estimated using the bootstrap method with 1,000 bootstrapped replications.

DISCRIMINATION

Discrimination refers to the ability of a prediction model to correctly distinguish those who will develop an event from those who will not. We quantified this by calculating the Harrell's C-statistic [15], which represents the probability of concordance amongst all pairs of subjects in which at least one had an event. Concordance refers to two subjects' predicted probabilities of survival and survival times going in the same direction, e.g. the person who has higher predicted probability of survival also survives longer in reality.

CALIBRATION

Calibration describes how closely the predicted probabilities agree with observed outcomes. We used two $\chi 2$ statistics to evaluate calibration. The first method was proposed by D'Agostino and Nam [16], which compared the predicted and observed probabilities by deciles based on the predicted risk. Plots were constructed showing predicted and actual probabilities of CVD events in each decile. A $\chi 2$ statistic exceeding 20 was used to indicate a significant lack of calibration ($P<0.01$) [7]. The second method used the $\chi 2$ statistics with cross-classified categories proposed by Cook [17], in which a reclassification table was built to divide participants into different risk categories based on predictions from both original and recalibrated Framingham models. The observed and predicted probabilities were compared for all cells with at least 20 individuals [17].

To investigate the validity of the recalibrated model, the repeat data-splitting (cross validation) method was used for internal validation. The original sample was randomly divided into five samples; the recalibration was conducted on all sets of four of these samples, and the resulting five recalibrated models were used to estimate the risk in the 5th omitted sample (i.e. those individuals not used in the model development). The C-statistic

and Nam-D'Agostino chi-square were computed on the estimated results. This data-splitting procedure was repeated 200 times to obtain stable results.

All statistical analyses were performed in Stata version 13.1(StataCorp LP, College Station, TX, USA).

CVD RISK CHART

A 5-year CVD risk chart for the Australian Indigenous population was generated based on the recalibrated 2008 Framingham model. To keep the chart simple and comparable to existing Australian CVD charts [18], we retained stratification by the total cholesterol: HDL ratio. This was achieved by fixing HDL at 1.2 mmol/L, the average level in this sample. We varied the HDL level by ±0.6 mmol/L (covering the maxim values of HDL in the study cohort) to test its effect on the predicted risk levels in a sensitivity analysis.

Results

The study cohort consisted of 1,448 people from the WPHC cohort (see flowchart, Appendix 3). Baseline risk factors of this cohort are provided in Table 1. Compared to the cohort used to estimate the 2008 Framingham model (6), our study cohort is slightly younger and has a higher proportion of smokers and diabetes patients at baseline. The 10-year baseline survival rates in the study cohort is much lower compared to the Framingham cohort (Table 1).

Table 1. Baseline risk factors and survival rates for people 30-75 years old in the Framingham (D'Agostino 2008) and WPHC Indigenous cohort

	Framingham *		**WPHC cohort (30-75) †**	
Risk factors	**Women n=4,522**	**Men n=3,969**	**Women n=748**	**Men n=700**
Age, mean (SD), years	49.1 (11.1)	48.5 (10.8)	45.2 (11.6)	44.9 (11.0)
Total-C, mean (SD), mmol/L	5.6 (1.1)	5.5 (1.0)	5.0 (1.0)	5.3 (1.1)
HDL-C, mean (SD), mmol/L	1.5 (0.4)	1.2 (0.3)	1.1 (0.3)	1.2 (0.3)
Systolic BP, mean (SD), mm Hg	125.8 (20.0)	129.7 (17.6)	133.0 (21.7)	136.8 (17.9)
Smoking, n (%)	1548 (34.2)	1398 (35.2)	352 (47.1)	442 (63.1)
Diabetes, n (%)	170 (3.8)	258 (6.5)	187 (25.0)	134 (19.1)
Baseline 5-year survival rate	NA	NA	0.931	0.916
Baseline 10-year survival rate	0.950	0.889	0.846	0.811

WPHC, Well Person's Health Check; SD, standard deviation; Total-C, total cholesterol; HDL-C, high-density lipoprotein cholesterol; BP, blood pressure; NA, not applicable.

** Adapted from data in D'Agostino 2008*

† The WPHC values were used for recalibration of the 2008 Framingham model

The 1,448 people contributed 15,221 person-years of follow-up in total, with a mean and maximal follow-up time of 10.5 and 16.4 years, respectively. Three-hundred and sixty-nine (25.5%) people developed at least one CVD event during the follow-up. The 5- and 10-year probabilities of CVD events were 10.0% (95% CI: 8.5-11.7) and 18.7% (95% CI: 16.7-21.0), respectively (Table 2).

Table 2. 5-year and 10-year probability of CVD events in the WPHC indigenous sample, observed and predicted probabilities using three different Framingham models

	Sample size, n	Observed probability, % (95% CI)	Predicted probability, % (95% CI)		
			Framingham 1991	Framingham 2008	Recalibrated Framingham 2008
5-year risk					
Total	1448	10.0 (8.5-11.7)	6.8 (6.4-7.2)	NA*	10.4 (9.9-10.9)
Gender					
Female	748	9.2 (7.3-11.6)	5.7 (5.2-6.2)	NA*	9.6 (8.9-10.3)
Male	700	10.8 (8.7-13.4)	8.0 (7.4-8.5)	NA*	11.3 (10.5-12.1)
Age group					
30-34	301	3.5 (1.9-6.5)	1.3 (1.2-1.5)	NA*	2.8 (2.6-3.1)
35-44	486	3.9 (2.5-6.2)	3.8 (3.4-4.1)	NA*	6.2 (5.8-6.6)
45-54	354	13.5 (10.3-17.6)	8.0 (7.5-8.6)	NA*	11.9 (11.1-12.7)
55-74	307	21.7 (17.4-26.8)	15.5 (14.5-16.5)	NA*	22.7 (21.2-24.2)
10-year risk					
Total	1448	18.7 (16.7-21.0)	14.2 (13.5-14.8)	12.0 (11.4-12.6)	21.2 (20.3-22.1)
Gender					
Female	748	17.3 (14.7-20.5)	12.2 (11.3-13.0)	8.9 (8.2-9.5)	19.6 (18.3-20.8)
Male	700	20.2 (17.2-23.6)	16.3 (15.3-17.3)	15.4 (14.3-16.4)	22.9 (21.5-24.3)
Age group					
30-34	301	8.1 (5.4-12.2)	3.5 (3.2-3.9)	3.3 (3.0-3.6)	6.5 (6.0-7.0)
35-44	486	12.0 (9.2-15.5)	8.9 (8.3-9.5)	7.3 (6.8-7.8)	13.7 (12.8-14.5)
45-54	354	24.5 (20.1-29.6)	17.2 (16.2-18.2)	14.1 (13.1-15.1)	25.0 (23.6-26.5)
55-74	307	32.8 (27.6-38.7)	29.5 (28.0-31.0)	25.6 (23.9-27.4)	43.1 (40.9-45.3)

CVD, cardiovascular disease; WPHC, Well Person's Health Check; NA, not applicable

** The baseline 5-year survival rate was not reported in the D'Agostino 2008 study. Therefore, we were unable to calculate 5-year risk using the original Framingham 2008 model.*

The overall predicted 5-year CVD risk using the 1991 Framingham model was 6.8% (95% CI: 6.4-7.2). The predicted 10-year risk was 14.2% (95% CI: 13.5-14.8) and 12.0% (95% CI: 11.4-12.6) using the 1991 and 2008 Framingham models, respectively. All predictions significantly underestimated the observed CVD probabilities in the WPHC cohort by around a third, with differences being 3.2% (95% CI: 1.9-4.5) for 5-year risk and 4.5% (95% CI: 2.9-6.3) to 6.7% (95% CI: 5.0-8.5) for 10-year risk. Sensitivity analyses which estimated predicted risk by adjusting the prevalence of ECG-LVH or hypertension produced similar results and were reported in supplementary materials.

After baseline risk recalibration, 5-year total and age- sex-specific predicted risks were similar to the observed results (Table 2). The predicted 10-year probability of CVD events using the recalibrated model was higher than the observed risk, mainly because of overestimation in the 55-74 year age group (Table 2). Compared to the predictions from the original 1991 Framingham model, after recalibration 165/1096 people in the low 5-year CVD risk (<10%) category and 146/186 people in the moderate 5-year CVD risk (10%-15%) category moved to a higher risk category; the predicted number of people with high 5-year CVD risk (>15%) almost doubled from 166 to 322 in the cohort (P<0.001). The probabilities of CVD events using the imputed data was similar and reported in supplementary materials.

Table 3. Performance of the original and recalibrated Framingham models in predicting 5-year and 10-year CVD events

	Original Framingham		**Recalibrated Framingham 2008**
	1991	**2008**	
5-year risk			
Discrimination			
C	0.671	NA †	0.674
95 % CI of C	(0.643-0.699)	NA †	(0.646-0.702)
Calibration			
Nam-D'Agostino χ^2 (9)	85.44	NA †	18.48
P value for Nam-D'Agostino χ^2	<0.001***	NA †	0.03*
Cook χ^2 (6)	43.84	NA †	11.82
P value for Cook χ^2	<0.001***	NA †	0.07*
10-year risk			
Discrimination			
C	0.671	0.668	0.674
95 % CI of C	(0.643-0.699)	(0.640-0.696)	(0.646-0.702)
Calibration			
Nam-D'Agostino χ^2 (9)	82.56	134.67	51.09
P value for Nam-D'Agostino χ^2	<0.001***	<0.001***	<0.001***
Cook χ^2 (6)	65.91	116.13	34.65
P value for Cook χ^2	<0.001***	<0.001***	<0.001***

CVD, cardiovascular disease; C, C-statistics; CI, confidence interval; NA, not applicable.

† The baseline 5-year survival rate was not reported in the D'Agostino 2008 study. Therefore, we were unable to calculate 5-year risk using the original Framingham 2008 model.

** not statistically significant*

**** highly statistically significant using a p-value cut-point of 0.05*

Table 3 contains the C-statistics and χ2 estimates for different models. The C-statistics were between 0.668 and 0.674, with no significant differences. We found a significant lack of calibration (P<0.001) for the original (5- and 10-year) and recalibrated (10-year) Framingham risk estimations.

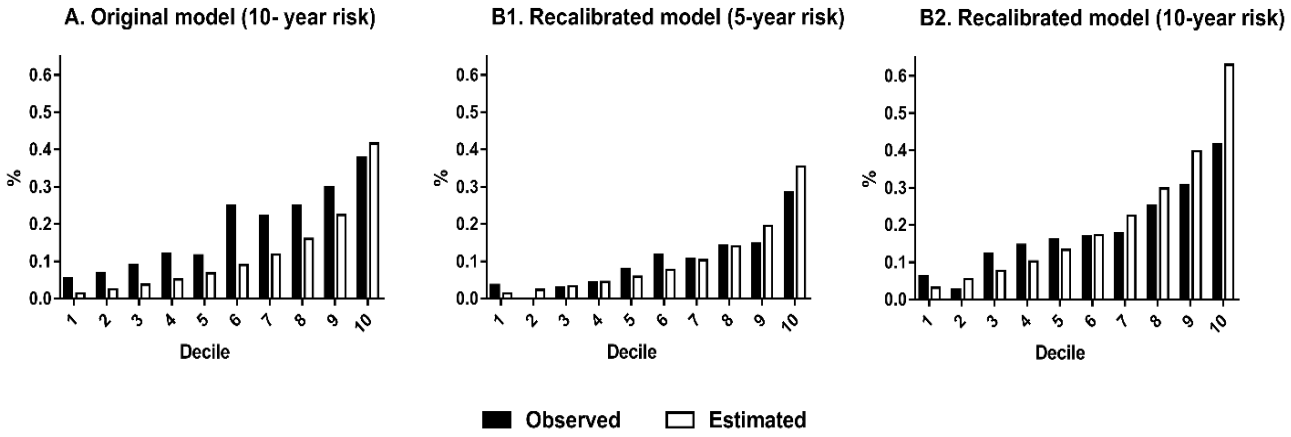


Figure 1. Calibration by decile for CVD risk of the original and recalibrated 2008 Framingham model. Horizontal-axes refer to decile of predicted risk based on the 2008 Framingham CVD model; vertical-axes refer to observed and model-based predicted probabilities of CVD event.

The recalibrated 5-year risk prediction showed improvement on calibration (Nam-D'Agostino $\chi2=18.48$, p=0.03, Cook $\chi2 =11.82$, p=0.07). Figure 1 compares predicted risks using the 2008 Framingham model and actual risks of CVD events for each decile of predicted risk. The original model (Figure 1A) shows poor calibration between estimated and observed risk in all deciles, except for the last decile. This was greatly improved after recalibration (Figure 1B); however, for the recalibrated 10-year risk large differences are still evident between the estimated and observed risks in the last decile. The performance of the recalibrated model on 5-year risk prediction did not change after internal validation, with a C-statistic of 0.678 (95% CI: 0.616-0.728) and Nam-D'Agostino $\chi2$ of 14.4 (95% CI: 10.0-20.9).

Based on the recalibrated 2008 Framingham model, a 5-year absolute CVD risk chart was built for the Australian Indigenous population (Figure 2). Predictions for people aged between 30 and 34 years were included because of the high CVD risk levels for certain populations in this age range (e.g. smokers with diabetes). A sensitivity analysis (reported in Appendix 4) showed that varying HDL levels by ±0.6 mmol/L produced only a small change of risk scores that would have a minimal impact on the classification of standard risk charts.

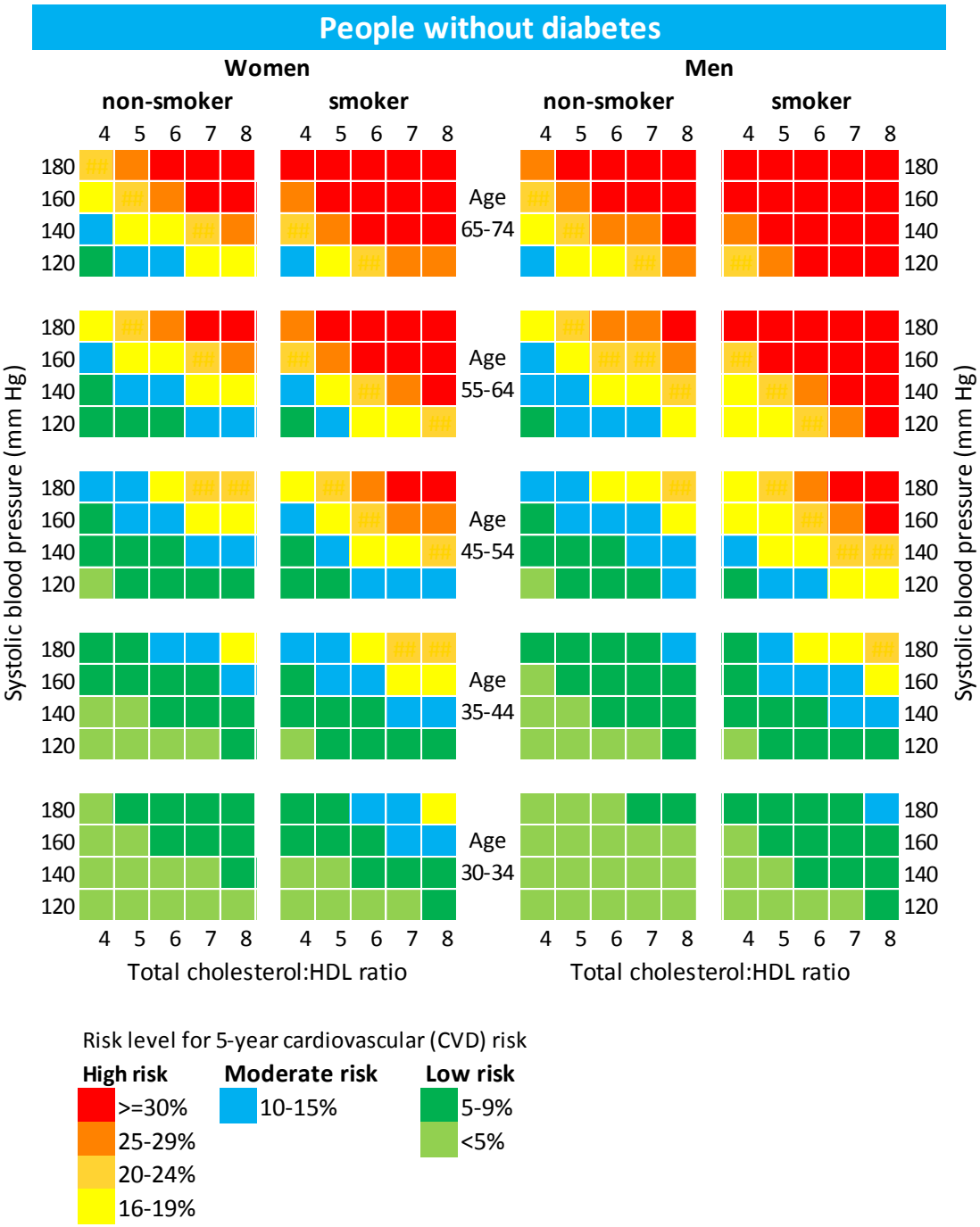


Figure 2 continues overleaf

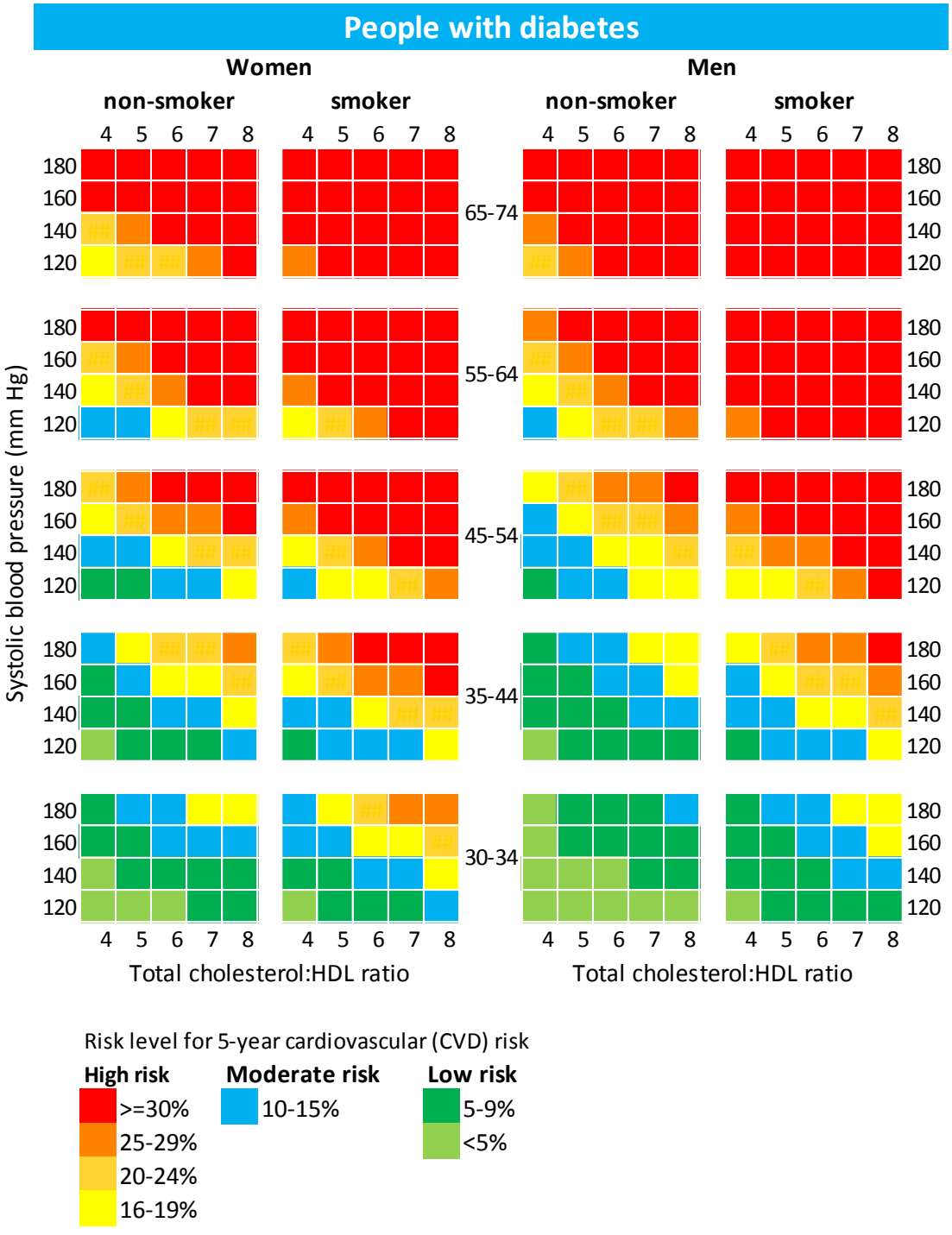


Figure 2 Five-year cardiovascular risk charts based on the recalibrated Framingham model.

**This chart is based on 2008 Framingham model that has been recalibrated using information on Aboriginal and Torres Strait Islander participants from the Well Persons Health Check study which recruited people from 26 remote communities from Far North Queensland. It has not been validated for use in other Indigenous populations.*

†For people under treatment for high blood pressure, 5% should be added to the risk on the chart.

[The Well Person's Health Check: a population screening program in indigenous communities in north Queensland[13]]

Discussion

The 1991 and 2008 Framingham CVD models substantially underestimated the absolute CVD risk in the Australian Indigenous cohort used in our study. Both models showed a lack of calibration to the observed CVD probabilities. Using the baseline risk from our study population we recalibrated the 5-year CVD risk to a level considered acceptable by the developers of the Framingham model [7]. The recalibrated equation was used to calculate the first CVD risk chart based on empirical validation using long-term follow-up data from an Indigenous Australian population.

The high CVD risk among Indigenous people presented in this study is not unique in Australia. The Strong Heart Study found that CHD rates in American Indians exceed rates in other US population and may more often be fatal [19]. A study in Canada found that Aboriginal people had a significantly higher frequency of CVD compared with Europeans (18.5% *vs* 7.6%) [20]. Although a variety of CVD risk models are available, few are based or calibrated using the observed risk in Indigenous populations. In the Strong Heart Study, both recalibration on the Framingham model and development of specific risk equation for the American Indians have been conducted to help better stratify the CHD risks in this population [7, 21]. As far as we are aware, there is no similar study for the Indigenous people in Australia. So, instead of trying more CVD risk models which were built based on other populations, we feel it is important to generate/recalibrate a prediction tool based on Indigenous Australian data. We hope this study represents a first step to producing a more accurate estimate on CVD risk for Indigenous people in Australia.

IMPLICATIONS

There is an important practical application of this research, as a previous study on CVD risk identification and management in Indigenous Australians showed that more than half of this population were not screened for CVD as recommended in national guidelines for cardiovascular risk management [19]. Current guidelines for remote health services in Indigenous communities use a CVD Risk Assessment tool which is an adaptation of the general Australian CVD charts based on the 1991 Framingham model [18], with a 5% upwards adjustment and additional estimates for a younger cohort aged 20-44 [23]. In this study we provide a risk chart derived from the recalibrated Framingham model based on the actual risk observed in a remote Indigenous population and so provides evidence that enables prediction of risk to be refined.

The more accurate calculation of CVD risk will enable better identification for Indigenous Australians in primary prevention. However, the assessment of CVD risk is just one component of a wide range of strategies to improve Indigenous health and thereby achieve the Council of Australian Governments' target on Closing the Gap in life expectancy [24]. To tackle the large numbers of high CVD risk patients more resources are required. This will need co-ordination with a range of primary care and allied health practitioners. Key to this will be the development of early risk intervention teams as well as broader community strategies such as improving infrastructure to promote healthy behaviours in Indigenous communities.

FUTURE STUDIES

After baseline calibration, there was no significant difference between the predicted and the observed 5-year CVD risks. However, overestimation occurred when using the recalibrated 10-year prediction model, mainly because of overestimation in the older age group. Hence we produced 5-year absolute CVD risk charts only. Unlike most of the guidelines in other

countries that dictated therapeutic intervention strategies based on 10-year CVD risk predictions [25], the current Australian guidelines for primary prevention of cardiovascular disease made the suggestions based on 5-year CVD risk estimations. So, we hope the 5-year risk charts produced in this study can provide reference to identify Indigenous Australian under high CVD risks.

However, future study should work on generating more accurate predictions for 10-year CVD risk as well, so clinical decisions can be made based on both short term and long-term risk estimations.

According to the data released by the Australian Bureau of Statistics, it was estimated that in 2011 there were 669,900 Indigenous people in Australia, accounting for 3% of the total Australian population; the largest population of Indigenous Australians lived in New South Wales (31.1%) and Queensland (28.2%); 34.8% of all Indigenous Australians lived in major cities of Australia, 43.8% people lived in inner or outer regional Australia and 21.4% lived in remote Australia [26]. Based on the Australian Aboriginal and Torres Strait Islander Health Survey in 2013, there was no significant difference on various health conditions and risk factors for Indigenous people across different states/territories in Australia [27]. While the study used a cohort from North Queensland, our results are consistent with a previous study that used an Indigenous cohort from remote regions of central Australia that also showed the Framingham model underestimated CHD rates [12]. Currently there is no corresponding study available on the CVD risk of Indigenous Australians in non-remote areas and so further external validation of our recalibrated model is required to determine the clinical utility of these risk charts in other Indigenous populations.

After adjustment of the baseline risk of the Framingham model in this study, calibration of the model largely improved but discrimination had no significant change. Previous studies have showed that other risk factors such as urinary albumin creatinine ratio and waist circumference and triglycerides also contributed to the development of CVD in a population with a high prevalence of diabetes [21, 28-29]. Another possible risk factor is rheumatic heart disease which can increase the risk for certain types of CVD (heart failure and stroke) and has high prevalence and mortality rate among Australian Indigenous people [30, 31]. This suggests that further recalibration which includes other predictors of CVD risk or the development of a new model specifically for the Australian Indigenous population which incorporates these risk factors should be a research priority. It would also be interesting to look at the population attributed risk for each of the tradition and new risk factors in the Indigenous population, which would provide evidence to promote more targeted strategies to reduce CVD risks.

STRENGTHS AND LIMITATIONS

Strengths of the present study include the use of relatively large baseline sample, long follow-up and objective measures (rather than self-reported) of baseline risk factors. There are also some limitations of our study. The participants of this study were from remote Indigenous communities in Far North Queensland who volunteered to participate in a population screening program. The mode of recruitment may have influenced the representativeness of the sample, which would impact on generalisability if other factors not contained in the Framingham models influences risk (e.g. body mass index) and differs between the sample and the population. External validation on our recalibrated Framingham model should be conducted in future studies involving other Indigenous populations (e.g. Indigenous Australians living in urban areas). Second, parametric uncertainty was not considered in this study when comparing observed and predicted CVD probabilities, as no variance-covariance matrices were reported for the original Framingham CVD models [32].

Conclusion

In conclusion, we found that both the 1991 and 2008 Framingham model underestimated the CVD risk in the Australian Indigenous population by about one third on average. A recalibrated equation was used to calculate the first risk chart based on empirical validation using long-term follow-up data from remote Indigenous communities in Far North Queensland.

References

1. Australian Institute of Health and Welfare. *Cardiovascular disease, diabetes and chronic kidney disease— Australian facts: Aboriginal and Torres Strait Islander people.* Report, Cardiovascular, diabetes and chronic kidney disease series no. 5. Cat. no. CDK 5. Canberra, 2015
2. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *American Heart Journal*. 1991;121(1):293-298.
3. The National Vascular Disease Prevention Alliance. *Guidelines for the management of Absolute cardiovascular disease risk,* https://heartfoundation.org.au/images/uploads/publications/Absolute-CVD-Risk-Full-Guidelines.pdf (2012, accessed 12 May 2017).
4. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases*. 1967;20(7):511-524.
5. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. *The American Journal of Cardiology*. 1976;38(1):46-51.
6. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care the Framingham Heart Study. *Circulation*. 2008;117(6):743-753.
7. D'Agostino Sr RB, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *Journal of the American Medical Association*. 2001;286(2):180-187.
8. Liu J, Hong Y, D'Agostino Sr RB, Wu Z, Wang W, Sun J, Wilson PW, Kannel WB, Zhao D. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. *Journal of the American Medical Association*. 2004;291(21):2591-2599.
9. Chow CK, Joshi R, Celermajer DS, Patel A, Neal BC. Recalibration of a Framingham risk equation for a rural population in India. *Journal of Epidemiology and Community Health*. 2009;63(5):379-385.
10. Knuiman MW, Vu H. Prediction of coronary heart disease mortality in Busselton, Western Australia: an evaluation of the Framingham, national health epidemiologic follow up study, and WHO ERICA risk scores. *Journal of Epidemiology and Community Health*. 1997;51(5):515-519.
11. Zomer E, Owen A, Magliano DJ, Liew D, Reid C. Validation of two Framingham cardiovascular risk prediction algorithms in an Australian population: the 'old' versus the 'new' Framingham equation. *European Journal of Cardiovascular Prevention and Rehabilitation.* 2011;18(1):115-120.
12. Wang Z, Hoy WE. Is the Framingham coronary heart disease absolute risk function applicable to Aboriginal people? *Medical Journal of Australia*. 2005;182(2):66-69.
13. Miller G, McDermott R, McCulloch B, Leonard D, Arabena K, Muller R. The Well Person's Health Check: a population screening program in indigenous communities in north Queensland. *Australian Health Review*. 2002;25(6):136-147.
14. Schafer JL. *Analysis of incomplete multivariate data*. CRC press; 1997.
15. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine.* 2004;23(13):2109-2123.
16. D'Agostino R, Nam B-H. Evaluation of the performance of survival analysis models: discrimination and calibration measures. *Handbook of Statistics*. 2003;23:1-25.

17. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical Chemistry*. 2008 Jan 1;54(1):17-23.

18. Heart Foundation. *Australian cardiovascular risk charts*, https://heartfoundation.org.au/images/uploads/publications/aust-cardiovascular-risk-charts.pdf (2012, accessed 12 May 2017).

19. Howard BV, Lee ET, Cowan LD, Devereux RB, Galloway JM, Go OT, Howard WJ, Rhoades ER, Robbins DC, Sievers ML, Welty TK. Rising tide of cardiovascular disease in American Indians. *Circulation*. 1999 May 11;99(18):2389-95.

20. Anand SS, Yusuf S, Jacobs R, Davis AD, Yi Q, Gerstein H, Montague PA, Lonn E, SHARE-AP Investigators. Risk factors, atherosclerosis, and cardiovascular disease among Aboriginal people in Canada: the Study of Health Assessment and Risk Evaluation in Aboriginal Peoples (SHARE-AP). *The Lancet*. 2001 Oct 6;358(9288):1147-53.

21. Lee ET, Howard BV, Wang W, Welty TK, Galloway JM, Best LG, Fabsitz RR, Zhang Y, Yeh J, Devereux RB. Prediction of coronary heart disease in a population with high prevalence of diabetes and albuminuria The Strong Heart Study. *Circulation*. 2006;113(25):2897-2905.

22. Peiris DP, Patel AA, Cass A, Howard MP, Tchan ML, Brady JP, De Vries J, Rickards B, Yarnold D, Hayman N, Brown A. Cardiovascular disease risk management for Aboriginal and Torres Strait Islander peoples in primary health care settings: findings from the Kanyini Audit. *Medical Journal of Australia*. 2009 Sep 21;191(6):304-309.

23. Centre for Remote Health. *CARPA Standard Treatment Manual,* 6th Edition, http://www.remotephcmanuals.com.au/publication/stm/Assessing_and_reducing_cardiovascular_risk.html (2014, accessed 12 May 2017).

24. Council of Australian Governments. *National Indigenous reform agreement (closing the gap)*. Report, Canberra, 2009.

25. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, Cooney MT, Corrà U, Cosyns B, Deaton C, Graham I. 2016 European Guidelines on cardiovascular disease prevention in clinical practice The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *European Journal of Preventive Cardiology.* 2016 Jun 27:2047487316653709.

26. Australian Bureau of Statistics. 3238.0.55.001 - *Estimates of Aboriginal and Torres Strait Islander Australians*, June 2011, http://www.abs.gov.au/ausstats/abs@.nsf/mf/3238.0.55.001 (2011, accessed 19 June 2017)

27. Australian Bureau of Statistics. 4727.0.55.001 - *Australian Aboriginal and Torres Strait Islander Health Survey: First Results, Australia, 2012-13*, http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4727.0.55.0012012-13?OpenDocument (2013, accessed 19 June 2017)

28. McDermott RA, McCulloch B, Li M. Glycaemia and albuminuria as predictors of coronary heart disease in Aboriginal and Torres Strait Islander adults: A north Queensland cohort. *Medical Journal of Australia*. 2011;194(10):514-518.

29. Adegbija O, Hoy WE, Wang Z. Waist circumference values equivalent to body mass index points for predicting absolute cardiovascular disease risks among adults in an Aboriginal community: a prospective cohort study. *BMJ Open*. 2015;5(11):e009185.

30. Marijon E, Mirabel M, Celermajer DS, Jouven X. Rheumatic heart disease. *The Lancet.* 2012 Mar 16;379 (9819):953-64.

31. Colquhoun, Samantha M., et al. Disparity in mortality from rheumatic heart disease in Indigenous Australians. *Journal of the American Heart Association* 4.7 (2015): e001282.

32. Ladapo JA, Goldfeld KS. Statistical uncertainty in 10-year Framingham risk of coronary heart disease and cardiovascular disease. *Journal of the American College of Cardiology*. 2014 Feb 4;63(4):377-8.

Appendix 1: ICD and procedure codes used to identify CVD events

- **CHD**, including

  MI

  angina pectoris

  coronary insufficiency

  → icd9: 4100-4149 (except 4110)
  icd10: I200-I259 (except I241)

  CHD death → death of cause

- **Stroke** → icd9: 4330-4359, 4370
  icd10: G450-G468 (except G454)
  I600-I672 (except I620-I629, I671)

- **Congestive heart failure** → icd9: 4280-4289
  icd10: I500-I509
  I110, I130, I132

- **Peripheral vascular disease** → icd9: 4400-4417 (except 4412, 4414, 4416)
  4439, 4440-4449
  icd10: I700-I718 (except I712, I714, I716)
  I739-I749

Procedure codes for CVD → icd9: 3600-3699
icd10: 3270000-3271801, 3273000-3275701, 3276300-3276303, 3276305- 3276314, 3276316-3276319, 3305000-3305500, 3307500-3310000, 3311200 -3313000, 3315100-3316300, 3317800-3355400, 3530306-3530501, 3530906 -3531501, 3845619, 3849700-3850900, 3863700, 9020100-9020103
9022900-9023000

Appendix 2: Recalibration of the 2008 Framingham CVD model

2008 Framingham COX equation:

$$P = 1 - S_0(t)^{exp(\sum \beta iXi - \sum \beta iMi)}$$

, where βi represents the regression coefficients, xi represents an individual's risk factors, Mi represents the means of the risk factors of the Framingham cohort. S0(t) is the Framingham baseline CVD rate at year 10.

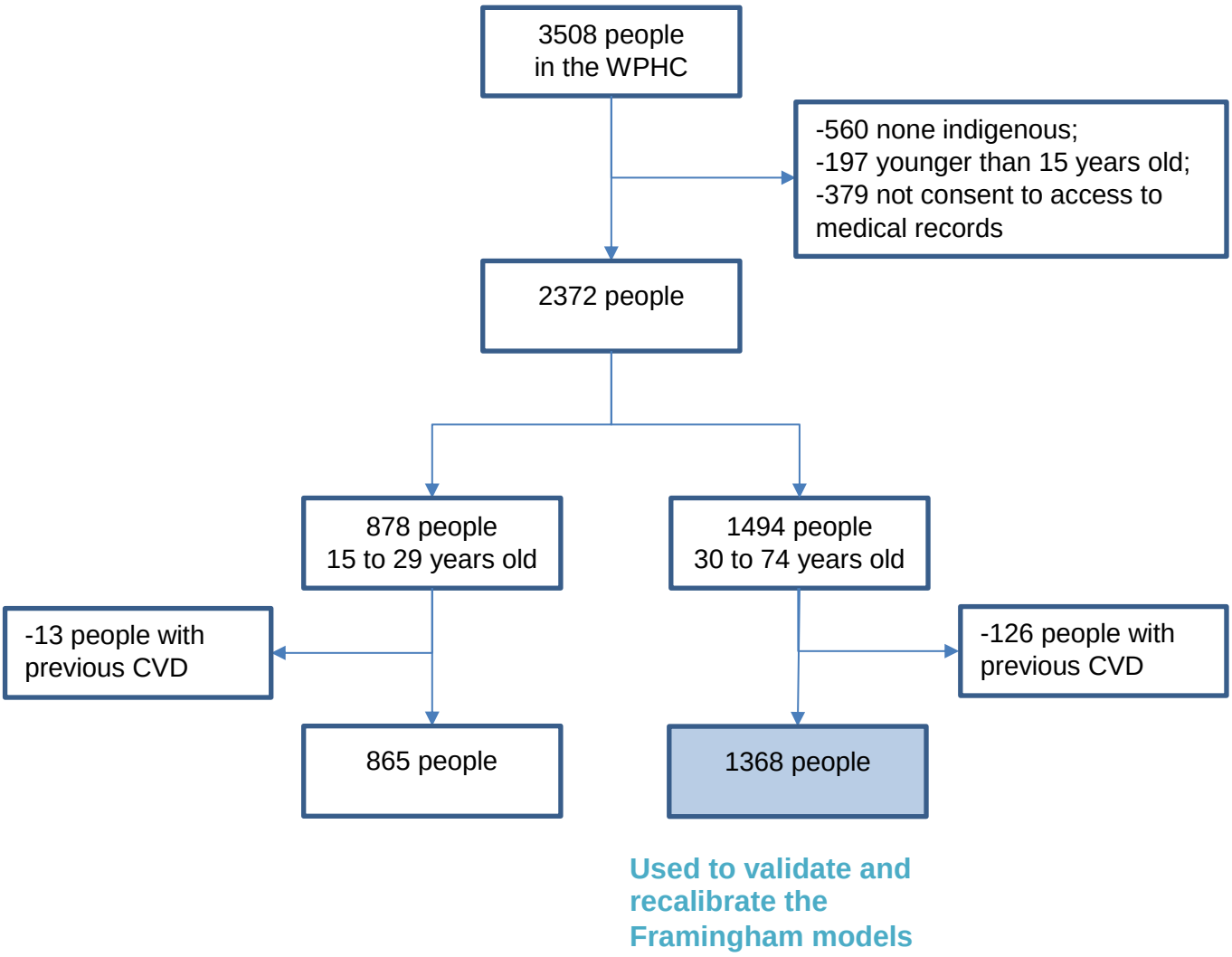
To recalibrate this Framingham equation, we replaced the Framingham means of the risk factors (Mi) by the means in our own cohort, while the Framingham baseline CVD rate S0(t) was replace by the cohort's own baseline 5-year CVD rate. The coefficients were kept the same as in the Framingham equation (table A).

Table A. Regression coefficients in the Framingham study

	Women	**Men**
Log of age	2.32888	3.06117
Log of total cholesterol	1.20904	1.12370
Log of HDL	-0.70833	-0.93263
Log of SBP if not treated*	2.76157	1.93303
Log of SBP if treated	2.82263	1.99881
Smoking	0.52873	0.65451
Diabetes	0.69154	0.57367

**High blood pressure treated rate in the Framingham cohort is 0.1176 and 0.1013 for women and men respectively*

Appendix 3: Flow chart of the sample selection process



Appendix 3: Recalibrated risk scores under different HDL levels – Sensitivity analysis

Using the recalibrated 2008 Framingham CVD model, the 5-year CVD risk of an "average" male or female (with risk factors equal to those reported in Table 1, total cholesterol five times of HDL level) can be found in the following Table B,

Table B. Recalibrated risk score of an average person under different HDL levels

HDL, mg/dL	Risk score for female	Difference from HDL of 45 mg/dL, female	Risk score for male	Difference from HDL of 45 mg/dL, male
20	6.1%	-2.9%	7.4%	-1.2%
25	6.8%	-2.2%	7.8%	-0.9%
30	7.4%	-1.6%	8.0%	-0.6%
35	8.0%	-1.0%	8.3%	-0.4%
40	8.5%	-0.5%	8.5%	-0.2%
45	9.0%	0.0%	8.6%	0.0%
50	9.5%	0.5%	8.8%	0.2%
55	9.9%	0.9%	9.0%	0.3%
60	10.4%	1.3%	9.1%	0.5%
65	10.8%	1.7%	9.2%	0.6%
70	11.1%	2.1%	9.4%	0.7%

We used the fixed HDL level of 45 mg/dL to build the Indigenous CVD chart. By varying the HDL level of ± 25 mg/dL, the change in risk score range from -2.9% to 2.1%, and -1.2% to 0.7% for female and male respectively, which would have a minimal impact on the classification for the chart.