

# Development and Use of Prediction Models for Classification of Cardiovascular Risk of Remote Indigenous Australians



An Tran-Duy, PhD <sup>a\*</sup>, Robyn McDermott, PhD <sup>b,c</sup>, Josh Knight, PhD <sup>a</sup>, Xinyang Hua, PhD <sup>a</sup>, Elizabeth L.M. Barr, PhD <sup>d,e</sup>, Kerry Arabena, PhD <sup>f</sup>, Andrew Palmer, MD, PhD <sup>a,g</sup>, Philip M. Clarke, PhD <sup>a,h</sup>

<sup>a</sup>Centre for Health Policy, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Vic, Australia

<sup>b</sup>Centre for Chronic Disease Prevention, James Cook University, Cairns, Qld, Australia

<sup>c</sup>School of Health Sciences, University of South Australia, Adelaide, SA, Australia

<sup>d</sup>Menzies School of Health Research, Darwin, NT, Australia

<sup>e</sup>Baker Heart and Diabetes Institute, Melbourne, Vic, Australia

<sup>f</sup>Centre for Health Equity, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Vic, Australia

<sup>g</sup>Menzies Institute for Medical Research, University of Tasmania, Hobart, Tas, Australia

<sup>h</sup>Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Received 2 September 2018; received in revised form 5 December 2018; accepted 5 February 2019; online published-ahead-of-print 22 February 2019

## Background

Cardiovascular disease (CVD) is the leading cause of death for Indigenous Australians. There is widespread belief that current tools have deficiencies for assessing CVD risk in this high-risk population. We sought to develop a 5-year CVD risk score using a wide range of known risk factors to further improve CVD risk prediction in this population.

## Methods

We used clinical and demographic information on Indigenous people aged between 30 and 74 years without a history of CVD events who participated in the Well Person's Health Check (WPHC), a community-based survey. Baseline assessments were conducted between 1998 and 2000, and data were linked to administrative hospitalisation and death records for identification of CVD events. We used Cox proportional hazard models to estimate the 5-year CVD risk, and the Harrell's c-statistic and the modified Hosmer-Lemeshow (mH-L)  $\chi^2$  statistic to assess the model discrimination and calibration, respectively.

## Results

The study sample consisted of 1,583 individuals (48.1% male; mean age 45.0 year). The risk score consisted of sex, age, systolic blood pressure, diabetes mellitus, waist circumference, triglycerides, and albumin creatinine ratio. The bias-corrected c-statistic was 0.72 and the bias-corrected mH-L  $\chi^2$  statistic was 12.01 (p-value, 0.212), indicating good discrimination and calibration, respectively. Using our risk score, the CVD risk of the Indigenous Australians could be stratified to a greater degree compared to a recalibrated Framingham risk score.

## Conclusions

A seven-factor risk score could satisfactorily stratify 5-year risk of CVD in an Indigenous Australian cohort. These findings inform future research targeting CVD risk in Indigenous Australians.

## Keywords

Aboriginal • Torres Strait Islanders • Prediction model • Risk score • Cardiovascular disease • Coronary heart disease

\*Corresponding author at: Centre for Health Policy, School of Population and Global Health The University of Melbourne, Level 4, 207 Bouverie Street, Carlton, Victoria 3053, Australia. Tel.: +61 3 90356387; Fax: +61 3 9348 1174., Email: [an.tran@unimelb.edu.au](mailto:an.tran@unimelb.edu.au)

## Introduction

Early prevention of cardiovascular disease (CVD) is recognised to be crucial for indigenous populations given their substantially higher CVD risk compared to non-indigenous people [1]. The excess CVD risk among Indigenous Australians is largely due to the very high prevalence of traditional CVD risk factors such as hypertension and diabetes. In addition, other highly prevalent risk factors such as albuminuria have also been shown to be associated with CVD in Indigenous Australians [2,3].

In CVD primary prevention, absolute CVD risk assessment is an important component that has been widely used and recommended [4,5]. This helps to ensure that higher-risk individuals are appropriately targeted for initiation of primary prevention strategies [6]. While the annual CVD risk assessment using a validated absolute risk algorithm has been recommended for Indigenous Australians [7], there is widespread belief that current tools have deficiencies in assessing CVD risk in this high risk population [6,8,9]. Several studies have shown that the equations derived from the Framingham study underestimate the CVD risk of Indigenous Australians [10,11]. One explanation for this underestimation may be that certain CVD risk factors for Indigenous people such as albuminuria and waist circumference are not included in the Framingham CVD equations [2,3,12,13]. While it is possible to recalibrate the Framingham models to correct for this underestimation [11], this does not account for a broader range of risk factors which are potentially useful in further stratifying risk.

On the other hand, there are also circumstances in which it is useful to stratify CVD risk using a restricted set of available risk factors [14]. For example, sometimes Indigenous participants were excluded from laboratory tests according to different rules [15]. When information on traditional risk factors is not available, a standard approach for risk stratification is to produce a model or risk chart based on a reduced set of factors [16].

In the present study, we aimed to develop and validate prediction models for 5-year risk of CVD using data from an Indigenous Australian cohort. We developed two models: a primary model that included any relevant factors and a reduced information model that does not contain laboratory variables. After internally validating these models, we used them to classify the study cohort as groups with high, moderate and low CVD risks alongside classification based on the Australian guidelines for CVD prevention [4] and the recalibrated 2008 Framingham equation [11], and compared the average risks in each predicted category with the observed risks. Then, we examined the extent to which the Australian guidelines may misclassify Indigenous Australians with and without high 5-year CVD risk.

## Methods

### Study Cohort

The cohort used in the present study was obtained from the Well Person's Health Check (WPHC) study [17] and has been

described in detail in the study by Hua et al. [11]. Briefly, the WPHC was conducted between 1998 and 2000 and included 3,508 people in 26 rural and remote Indigenous communities in North Queensland. For our analysis, we included Indigenous men and women aged between 30 and 74 years who could be linked to hospitalisation and death records for identification of the CVD events occurring after the study entry. We excluded people with a history of CVD events. In contrast to the study by Hua et al. [11], we did not exclude people with missing risk factors at baseline. Instead, we used imputations to replace the missing values with the estimates. As a result, the study cohort consisted of 1,583 individuals (see supplementary document - Figure S1 for a flow chart of the cohort selection).

### Outcome Events and Follow-Up Time

We used the definition of CVD from the Framingham Heart Study for our outcome events, which included a composite of coronary heart disease (coronary insufficiency, myocardial infarction, and angina pectoris), cerebrovascular events (ischaemic stroke, haemorrhagic stroke and transient ischaemic attack), congestive heart failure, peripheral vascular disease, and coronary deaths [18]. These events were identified using the International Classification of Diseases and procedure codes (versions ICD-9-CM and ICD-10-AM; for details, see Hua et al.) [11]. The follow-up time was calculated as the duration between baseline measurement and hospitalisation of a CVD event, coronary death, non-coronary death or last date of data collection (1 December 2014), whichever came first. To maintain alignment with the CVD risk calculators recommended by the Australian guidelines for prevention of CVD [4], a 5-year follow-up period was generated for all individuals. This was done by considering an individual as right censored if there was no recorded CVD event within the 5-year follow-up period.

### Statistical Analysis

#### Missing Data and Multiple Imputation

The percentage of observations with missing data ranged from 0% to 15.8% depending on variables recorded in the data set. We used multiple imputation [19] to replace missing data with appropriate estimates (see detailed methods in supplementary document - section 2).

#### Model Selection

We used Cox proportional hazard models to estimate the 5-year CVD risk. Two models were developed: (i) the primary model with variables selected from risk factors included in the Framingham CVD models [18] as well as other factors such as waist circumference [12] and albumin-to-creatinine ratio (ACR) [2,3,20] which have been shown to be associated with CVD risk in Indigenous populations; and (ii) the reduced information model that excluded risk factors requiring laboratory measurements. Initial covariates considered in the primary model included sex, age, total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, total cholesterol to HDL ratio,

triglycerides, systolic blood pressure (SBP), diastolic blood pressure (DBP), current smoking, diabetes status, fasting blood glucose, body mass index (BMI), waist circumference and ACR (see supplementary document - section 3 for methods of risk factor measurements). Of note, the diabetes condition was identified based on the patient self-reports and confirmed by the primary care medical records. In the reduced information model, only sex, age, SBP, DBP, current smoking, diabetes status, BMI and waist circumference were considered. To select the optimal set of predictors, we performed both stepwise and manual procedures, and included variables with a p-value smaller than 0.15 in the final models (see supplementary document - section 4 for detailed methods). We used the scaled Schoenfeld residuals [21] to test the proportional hazard assumption.

### Risk Equation Development and Performance Assessment

The equation for 5-year CVD risk based on the Cox proportional hazard model takes the following form [22]:

$$p = 1 - S(5) \exp \left\{ \sum_{i=1}^n \hat{\beta}_i (x_i - \bar{x}_i) \right\}$$

where  $p$  is the 5-year CVD risk,  $S(5)$  is the survival rate at 5 years of a person with mean values of risk factors,  $\hat{\beta}_i$  is the estimated coefficient for the  $i$ -th risk factor, and  $x_i$  and  $\bar{x}_i$  are the actual and mean values, respectively, of the  $i$ -th risk factor. We estimated  $S(5)$  and  $\hat{\beta}_i$  and their 95% CIs by fitting the final model to each of the imputed data sets and combined the estimates using Rubin's rule (see supplementary document - section 5 for detailed methods).

We used the Harrell's c-statistic [23] to evaluate the model discrimination, i.e. the model's ability to separate persons with events from those without events. The c-statistic is defined as the proportion of all possible pairs of persons, at least one of whom has an event, in which the predicted survival time is larger for the person who lived longer. A c index of 0.7 or larger indicates good discrimination.

To quantify the model calibration, i.e. the agreement between the observed and predicted risks, we used the modified Hosmer-Lemeshow (mH-L)  $\chi^2$  statistic proposed by D'Agostino and Nam for survival models [24]. The participants were divided into deciles based on their predicted CVD risks, and the Kaplan-Meier estimate for failure within each decile was compared with the mean predicted risk in the same decile. A mH-L  $\chi^2$  statistic larger than 20 indicates a poor calibration [25].

To internally validate the models, we calculated the apparent c and mH-L  $\chi^2$  statistics and corrected these for the optimism that might arise from overfitting using the algorithm developed by Harrell [23] (see supplementary document - section 6 for detailed methods). The corrected c and mH-L  $\chi^2$  statistics were shown to represent nearly unbiased assessment of the model performance [23].

### Risk Classification

We classified 1,583 individuals as either a high (>15%), moderate (10–15%) or low 5-year CVD risk (<10%) using

the algorithm provided by the Australian guidelines for CVD prevention [4], the recalibrated 2008 Framingham equation [11] and the new primary and reduced information equations, then computed mean risks and their 95% CIs within each category using Rubin's rule [19,26]. We used the Kaplan-Meier estimator to compute 5-year CVD risks of people within the same category in different imputed data sets and combined them to estimate the observed risks. The Australian guidelines automatically classify people with one or more of the following conditions as high-risk people: diabetes and age >60 years, diabetes with microalbuminuria, moderate or severe chronic kidney disease, a history of familial hypercholesterolaemia, a SBP  $\geq$  180 mmHg, serum total cholesterol >7.5 mmol/L, or age >74. For people not automatically classified as high risk, the 1991 Framingham equation [27] is recommended by the Australian guidelines for calculating the absolute risks, which are used to classify these people into high (>15%), moderate (10–15%) and low (<10%) risks. Therefore, people classified as high risk by the guidelines may have or not have any of the above-mentioned conditions. We used the same cut-off points for risk categories as those in the guidelines to classify people using our new equations. Because the Australian guidelines do not recommend any equation for the calculation of the actual risk of those automatically classified as high risk, we used the 1991 Framingham equation [27] to examine their absolute risks in reference to the observed risks and the risks predicted by our new models and the recalibrated 2008 Framingham model.

To examine the potential misclassification of 5-year CVD risk using the Australian guidelines, we calculated the proportions of people classified as high risk and medium-low risk by the Australian guidelines, and within each of these groups determined the proportions of people with a contrasting risk category based on our new models.

## Results

Of 1,583 people, 142 developed CVD within 5 years after the first screening. We found that sex, age, SBP, waist circumference, diabetes status, triglycerides and ACR ratio were associated with 5-year CVD risk. Table 1 provides summary statistics of these risk factors and the 5-year survival rate of an average person. Current smoking and cholesterol-related variables did not significantly contribute to the prediction (p-values >0.5; see supplementary document - section 7 for coefficients of all variables included in the full model). Table 2 shows the regression coefficients and hazard ratios (HRs) for CVD events associated with the risk factors in the primary and reduced information models. The proportional hazard assumption was satisfied in fitting the models to any of the imputed data sets.

Table 3 provides performance indices of our new models and the recalibrated 2008 Framingham model [11]. The bias-corrected c-statistics of the primary and reduced information models were 0.722 and 0.719, respectively, indicating good

**Table 1** Baseline risk factors and 5-year survival rate of the study cohort (N = 1,583).

Risk factor	Value*
Men, % (SE) <sup>†,‡</sup>	48.1 (1.3)
Age, mean (SE) <sup>†,‡</sup>	45.0 (0.3)
Systolic BP, mean (SE), mmHg <sup>†,‡</sup>	134.5 (0.5)
Diastolic BP, mean (SE), mmHg	75.6 (0.3)
Current smoking, % (SE)	53.4 (1.2)
BMI, mean (SE), kg/m <sup>2</sup>	28.8 (0.18)
Waist circumference, mean (SE), cm <sup>†,‡</sup>	99.2 (0.4)
Diabetes, % (SE) <sup>†,‡</sup>	23.2 (1.1)
Fasting blood glucose, mean (SE), mg/dL	114.4 (1.4)
Total cholesterol, mean (SE), mg/dL	198.9 (1.0)
HDL cholesterol, mean (SE), mg/dL	44.5 (0.3)
LDL cholesterol, mean (SE), mg/dL	121.2 (0.9)
Total cholesterol:HDL ratio	4.7 (0.04)
Triglycerides, mean (SE), mg/dL <sup>†</sup>	179.8 (3.7)
ACR, mean (SE) <sup>†</sup>	19.2 (1.6)
Underlying 5-year survival rate <sup>§</sup>	
Primary equation	0.9298
Reduced information equation	0.9344

Abbreviations: ACR, Albumin creatinine ratio; BMI, body mass index; BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SE, standard deviation of the mean.

\*Pooled values using Rubin’s rule [19] from estimates for 20 imputed data sets.

<sup>†</sup>Risk factors in the primary model.

<sup>‡</sup>Risk factors in the reduced information model.

<sup>§</sup>Estimated using the fitted Cox proportional hazard models for a person with mean risk factors.

discrimination. The bias-corrected mH-L  $\chi^2$  statistic (p-values) of the primary and reduced information models were 12.01 (0.212) and 12.49 (0.188), respectively, indicating good calibration. Compared to the recalibrated 2008 Framingham model [11], which had a bias-corrected c-statistic of 0.727 and a bias-corrected mH-L  $\chi^2$  statistic (p-value) of 29.47 (0.0005), our new models perform similarly in terms of discrimination, but far better in terms of calibration. The observed and predicted risks in deciles of the primary and reduced information models were similar (Figure 1).

Table 4 shows mean observed and predicted absolute risks for the whole cohort and different groups stratified by sex, age and predicted risk levels. The predicted mean 5-year risks for the whole cohort, women and men were 9.4%, 8.8% and 10.1%, respectively, using the primary model, and were 9.4%, 8.6% and 10.2%, respectively, using the reduced information model. These predicted values were almost identical to the observed 5-years risks (9.4%, 8.7% and 10.1% for the whole cohort, women and men, respectively).

When the cohort was stratified into high, moderate and low risks based on the predicted values, the 1991 Framingham equation [27] slightly overestimated the mean risk of the moderate-risk group (12.4% against 11.7%) and underestimated the mean risk of the low-risk group (3.1% against 5.7%). In the high-risk group classified by the guidelines’ algorithm, the 1991 Framingham equation [27] underestimated the mean risk of the whole group (13.8% against 17.1%), and overestimated the mean risk of persons who did not have any of the conditions pre-defined in the guidelines (20.9% against 17.1%). The recalibrated 2008 Framingham model [11] overestimated the mean risk in the high-risk

**Table 2** Estimated coefficients and hazard ratios from Cox proportional hazard models.

Variable	Primary model		Reduced information model	
	Coefficient (p-value) [95% CI]	Hazard ratio [95% CI]	Coefficient (p-value) [95% CI]	Hazard ratio [95% CI]
Men	0.272 (0.074) [-0.029, 0.573]	1.328 [0.925, 1.731]	0.261 (0.076) [-0.031, 0.552]	1.312 [0.927, 1.697]
Age, years	0.050 (<0.001) [0.038, 0.063]	1.052 [1.039, 1.065]	0.048 (<0.001) [0.036, 0.061]	1.050 [1.037, 1.062]
Systolic BP, mmHg	0.010 (0.012) [0.002, 0.018]	1.010 [1.002, 1.018]	0.014 (<0.001) [0.006, 0.022]	1.014 [1.006, 1.022]
Waist circumference, cm	0.009 (0.110) [-0.002, 0.020]	1.010 [0.999, 1.020]	0.009 (0.09) [-0.001, 0.020]	1.010 [0.998, 1.020]
Diabetes*	0.268 (0.104) [-0.015, 0.592]	1.325 [0.974, 1.755]	0.405 (0.009) [0.101, 0.709]	1.517 [1.055, 1.980]
Triglycerides, mg/dL	0.001 (0.142) [0.000, 0.001]	1.001 [1.000, 1.001]	-	-
ACR	0.003 (0.001) [0.001, 0.005]	1.003 [1.001, 1.004]	-	-

Abbreviations: ACR, albumin creatinine ratio; BP, blood pressure; CI, confidence interval.

\*Self-reported diabetes confirmed by the primary care medical records, used in both primary and reduced information models.

**Table 3** Performance of the new models and the recalibrated 2008 Framingham model for 5-year CVD risk prediction.

	New model		Recalibrated 2008 Framingham model
	Primary	Reduced information	
Discrimination			
Apparent c-statistic	0.735	0.731	0.728
[95% CI]	[0.698, 0.782]	[0.701, 0.783]	[0.681–0.776]
Optimism	0.013	0.012	0.001
Bias-corrected c-statistic	0.722	0.719	0.727*
Calibration			
Apparent mH-L $\chi^2$ statistic (p-value)	5.70 (0.770)	7.46 (0.589)	19.17 (0.024)*
Optimism	6.31	5.03	10.30
Biased-corrected mH-L $\chi^2$ statistic (p-value)	12.01 (0.212)	12.49 (0.188)	29.47 (<0.001)

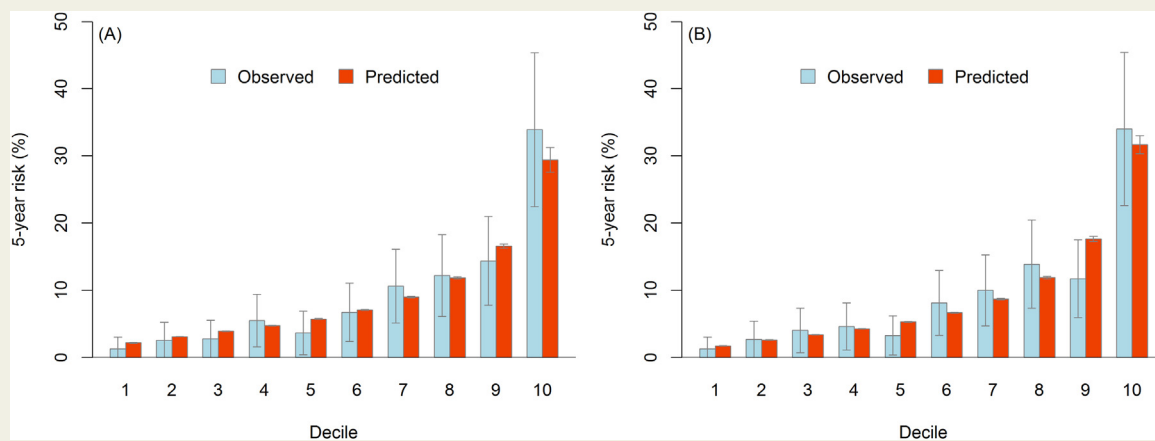
\*These values are slightly different from those in the study by Hua et al. [11] because in the present study we used multiple imputation to replace missing values with the estimates and evaluated performance of the recalibrated 2008 Framingham model on the 20 imputed data sets, while in the previous study [11] the recalibrated 2008 Framingham model was evaluated on the cohort with complete cases.

group (26.4% against 19.5%), but slightly underestimated the mean risk in the moderate- (12.3% against 13.2%) and low-risk (4.5% against 5.1%) groups. In contrast, mean risks predicted by the primary model were relatively close to the observed risks in the high- (25.0% vs 25.4%), moderate- (12.3% vs 12.6%) and low-risk (4.9% vs 4.6%) groups. Similar mean predicted and observed risks in three risk categories were observed using the reduced information model.

When risk was calculated separately for women and men (see Tables S5 and S6 in the supplementary document), we found that, in women, the CVD risk predicted by the 1991 Framingham model [27] was only one third of the observed CVD risk in the youngest group (30–34 years old), while this ratio was much larger in other age groups of women (0.53–0.86) and in any age groups of men (0.6–1.1). When the cohort was classified into major CVD risk categories, the 1991 Framingham equation [27] underestimated CVD risk

in all categories in women, while this was not the case in men with moderate risk. In both women and men, the agreement between the observed and predicted CVD risk using the recalibrated 2008 Framingham model [11] and our new models was much better compared to the 1991 Framingham model [27] in every age group and major risk category.

Using the primary model, the reduced information model, the recalibrated 2008 Framingham model [11] and the guidelines for risk classification, the proportions of cohort (i) with high risk were 17.1%, 18.3%, 22.0% and 28.4%, respectively, (ii) with moderate risk were 13.8%, 11.6%, 13.4% and 6.8%, respectively, and (iii) with low risk were 69.1%, 70.2%, 64.7% and 64.8%, respectively (Figure 2). Given the good calibration and the predicted mean risks almost identical to the observed mean risks in any risk categories of the new models, these figures suggest that



**Figure 1** Mean observed and predicted 5-year cardiovascular risks in deciles of the predicted risks using (A) the primary model and (B) the reduced information model. The error bars represent 95% confidence intervals.

**Table 4** Mean 5-year cardiovascular risk estimated using Kaplan-Meier method, 1991 Framingham equation [27], recalibrated 2008 Framingham model [11] and the new models for the whole cohort and groups classified by sex, age and predicted risk levels.

	Sample size	Observed risk based on Kaplan-Meier method, % [95% CI] <sup>†</sup>	Predicted risk, % [95% CI] <sup>†</sup>			
			1991 Framingham equation [27]	Recalibrated 2008 Framingham model [11]	New model Primary      Reduced information	
Total	1583 <sup>*</sup>	9.4 [7.9–10.8]	6.7 [6.4–7.1] <sup>¶</sup>	10.4 [9.8–10.9]	9.4 [8.9–9.8]	9.4 [8.9–9.8]
Sex						
Women	822 <sup>*</sup>	8.7 [6.7–10.7]	5.7 [5.2–6.2] <sup>¶</sup>	9.6 [8.9–10.3]	8.8 [8.1–9.4]	8.6 [8.1–9.3]
Men	761 <sup>*</sup>	10.1 [7.9–12.2]	7.9 [7.3–8.4] <sup>¶</sup>	11.2 [10.4–11.9]	10.1 [9.4–10.6]	10.2 [9.5–10.8]
Age groups (years)						
30–34	327 <sup>*</sup>	3.2 [1.2–5.2]	1.3 [1.2–1.5] <sup>¶</sup>	2.9 [2.6–3.1]	3.3 [3.1–3.5]	2.8 [2.7–3.0]
35–44	541 <sup>*</sup>	3.7 [2.1–5.3]	3.7 [3.4–4.0] <sup>¶</sup>	6.1 [5.7–6.5]	5.6 [5.4–5.9]	5.1 [4.9–5.3]
45–54	377 <sup>*</sup>	13.0 [9.5–16.4]	7.9 [7.3–8.5] <sup>¶</sup>	11.8 [11.0–12.5]	9.9 [9.2–10.5]	9.3 [8.9–9.8]
55–74	338 <sup>*</sup>	20.3 [15.8–24.6]	15.5 [14.5–16.5] <sup>¶</sup>	22.8 [21.4–24.3]	20.6 [19.5–21.7]	22.7 [21.623.8]
Risk categories based on the guidelines [4]						
High (>15%)	450 <sup>‡</sup>	17.1 [11.0–23.2]	20.9 [19.6–22.2] <sup>§</sup> / 13.8 [12.8–14.8] <sup>¶</sup>	20.4 [18.9–21.9]	16.5 [15.4–17.6]	16.3 [15.2–17.4]
Moderate (10–15%)	108 <sup>‡</sup>	11.7 [1.9–22.8]	12.4 [11.5–13.2]	17.1 [16.1–18.1]	14.1 [13.0–15.2]	15.8 [14.3–17.3]
Low (<10%)	1025 <sup>‡</sup>	5.7 [3.0–8.5]	3.1 [2.9–3.2]	5.3 [5.0–5.5]	5.7 [5.0–6.0]	5.7 [5.4–6.0]
Risk categories based on the recalibrated 2008 Framingham equation [11]						
High (>15%)	348 <sup>‡</sup>	19.5 [13.5–25.5]	15.8 [14.9–16.7] <sup>§</sup> / 18.1 [17.1–19.1] <sup>¶</sup>	26.4 [24.9–27.9]	20.4 [19.2–21.6]	21.4 [20.2–22.5]
Moderate (10–15%)	212 <sup>‡</sup>	13.7 [1.1–25.3]	8.4 [7.6–9.2]	12.3 [11.3–13.3]	11.6 [11.0–12.3]	11.8 [11.0k12.6]
Low (< 10%)	1023 <sup>‡</sup>	5.1 [1.4–9.3]	2.4 [2.2–2.6]	4.5 [4.2–4.8]	5.2 [4.7–5.3]	4.8 [4.6–5.0]
Risk categories based on the primary equation						
High (>15%)	271 <sup>‡</sup>	25.4 [15.8–35.1]	14.7 [13.3–16.1] <sup>§</sup> / 18.0 [17.1–19.2] <sup>¶</sup>	26.6 [25.0–28.2]	25.0 [23.1–25.7]	25.9 [24.7–27.1]
Moderate (10–15%)	218 <sup>‡</sup>	12.6 [6.6–19.9]	10.3 [9.7–10.9]	14.9 [14.1–15.8]	12.3 [12.1–12.5]	12.5 [12.1–12.8]
Low (<10%)	1094 <sup>‡</sup>	4.6 [2.3–7.0]	3.2 [3.0–3.4]	5.4 [5.2–5.6]	4.9 [4.9–5.1]	4.7 [4.5–4.8]
Risk categories based on the reduced information equation						
High (>15%)	289 <sup>‡</sup>	25.2 [14.9–32.9]	13.5 [12.3–14.7] <sup>§</sup> / 17.0 [16.4–18.5] <sup>¶</sup>	25.7 [24.1–27.2]	23.4 [22.1–24.6]	25.6 [24.5–26.7]
Moderate (10–15%)	183 <sup>‡</sup>	12.9 [5.2–20.6]	10.4 [9.7–11.1]	15.0 [14.0–15.9]	12.3 [11.8–12.8]	12.3 [12.1k12.5]
Low (<10%)	1111 <sup>‡</sup>	5.0 [3.0–7.1]	3.3 [3.1–3.5]	5.6 [5.4–5.9]	5.2 [5.0–5.4]	4.7 [4.5–4.8]

<sup>\*</sup>Number of persons in each of 20 imputed data sets.

<sup>†</sup>Pooled values using Rubin’s rule [19] from estimates for 20 imputed data sets.

<sup>‡</sup>Average number of persons across imputed data sets.

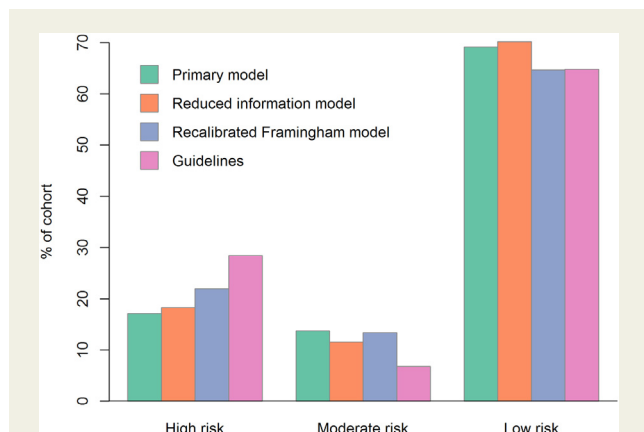
<sup>§</sup>Estimated for persons not automatically classified as those with high CVD risk by the guidelines [4], i.e. persons do not have any of the following conditions: (1) Diabetes and age>60 years, (2) Diabetes with albumin creatinine ratio >2.5 mg/mmol for males or >3.5 mg/mmol for females, (3) moderate or severe chronic kidney disease, (4) a previous diagnosis of familial hypercholesterolaemia, (5) Systolic blood pressure (BP)≥180 mmHg or diastolic BP > 110 mmHg, and (6) total cholesterol >7.5 mmol/L.

<sup>¶</sup>Estimated for all people including persons automatically classified as those with a high cardiovascular risk by the guidelines. See criteria in the previous note.

the guidelines overestimate the proportion of people with high risk, underestimate the proportion of people with moderate and low risk, and potentially misclassify a small proportion of people with actual high, moderate or low risk. The agreement between the guidelines and the new models in classification of proportions with different risk categories

was higher in the low-risk category than in the moderate- and high-risk categories.

Using our new models as a standard tool for identification of a person with a high 5-year CVD risk, we found potential misclassification of the high risk and moderate-low risk people by the Australian guidelines (Figure 3). Sixteen



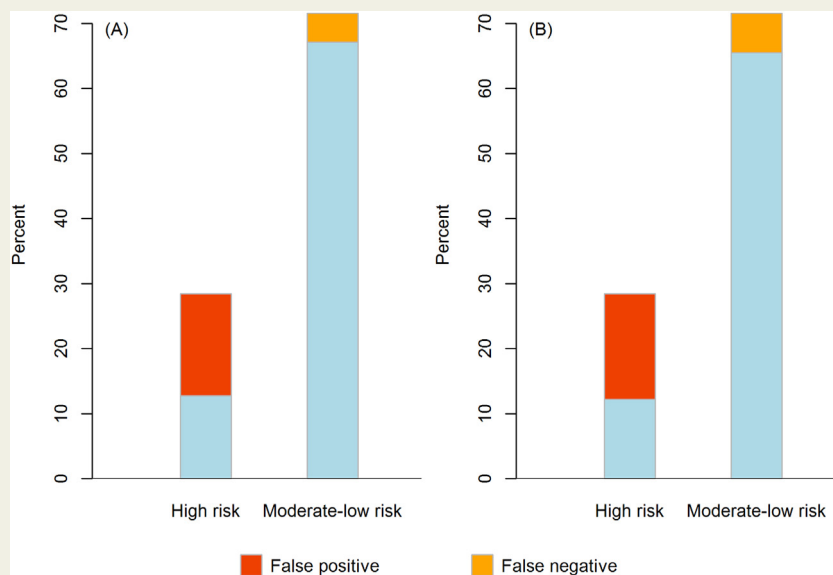
**Figure 2** Percentage of cohort with high (>15%), moderate (10–15%) and low 5-year cardiovascular risk (<10%) classified by the new models, the recalibrated 2008 Framingham model and the guidelines.

per cent (16%) of the population were false positives, i.e. the individuals classified as high risk by the guidelines but as moderate-low risk by either the primary or reduced information model. Four per cent (4%) or 6% of the population were false negatives on the basis of the primary or reduced information model, respectively, i.e. the individuals classified as moderate-low risk by the guidelines but as high risk by the primary or reduced information model. The estimated sensitivity and specificity of guidelines were 74% and 81%, respectively, based on the primary model, and were 67% and 80%, respectively, based on the reduced information model.

## Discussion

In this study, we developed the first 5-year CVD risk prediction models based on Indigenous Australian data. Both the primary model and the reduced information model performed well, and had a far better calibration (mH-L  $\chi^2$  statistic) than the recalibrated 2008 Framingham model [11]. This was expected as in the recalibrated model only the baseline survival rate and means of risk factors were derived from the Indigenous data, while the predictors and their coefficients were obtained from the Framingham equation [18]. As the reduced information model performed similarly to the primary model, it provides a convenient tool for a quick assessment of 5-year CVD risk where no laboratory variables are needed. Using multiple imputation combined with bootstrapping, we confirmed the findings in the study by Hua et al. [11] that the 1991 Framingham model [27] underestimated mean 5-year risk of the Indigenous people, and the recalibrated 2008 Framingham model [11] corrected the under prediction of the original model. We also showed that the 1991 Framingham equation [27] underestimated the CVD risk more severely in younger women (aged 30–34 years) compared to older women and to men. Despite this, the observed CVD risk in this age group is low (mean, 3.2; 95% CI, 1.2–5.2) and, therefore, this underestimation would not have impacted the cardiovascular health of younger women more strongly compared to older women and men.

In terms of discrimination, the primary model and the reduced information model performed similarly to the recalibrated 2008 Framingham model [11]. This might be because the c-statistic is not always sensitive to model fit [28,29]. Another explanation could be that a number of risk factors



**Figure 3** Potential misclassification of the high risk and moderate-low risk by the Australian guidelines for prevention of CVD based on the (A) primary model and (B) reduced information model for 5-year CVD risk prediction. The bars represent proportions of people classified as high risk (>15%) and moderate-low risk classified by the guidelines, within each of which the upper part represents the proportion of people that are misclassified on the basis of the classification by the new models.

that are associated with CVD such as family history of CVD [30], depression [31] and history of acute rheumatic fever [32,33] (which is still common in many Australian Indigenous communities [34]) were not collected and therefore could not be controlled for in our analysis. Our data set also did not contain information on socioeconomic deprivation, a risk factor for CVD mortality in an Australian cohort [35]. Therefore, socioeconomic deprivation might be a potential risk factor for CVD including both fatal and non-fatal cardiovascular events as defined in our study. It is important to note that the distributions of risk factors in our study sample are different from those in the Framingham cohort [18]. For example, participants in our study sample have a lower mean total cholesterol (198.9 vs 213.9 mg/dL), but higher mean SBP (134.5 vs 127.6 mmHg) and substantially higher mean rates of smoking (53.4% vs 34.7%) and diabetes (23.2% vs 5.0%) compared to those in the Framingham study [18]. Our models include several factors such as triglycerides and ACR that are not included in the Framingham risk equations [18,27]. The results from our analysis confirmed previous findings that ACR is an important CVD risk factor in Indigenous populations [2,3]. Two studies of Indigenous Australians showed that microalbuminuria (ACR between 3 and 30 mg/mmol) and macroalbuminuria (ACR  $\geq$  30 mg/mmol) increased the risk of coronary heart disease by two to three times compared to normal albuminuria, after adjusting for other CVD risk factors [3,20]. We did not categorise ACR because this reduced the model calibration.

The outcomes from our model fitting showed that, in the presence of waist circumference, BMI was not significantly associated with CVD risk. This is in line with the results from the study by Wang and Hoy [12], which found that waist circumference was a better predictor for CVD risk than BMI in Indigenous Australians. This may be explained by the fact that Indigenous Australians have a greater amount of abdominal fat compared to the European Australians and the use of BMI for weight status classification has been found to be inappropriate in Indigenous populations [36,37].

Surprisingly, smoking was not a significant predictor of 5-year CVD risk in our study. An explanation for this might be that smoking was masked by SBP and triglycerides, as smoking was strongly correlated with the latter. In the study by Luke et al. [38], smoking was also found to be significantly associated with dyslipidaemia and SBP. Although smoking did not significantly contribute to the prediction of 5-year CVD risk, we found it significant when fitting a model without the constraint on the 5-year follow-up time. Hence, the fact the smoking is not included in our models for 5-year CVD risk prediction does not mean that reducing smoking is not an important component of the CVD prevention strategy.

The agreement between the Australian guidelines' algorithm [4] and our new models in risk classification were only moderate. The false positives represented the largest disagreement, which can lead to unnecessary treatment and increased health care costs. The consequences of false negatives are potentially severe with individuals who would otherwise be medicated not receiving treatment to reduce

CVD risk. Although the proportion of people with false negatives was relatively small (about 4–6% of the population), this is roughly equivalent to a large number of 31,000–47,000 Indigenous people with high CVD risk but could potentially be untreated, assuming that the distributions of the risk factors in the WPHC was similar to those in the Indigenous Australian population [39].

As our models were able to better classify the CVD risk of Indigenous people compared to the algorithm recommended by the Australian guidelines [4], they should be considered as a contemporarily evidence-based tool for CVD risk assessment in remote Indigenous Australians. Since CVD is the leading cause of death in Indigenous Australians [40], translation of our findings to clinical practice is an important step towards addressing the target set by the Commonwealth of Australia, that is to close the gap in life expectancy between Indigenous and non-Indigenous Australians within a generation [41,42]. Detailing strategies for this translation is beyond the scope of the present study, but it is worth noting that developing a tool for risk assessment that is readily available and acceptable to health professionals and Indigenous communities is crucial. Communication of this tool and our findings with physicians can be done in collaboration with the Heart Foundation and other relevant organisations via the internet, scientific meetings and media.

## Strengths and Limitations

In the present study, we used a relatively large community-based cohort and high-quality data for model development, multiple imputation to deal with missing data, and rigorous methods to select the predictors and assess the model performance. However, our study has some limitations. The cohort used for model development might not represent the Indigenous population in North Queensland because the participants in the screening program were volunteers. Because our models were developed for remote Indigenous people and have not been tested in other populations, their generalisability is currently inconclusive. However, they provide an additional evidence-based tool for risk assessment that can be used alongside existing risk scores such as the recalibrated Framingham model [11]. As data on treatment for hypertension were not ascertained in our study, our models did not include treatment as a covariate, which is aligned with the 1991 Framingham equation [27]. However, we note that the 2008 Framingham equation [18] accounted for the effect of anti-hypertensive treatment and therefore it is important to explore this treatment effect in future studies for risk stratification in Indigenous populations. Other potential risk factors such as family history of CVD [30], socioeconomic deprivation [35] and rheumatic heart disease [32,33] were not collected, and thus additional relevant predictors might be missed to be included in the models. Our equations were designed as a risk prediction tool and thus they may not represent all key risk factors associated with the aetiology of CVD. An example of this is the inclusion of triglycerides as the sole lipid fraction in our primary model while there is a well-established strong correlation between CVD and other



lipid fractions such as total and HDL cholesterol [30]. In this respect the role of triglycerides in our equation represents both the direct impact as well as the indirect impact on CVD via its correlation with other lipid fractions in Indigenous Australians [43,44]. Although we rigorously assess the model performance internally, external validation of the models using other data sources is needed and this warrants further study in the future.

## Conclusions

The algorithm for classification of people as high, moderate and low 5-year CVD risk recommended by the Australian guidelines for prevention of CVD [4] is subject to misclassification of Indigenous people in these major risk categories, and is not accurate in estimating mean absolute risk in each of these categories. The prediction models we developed improved these. The reduced information model, with good discrimination and calibration, provides a convenient and reasonable tool for a quick assessment of 5-year CVD risk where no laboratory variables are required.

## Funding Source

This study is supported by the Australia National Health and Medical Research Council (project number NHMRC/1107140). The sponsor has no role in study design; or collection, analysis and interpretation of data; or writing of the manuscript; or the decision to submit the article for publication.

## Conflicts of Interest

The authors declared that there is no potential conflict of interest.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jhlc.2019.02.005>.

## References

- [1] International Diabetes Federation. Expert meeting on indigenous peoples, diabetes and development report. Copenhagen: International Diabetes Federation; 2012 . Accessed 9 August 2018 [https://www.worlddiabetesfoundation.org/sites/default/files/Indigenous%20Peoples%2C%20Diabetes%20and%20Development\\_IP%20Expert%20Meeting%202012.pdf](https://www.worlddiabetesfoundation.org/sites/default/files/Indigenous%20Peoples%2C%20Diabetes%20and%20Development_IP%20Expert%20Meeting%202012.pdf).
- [2] 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. *Med J Aust* 2011;194:514–8.
- [3] Barr E, Cunningham J, Tatipata S, Dunbar T, Kangaharan N, Guthridge S, et al. Associations of mortality and cardiovascular disease risks with diabetes and albuminuria in urban Indigenous Australians: the DRUID follow-up study. *Diabet Med* 2017;34:946–57.
- [4] National Vascular Disease Prevention Alliance. Guidelines for the management of absolute cardiovascular disease risk; 2012 . Accessed 9 August 2018 <https://www.heartfoundation.org.au/images/uploads/publications/Absolute-CVD-Risk-Full-Guidelines.pdf>.
- [5] Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'agostino RB, Gibbons R, et al. 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. *J Am Coll Cardiol* 2013;2014(63):2935–9.
- [6] Reath J, Brown N. Managing cardiovascular disease in Aboriginal and Torres Strait Islander people. *Aust Prescr* 2010;33:72–5.
- [7] Brown A, O'Shea RL, Mott K, McBride KF, Lawson T, Jennings GL. Essential service standards for equitable national cardiovascular care for Aboriginal and Torres Strait Islander people. *Heart Lung Circ* 2015;24:126–41.
- [8] Chan A. Making the case for a more accurate cardiovascular disease risk assessment tool for Indigenous Australians. *Contemp Nurse* 2015;50:92–3.
- [9] Vasant BR, Matthews V, Burgess CP, Connors CM, Bailie RS. Wide variation in absolute cardiovascular risk assessment in Aboriginal and Torres Strait Islander people with type 2 diabetes. *Front Public Health* 2016;4:37.
- [10] Wang Z, Hoy WE. Is the Framingham coronary heart disease absolute risk function applicable to Aboriginal people? *Med J Aust* 2005;182:66–9.
- [11] Hua X, McDermott R, Lung T, Wenitong M, Tran-Duy A, Li M, et al. Validation and recalibration of the Framingham cardiovascular disease risk models in an Australian Indigenous cohort. *Eur J Prev Cardiol* 2017;24:1660–9.
- [12] Wang Z, Hoy W. Waist circumference, body mass index, hip circumference and waist-to-hip ratio as predictors of cardiovascular disease in Aboriginal people. *Eur J Clin Nutr* 2004;58:888–93.
- [13] Adegbija O, Hoy W, Wang Z. Prediction of cardiovascular disease risk using waist circumference among Aboriginals in a remote Australian community. *BMC Public Health* 2015;15:57.
- [14] Modesti PA, Agostoni P, Agyemang C, Basu S, Benetos A, Cappuccio FP, et al. Cardiovascular risk assessment in low-resource settings: a consensus document of the European Society of Hypertension Working Group on Hypertension and Cardiovascular Risk in Low Resource Settings. *J Hypertens* 2014;32:951–60.
- [15] Rowley KG, Thorpe AH. Research, information and consent for the Australian Health Survey: a separate standard for Indigenous people? *Med J Aust* 2011;195:158–9.
- [16] World Health Organization. Prevention of cardiovascular disease : guidelines for assessment and management of cardiovascular risk. Geneva: World Health Organization; 2007 Accessed 9 August 2018 [http://www.who.int/cardiovascular\\_diseases/guidelines/Full%20text.pdf](http://www.who.int/cardiovascular_diseases/guidelines/Full%20text.pdf).
- [17] 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. *Aust Health Rev* 2002;25:136–47.
- [18] D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care — the Framingham Heart Study. *Circulation* 2008;117:743–53.
- [19] Rubin D. Multiple imputation for nonresponse in surveys. New York: John Wiley & Son; 2004.
- [20] Wang Z, Hoy WE. Albuminuria and incident coronary heart disease in Australian Aboriginal people. *Kidney Int* 2005;68:1289–93.
- [21] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994;515–26.
- [22] Woodward M. Epidemiology: study design and data analysis, 3rd ed. Boca Raton: CRC Press; 2014.
- [23] Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- [24] D'Agostino R, Nam B-H. Evaluation of the performance of survival analysis models: discrimination and calibration measures. *Handbook of statistics*, 23. 2003;p. 1–25.
- [25] D'Agostino Sr RB, Grundy S, Sullivan LM, Wilson P, CRP Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA* 2001;286:180–7.
- [26] Rubin DB, Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Stat Assoc* 1986;81:366–74.
- [27] Anderson K, Odell P, Wilson P, Kannel W. Cardiovascular disease risk profiles. *Am Heart J* 1991;121:293–8.

- [28] May M, Royston P, Egger M, Justice AC, Sterne JA, AC Collaboration. Development and validation of a prognostic model for survival time data: application to prognosis of HIV positive patients treated with antiretroviral therapy. *Stat Med* 2004;23:2375–98.
- [29] Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928–35.
- [30] Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353:i2416.
- [31] Goldston K, Baillie AJ. Depression and coronary heart disease: a review of the epidemiological evidence, explanatory mechanisms and management approaches. *Clin Psychol Rev* 2008;28:288–306.
- [32] Carapetis JR, McDonald M, Wilson NJ. Acute rheumatic fever. *Lancet* 2005;366:155–68.
- [33] Marijon E, Mirabel M, Celermajer DS, Jouven X. Rheumatic heart disease. *Lancet* 2012;379:953–64.
- [34] RHD Australia (ARF/RHD writing group) NHFoAatCSoAaNZ. The Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease, 2nd ed. National Heart Foundation of Australia and the Cardiac Society of Australia and New Zealand; 2012.
- [35] Backholer K, Hirakawa Y, Tonkin A, Giles G, Magliano DJ, Colagiuri S, et al. Development of an Australian cardiovascular disease mortality risk score using multiple imputation and recalibration from national statistics. *BMC Cardiovasc Disord* 2017;17:17.
- [36] Piers L, Rowley K, Soares M, O'Dea K. Relation of adiposity and body fat distribution to body mass index in Australians of Aboriginal and European ancestry. *Eur J Clin Nutr* 2003;57:956–63.
- [37] World Health Organization. Obesity: preventing and managing the global epidemic. WHO Technical Report Series 894. Geneva: World Health Organization; 2000.
- [38] Luke JN, Brown AD, Brazionis L, O'Dea K, Best JD, McDermott RA, et al. Exploring clinical predictors of cardiovascular disease in a central Australian Aboriginal cohort. *Eur J Prev Cardiol* 2013;20:246–53.
- [39] Australian Bureau of Statistics. 2940.0 — Census of population and housing: details of overcount and undercount. Australia: Australian Bureau of Statistics; 2016 Accessed 9 August 2018 <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2940.0Main+Features12016?OpenDocument>.
- [40] Australian Bureau of Statistics. 3303.0 — Causes of death. Australia: Australian Bureau of Statistics; 2015, <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/3303.0~2015~Main%20Features~Australia's%20leading%20causes%20of%20death,%202015~3>.
- [41] Australian Institute of Health and Welfare. Chapter 5: Health of population groups. In: Australia's health 2016. Australia's health series no. 15. Cat. no. AUS 199. Canberra: AIHW; 2016 Accessed November 20, 2018 <https://www.aihw.gov.au/getmedia/5d39a104-a2d5-4ab5-900c-697ee0e5a1d8/ah16-5-8-main-contributors-indigenous-life-expectancy-gap.pdf.aspx>.
- [42] Commonwealth of Australia-Department of the Prime Minister and Cabinet. Closing the gap Prime Minister's report; 2018 Accessed July 31, 2018 <https://closingthegap.pmc.gov.au/sites/default/files/ctg-report-2018.pdf?a=1>.
- [43] Australian Bureau of Statistics. 4727.0.55.003 — Australian Aboriginal and Torres Strait Islander Health Survey: biomedical results, 2012–13; 2014 Accessed 30 November 2018 <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4727.0.55.003~2012-13~Main%20Features~Triglycerides~122#5>.
- [44] Li M, McCulloch B, McDermott R. Metabolic syndrome and incident coronary heart disease in Australian indigenous populations. *Obesity* 2012;20:1308–12.