An external validation of Thais' cardiovascular 10-year risk assessment in the southern Thailand

Ponlapat Satian[1], MD; Suthara Aramcharoen[2], MD; Ponlachart Chotikarn[3], PhD; Sipat Triukose[4,5,] PhD

[1] Lansaka Hospital, Nakhonsrithammarat

[2] Thungsong Hospital, Nakhonsrithammarat

[3] Marine and Coastal Resources Institute, Faculty of Environmental Management, Prince of Songkla University

[4] Chulalongkorn University Big Data Analytics and IoT Center (CUBIC), Chulalongkorn University,

[5] Research group on Applied Digital Technology in Medicine (ATM),

Chulalongkorn University

Abstract

Background: Cardiovascular diseases (CVDs) are the number one cause of death globally. WHO estimated that CVD was a cause of 17.9 million deaths (or 31% of all global deaths) in 2019. It may seem surprising, but we can easily prevent CVDs by altering our lifestyles to avoid associated risk factors. The only requirement needed is to know one's risk apriori. Various versions of cardiovascular risk scores are adopted to assess cardiovascular risk and associated risk factors in many countries. Thailand developed as the Thai CV risk score as a reliable tool to forecast the risk of having a cardiovascular event in the future for Thais.

Objective: This study is an external validation of the Thai CV risk score. We aim to answer two key questions. Firstly, Can the Thai CV Risk score developed using the dataset of people from the central and northwestern regions of Thailand applies to people from other parts of the country? Secondly, Can the Thai CV Risk score, developed for general people, work for hospital patients who tend to have a higher risk?

Method: This work is a retrospective study set out to answer two questions using a dataset of 1,025 patients from Lansaka Hospital in southern Thailand. We calculated 10-years cardiovascular risk using the following factors – sex, age, smoke, diabetes, systolic blood pressure, LDL, and HDL -- of all patients in 2008 and compare the predictive result to observe their actual cardiovascular events from 2008 to 2017.

Result: We are able to find answers to both questions in this study. We find that the Thai CV risk score works for the southern Thais population, including patients in the hospital. It generally works well for the low CV risk group.

Conclusion: Even though the Thai CV risk score is applicable for hospital settings, it tends to overestimate moderate and high risks. Fortunately, this poses no serious concern for general people as it only makes people more careful about their lifestyle. The doctor should be careful when using the score with other factors to make a treatment decision.

Keyword: An external validation, cardiovascular diseases, cardiovascular events, CV risk model, Thai CV risk score.

Introduction

Cardiovascular diseases (CVDs) are the number one cause of death globally. In 2019, WHO estimated that 17.9 million people died from CVDs (or 31% of all global deaths). The deaths from CVDs are from individual or combined effects of high blood pressure, serum cholesterol, smoking, high blood glucose, and high body mass index, as shown in figure 1 [1]. We can easily prevent most CVDs by modifying their lifestyle to avoid behavioral risk factors or preventive medical treatment. Therefore, it is essential to have a model that can predict each individual's cardiovascular (CV) risk.

The World Health Organization (WHO) estimates that over 75% of premature CVD is preventable, and risk factor amelioration can help reduce the growing CVD burden on both individuals and healthcare providers.

WHO predicted that small changes in cardiovascular risk factors could halve mortality from coronary heart disease (CHD) in the United Kingdom[12]:

- A 1% decrease in cholesterol in the population could lead to a 2–4% CHD mortality reduction.
- A 1% reduction in smoking prevalence could lead to 2000 fewer CHD deaths per year.
- A 1% reduction in population diastolic blood pressure could prevent around 1500 CHD deaths each year.

From 1972 to 1992, the decline of the major cardiovascular risk factors in Finland contributed to 80% of the CHD mortality reduction. Similarly, in Ireland, almost half (48.1%) of the reduction in CHD mortality rates during 1985–2000 among those aged 25–84 years has been attributed to favorable trends in population risk factors. In both countries, the most significant benefits appear to have come from reductions in mean cholesterol concentrations, smoking prevalence, and blood pressure levels [12] [13].

The world-first CV risk model, Framingham Risk Score, was developed based on findings from Framingham Heart Study [2]. The original cohort study, founded in 1948, consisted of 5,209 men and women. The study findings revealed how cardiovascular health affects the rest of the body. The study found high blood pressure and high blood cholesterol to be significant risk factors for cardiovascular disease.

The Framingham Risk Score is one of many scoring systems used to determine an individual's cardiovascular disease risk. Cardiovascular risk scoring systems estimate the probability that a person will develop cardiovascular disease within a specified amount of time, usually 10 to 30 years.

In Thailand, the Thai CV Risk Score[3] was developed based on cohorts collected from The Electricity Generating Authority of Thailand (EGAT) study [4]. The model was derived from the 1st EGAT cohort comprising a survey and blood sample from 3,499 EGAT employees in the Bangkok area in 1985. The model was later adjusted and validated using 2nd and 3rd EGAT cohorts comprising 2,999 and 2,584 samples from northwestern and central regions of Thailand, respectively.

This work is an external validation of the Thai CV risk score to address two main issues and one far-reaching goal with

cardiovascular diseases. Firstly, the current Thai CV Risk Score was developed using a cohort of people from Thailand's central and northwestern regions. How well does the score predict the CV risk for people from other parts of the country? Secondly, the CV risk is currently being used as a guideline to prescribe drugs in the statins group for patients in hospitals. A question remains unanswered about how well the score developed for general people works for hospital patients who typically have a higher risk.

Finally, it is rather convenient to gather a large set of patients' medical records in the digital era, such as records from a hospital database. The medical community in the UK exploits big data by using a dataset of roughly 10 million people to develop their recent CV Risk Score, QRISK3 [5]. However, the characteristics of this large dataset [8] are in many ways different from the traditional cohort. Therefore, different methods to use big data in medical research need to be studied and developed. This work is considered the first step in our journey to use big data to develop a cardiovascular risk score in the Thailand context [14].

Methods

Sample size calculation.

In this study, we calculated the sample size based on the Lansaka population in 2008. We used Krejcie and Morgan formular.

$$n = \frac{X^2 Np(1-p)}{e^2(N-1) + X^2 p(1-p)}$$

Where

n is the required sample size

X^2 is the table value of Chi-square for 1 degree of freedom at the confidential 95%

N is the population size

p is the population proportion

e^2 is the degree of accuracy expressed as a proportion

Determine the sample size required for the study with 95% confidence and a margin of error of 5%. Assume a population proportion of 0.5 and a population size of 40,291 (Lansaka area population in 2008).

$$n = \frac{3.841 \times 40{,}291 \times 0.5 \times 0.5}{(.05)^2 \times (40{,}291 - 1) + 3.841 \times 0.5 \times 0.5} = 380.4822$$

Thus, a sample size of at least 380 people is needed. However, we decided to use all available data on patients' health records at Lansaka Hospital, which gave us a dataset three times larger than the needed sample size.

Table 1: Characteristics of the Lansaka's cohort

Characteristics	Lansaka's Cohort (1,025 patients, from 2008 – 2017)	
	Male	Female
Age		
Range	35 - 70	35 - 70
Mean (SD)	54.8 (8.98)	54.7 (9.13)
Gender		
Count (%)	319 (31)	706 (69)
Smoke (count, %)		
Smoke	175 (54.86)	26 (3.68)
Unknown	52 (16.3)	68 (9.63)
Physical Exam (mean, SD)		
BMI (kg/m^2)	24.57 (4.35)	25.41 (4.15)
Systolic BP	139.2 (18.87)	136.42 (19.45)
Prevalence of disease (%)		
Diabetes	18.18%	24.08%
General chemistry (mean, SD)		
LDL	130.7 (32.84)	137.1 (32.15)

Table 1 characterizes the dataset of 1,025 patients (319 male and 706 female) without cardiovascular events prior to 2008 used in our study.The age range for both genders are from 35- to 70-year-old, the average (and standard deviation) of age for both genders are 54.8 (8.98) and 54.7 (9.13) for male and female, respectively, and no significant difference in age distribution among genders. A male patient group has a significant portion of smoking history, 54.86 vs 3.68 percent for males and females, respectively.

We found no significant difference in both body mass index and systolic blood pressure among both genders. The prevalence of diabetes is higher in females and males, 24% and 18%, respectively. An average (and standard deviation) of low-density lipoprotein are 130.7 (32.84) and 137.1 (32.15) for male and female participants, respectively, with a slightly high LDL in females.

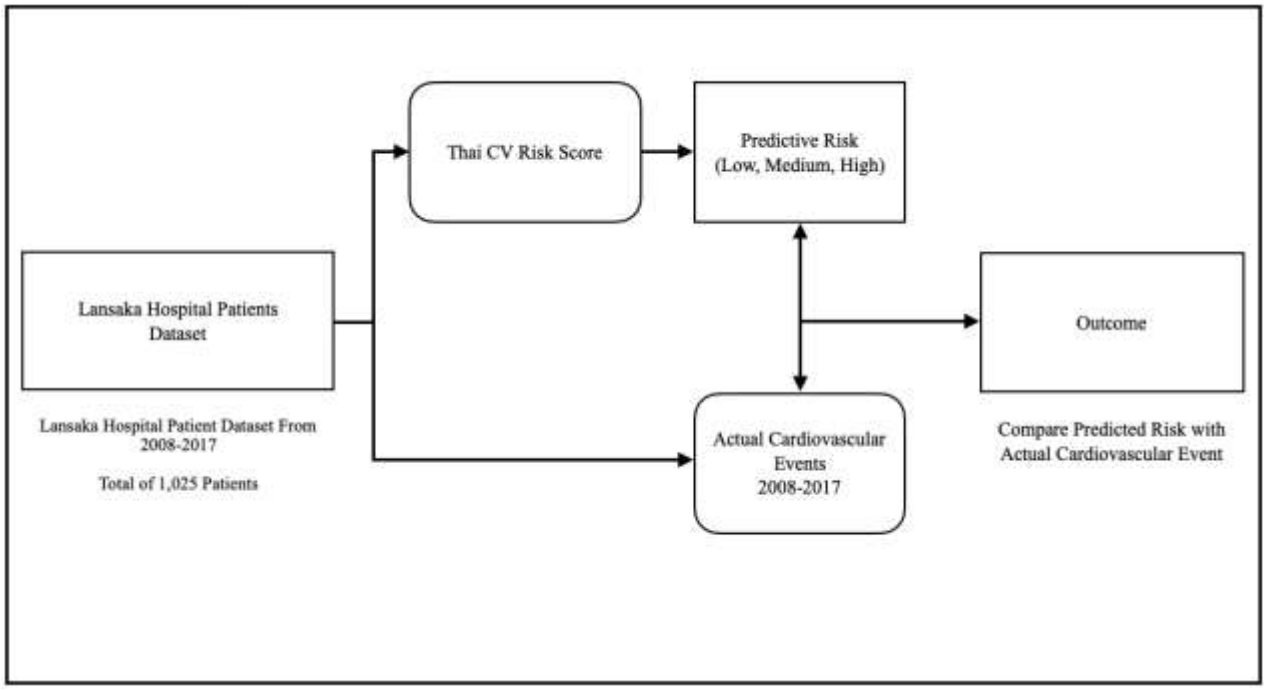

Figure 1: Methodology used in this study

Table 2: Number of patients presented with cardiovascular events during 2009–2017

Hazard Event	Subtype	ICD 10 Code	ICD 10 Group	A Number of Cases
Coronary Revascularization		Z95	Z95.5	N/A
Angina Pectoris		I209	-	5
Unstable Angina		I200	I200-I209	8
Myocardial Infarction	STEMI	I214	I210-I219	16
	NSTEMI	I213		
CHD Death		I519	I510-I519	11
Stroke	Hemorrhagic Stroke	I619	I600-I64	43
	Ischemic Stroke	I64		
Cardiac Failure		I509	I500-I509	N/A

Figure1 illustrates our research methodology. We first calculated 10-years cardiovascular risk using the following factors – sex, age, smoke, diabetes, systolic blood pressure, LDL, and HDL -- of all patients in 2008. We then observed their actual cardiovascular events from 2008 to 2017. Specifically, in this study, we focus on cardiovascular events comprising fatal and non-fatal myocardial infarction and stroke, as shown in Table 2. Stroke is the most frequent event among our target patients, while myocardial infarction and CHD death are the second and third most frequent events.

Ethics approval

This study was approved by the Nakhonsrithammarat Provincial Health Office Research Ethics Committee (Reference Number NSTPH 24/2561). Permission to use anonymized data from Lansaka hospital's patient records.

Results

Table 3: The validation outcome of the Thai CV risk score

Risk Level	Predicted (Group's Average)	Actual	Estimation
Low Risk [0,10%)	0.0529	0.0707	As Expected
Moderate Risk [10%,20%)	0.1522	0.0884	Over Estimate
High Risk [20,30%)	0.2523	0.1364	Over Estimate

Table 3 presents the validation outcome of the Thai CV risk score. The risk score indicates chances of having cardiovascular events in 10-year time. The scores below 0.1, between 0.1 and 0.2, and more than 0.2 are interpreted as low, moderate, and high risks [3].

Therefore, we grouped Lansaka's patients into three groups based on their predicted risk level and calculated the actual risk for each group from cardiovascular events between 2008 and 2017. Essentially, we divided the predicted risk of all participants into 20 bins with a width of 0.05 each. Then for each bin, we analyze the actual risk based on actual cardiovascular events and then calculate an error by comparing how much the actual risk is higher or lower than the upper or lower bound of the bin. Errors of all risk levels -- low, medium, high -- are average errors of bins in respective risk levels. On average, in the low-risk group, the risk score slightly underestimated the actual risk by 0.0068. However, in both moderate and high-risk groups, the score, on average, overestimated the actual risk by 0.037 and

0.088, respectively. This could partly be due to the right censoring issue, discussed later in the limitations of this study.

Although this study showed that over-predict incidents occurred in Lanska's patients in some groups, the finding was still consistent with several studies that used Framingham risk score prediction, which has been widely used in western countries [15,16,17].

Discussion

Here we discuss the results of our experiment, particularly along the main objectives of this study. In the end, we sketch ideas to improve the score.

This study aims to answer two key questions. First, "Can the Thai CV Risk score developed using the dataset of people from the central and north-western regions of Thailand applies to people from other parts of the country?" Second, "Can the Thai CV Risk score, developed for general people, work for hospital patients who tend to have a higher risk?" The result of this study reveals that in the low-risk range, the score works quite well; however, in the moderate and high risks, the score tends to overestimate.

We believe there are two reasonable explanations for the overestimation. First, the dataset used in this study is the Lansaka hospital's patient records from 2008 to 2017; therefore different from the cohort dataset. Due to several factors, such as migration, this kind of dataset might be incomplete or present a censoring problem. For example, suppose the patient moved to another city after 2008 and had a cardiovascular event later. In that case, Lansaka's dataset will indicate that this patient never had any CVD events, which is inaccurate. This type I censoring scenario leads to the right censoring condition [7]. In our study, patients likely to fall into this censoring scenario were excluded to mitigate this problem, even if they stayed with Lansaka hospital for the entire 10-year period.

The second explanation is that a small fraction of patients has complete information to calculate the 10-year risk score. One of the major contributions is the missing cholesterol values, which is understandable since a blood draw is required to acquire this data.

Even though the Thai CV risk score is applicable for most Thais, it is still not perfect. We portray ideas to improve the score further.

A set of risk factors used in the current Thai CV risk score is similar to the original Framingham risk score [2]. However, the more recent CV risk score, such as QRISK3 [5], uses more extensive factors. Our research team aims to extend the current Thai CV risk score to include additional factors that could help improve the CV risk assessment for all Thais.

We are exploring a non-invasive approach to obtain these missing values to compensate for the likelihood of missing lab results. For example, we are looking into deep learning on retinal fundus images [6] to infer the missing information. Along this line of thinking, we believe modern digital technology could help compensate for a shortfall in healthcare data.

Conclusion

In summary, the Thai CV risk score works for the southern Thais population, including patients in the hospital. It generally works well for the low CV risk group. However, the score tends to overestimate moderate and high risks, and the error increases as the risk increases. Fortunately, this poses no serious concern for general people as it only makes people more careful about their lifestyles [18]. The doctor should be careful when using the score with other factors to make a treatment decision.

Reference

[1] Tzoulaki I, Elliott P, Kontis V, Ezzati M. Worldwide exposures to cardiovascular risk factors and associated health effects: current knowledge and data gaps. Circulation. 2016;133(23):2314-33.

[2] Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. The Lancet. 2014;383(9921):999-1008.

[3] Faculty of Medicine Ramathibodi Hospital, Mahidol University[online]. 2019 Feb 12 [cited 2022 Mar 10]; Available from: URL: https://med.mahidol.ac.th/cardio_vascular_risk/thai_cv_risk_score

[4] Vathesatogkit P, Woodward M, Tanomsup S, Ratanachaiwong W, Vanavanan S, Yamwong S, et al. Cohort profile: The electricity generating authority of Thailand study. Int J Epidemiol. 2012;41(2):359-65.

[5] Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. 2017;357:j2099.

[6] Poplin R VA, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering. 2018;2:158-64.

[7] Wikipedia.org [online]. 2021 Dec 30 [cited 2022 Mar 19]; Available from: URL: https://en.wikipedia.org/wiki/Censoring_(statistics)

[8] Hilbert, Martin. Big data for development: A review of promises and challenges. Development Policy Review 34. 2016;1: 135-174.

[9] Albarqouni L, Doust JA, Magliano D, Barr ELM, Shaw JE, Glasziou PP. External validation and comparison of four cardiovascular risk prediction models with data from the Australian diabetes, obesity and lifestyle study. Med J Aust. 2019;

[10] Stewart J, Manmathan G, Wilkinson P. Primary prevention of cardiovascular disease: A review of contemporary guidance and literature. JRSM Cardiovasc Dis. 2017;

[11] Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, et al. External validations of cardiovascular clinical prediction models: A large-scale review of the literature. Circulation: Cardiovascular Quality and Outcomes. 2021.

[12] WHO [online]. 2016 Oct 10 [cited 2022 Feb 22]; Available from: URL: http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cardiovascular-diseases/data-and-statistics
[13] Liew SM, Blacklock C, Hislop J, Glasziou P, Mant D. Cardiovascular risk scores: Qualitative study of how primary care practitioners understand and use them. Br J Gen Pract. 2013;
[14] Lindbohm J V., Sipilä PN, Mars N, Knüppel A, Pentti J, Nyberg ST, et al. Association between change in cardiovascular risk scores and future cardiovascular disease: analyses of data from the Whitehall II longitudinal, prospective cohort study. Lancet Digit Heal. 2021;
[15] Kingkaew N, Antadech T. Cardiovascular risk factors and 10-year CV risk scores in adults aged 30-70 years old in Amnat Charoen Province, Thailand. Asia-Pacific Journal of Science and Technology. 2019;24(04):1-10.
[16] Collins G.S, Altman D.G. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ. 2010;340:2442.
[17] Barroso L.C, Muro E.C, Herrera N.D, Ochoa G.F, Hueros J.I.C, Buitrago F. Performance of the framingham and SCORE cardiovascular risk prediction functions in a non-diabetic population of a Spanish health care centre: a validation study. Scand J Prim Health Care. 2010;28(4):242-248.
[18] Duangrithi D, Wattanasermkit R, Rungwijee S, Khunsom N. Thai CV risk score and primary prevention in impaired fasting plasma glucose or diabetes mellitus versus normoglycemia in patients with metabolic syndrome. Int J Prev Med. 2020;