Project title:

Harnessing Evidence And Real-world Data To Improve Cardiovascular Health and Services (HEARTWISE)-The Personalised - CARdiovascular Disease Risk Assessment for Chinese (P-CARDIAC) Real World Evidence Study - Population-based Study

Brief Title:
HEARTWISE - P-CARDIAC for Chinese: Population-based Study

HKU/HA HKW IRB: UW 24-064
Date: 31 October 2023

Study Protocol

Abstract

Background: Cardiovascular disease (CVD) is one of the prominent diseases that affect many people. One cost-effective solution is to identify people at higher risk of CVD by CVD risk prediction model. China-PAR, TRS-2P, and SMART2 are common risk prediction models for prevention. However, these risk scores were mostly based on the routinely self-check health information and multivariable regression without time-varying consideration. We developed a Machine Learning (ML) based risk prediction model, Personalized CARdiovascular DIsease risk Assessment for Chinese (P-CARDIAC) among a predominantly Chinese population in Hong Kong to estimates the 10 years of secondary recurrent CVD risk for the high-risk individuals. Objective: The objective is to evaluate the accuracy of the P-CARDIAC performance in practice among a large-scale Hong Kong population in medicine specialist outpatient clinic (SOPC) and cardiac clinic.

Participants: Individuals aged over 18 years old and above who visited the medical or cardiology SOPC clinics under HA.

Method: The study has 2 parts. Part 1 is a prospective population-based cohort study. We hypothesize that patients with a higher frequency or larger number of types of cardiac event symptoms are likely to have a higher P-CARDIAC risk score. The results will reassure cardiologists that the P-CARDIAC risk score is sensitive to the heart disease symptoms. Part 2 of the study will involve Delphi technique to determine P-CARDIAC risk thresholds with reference to current clinical management guideline such as AHA/ACC and ESC/EAS. Researchers will collect insights from cardiologists regarding the interpretability of performance metrics and acceptable threshold of model performance for clinical practice.

Expected results: We anticipate that the results may help us to facilitate P-CARDIAC in clinical setting and provide more practical information with the development of P-CARDIAC.

1. Introduction

1.1 Background

Cardiovascular disease (CVD) is one of the prominent diseases that affect many people in their middle or old age and result in fatal complications worldwide [1]. It has been regarded as a global disease burden and a leading cause of mortality, especially in developing countries [2, 3]. From 2009 to 2021, the number of persons diagnosed with heart disease increased by 21% according to the Census and Statistics Department's Thematic Household Surveys in Hong Kong [4]. According to a report from World Health Organization, 80% of premature heart attacks and strokes are preventable [5]. To prevent heart disease, one cost-effective solution is to identify people at higher risk of CVD and offer them appropriate advice for earlier prevention [6]. China-PAR (Prediction for ASCVD Risk in China) [7], Thrombolysis in Myocardial Infarction risk score for secondary prevention (TRS-2P) [8], and Secondary Manifestations of ARTerial disease (SMART2) [9] risk score are common risk prediction models for the prevention and management of CVD. These risk prediction models will serve as a framework for integrating electronic patient health records with the use of standard statistical modelling and then simplified into a points system for clinical use. Results from these risk scores provide an important reference to help clinicians give appropriate care to high-risk individuals. By initiating preventive treatment, risk prediction outcomes potentially will also lower the burden on the public healthcare sector. However, these risk scores were mostly based on the routinely self-check health information and multivariable regression without time-varying consideration which is not feasible for providing accurate prediction. A systematic review recently published in 2022 that summarized 18 CVD risk prediction models from 22 studies based on Chinese population [10]. The study concluded that although most of the models were developed among Chinese population after 2015, their usefulness still remained unclear due to lacking in external validation and comparison for the local setting and recurrent episode.

To address this issue, we developed a Machine Learning (ML) based risk prediction model, Personalized CARdiovascular DIsease risk Assessment for Chinese (P-CARDIAC) among a predominantly Chinese population in Hong Kong. It estimates the 10 years of secondary recurrent CVD risk for the high-risk individuals with consideration of an array of risk variables and concurrent medication effect under time-varying effect.

1.2 Rational and research gap of the risk prediction models validation on local clinical setting

In 2016, a developed risk prediction model designed for Chinese adults in multiple contemporary cohorts, China-PAR, was published [7]. China-PAR were recommended by the updated guideline on the assessment and management of cardiovascular risk, however, a study conducted among a large-scale Chinese population in real-world

indicated that it underestimated the risk for woman in southern Chinese population due to the geographic developed economic level. China-PAR fairly estimated 5-year ASCVD risk in low- and medium-risk but underestimated the high-risk participants with older age, socioeconomically disadvantaged, higher prevalence of diabetes mellitus, and with higher proportion of lipid-lowing medications. It implied that this guideline among southern Chinese population might not be appropriate for the individuals with established CVD or high-risk risk factors.

The risk of recurrent CVD events and death increases with age, smoking, hypertension, and diabetes mellitus in individuals with established CVD [11]. Symptoms of a heart attack includes pain or discomfort in the centre of the chest, pain or discomfort in the arms, the left shoulder, elbows, jaw, or back. Moreover, individuals may experience difficulty in breathing or shortness of breath; nausea or vomiting; light-headedness or faintness; a cold sweat; and turning pale. Other common symptom of a stroke is sudden weakness of the face, arm, or leg, most often on one side of the body [12]. Recurrent CVD events are also more likely to occur in patients who survive non-ST-elevation myocardial infarction (NSTEMI) [13-15]. However, limited data exist on long-term secondary atherothrombotic risk stratification. A recently observational study conducted to validated TRS 2P among post-NSTEMI patients for risk of recurrent cardiovascular events among Hong Kong population for its application in real world [16]. Results showed that TRS 2P was able to stratified the patient with high risk of recurrent CVD, however, the outcomes didn't include full spectrum of composite CVD, such as non-fatal ischaemic stroke.

Since P-CARDIAC is a newly developed model, it had not been sufficiently validated externally in the real-world yet, and the performance in diverse geographical populations for secondary prevention is in doubt. For the availability and feasibility of P-CARDIAC performance in local clinical setting, a validation in the local clinical setting is necessary to be conducted.

2. Research aims and objectives

The objective of this study is to evaluate the accuracy of the P-CARDIAC performance in practice among a large-scale Hong Kong population in medicine specialist outpatient clinic (SOPC) and cardiac clinic managed by the Hospital Authority (HA) in Hong Kong. We anticipate that the results may help us to facilitate P-CARDIAC in clinical setting and provide more practical information with the development of P-CARDIAC. The main objectives are:

1) To map P-CARDIAC score with 2-year CVD outcomes using electronic health record and patients reported symptoms.
2) To map P-CARDIAC score with clinician's judgement, international treatment guideline identifies thresholds.

3. Methods

3.1 Study design

The study has two parts:

Part 1

Part 1 is a prospective population-based cohort study which will address objective 1. We hypothesize that patients with a higher frequency or larger number of types of cardiac event symptoms are likely to have a higher P-CARDIAC risk score. The correlation between P-CARDIAC risk score and the increasing in number of types and in number of times a patient reported to have heart disease symptoms, e.g. chest pain, shortness of breath, fatigue, swelling in legs, ankles or feet will be examined [12]. The results reassure cardiologists that the P-CARDIAC risk score is sensitive to the heart disease symptoms. Patients who are at a higher risk of recurrent CVD event are likely to have more medications and to have a sooner follow up appointment. At such, we hypothesize a positive relationship between P-CARDIAC risk score and the number of class of medications prescribed and a negative correlation between P-CARDIAC risk score and the time to next SOPC follow up appointment.

The application of P-CARDIAC will be firstly introduced to our collaborated clinicians. Collaborating clinicians will help to randomly select patients who visited medical / cardiology SOPC to identify eligible patients with stabled CVD risk or of low CVD risk. If time allows, the participating clinicians may introduce the project to the participants during the consultation. Research personnel will then recruit and obtain the written consent from the eligible patients, our researchers (not limited to: research personnel, pharmacy or nursing students or study helpers) will ask any CVD related symptom such as chest pain and its frequency, and medication for the baseline records through structured interview using a questionnaire. We will also obtain consent to collect their personal identifier (such as Hospitalization Number, Accidence and Emergency Number, HKID or Out-Patient appointment reference) to access their electronic health records (eHR) from the Clinical Management System and Hospital Authority Clinical Data Analysis and Reporting System (CDARS) and inputted to the P-CARDIAC model for recurrent CVD risk calculation. The medical records of recruited patients will be collected every 3 to 6 months after until end of follow-up period to observe any CVD outcomes and relevant change in risk factors.

Part 2

Part 2 of the study will involve Delphi technique [17] to determine P-CARDIAC risk thresholds with reference to current clinical management guideline such as American Heart Association (AHA)/American College of Cardiology (ACC) Multisociety

Guideline (AHA/ACC) and European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS) (ESC/EAS), and this part will not involve patient contact. This part of the study will address Objective 2. We will compare the P-CARDIAC risk score and the clinician's rating of a patient's recurrent CVD risk on a 10% random subgroup of patients from part 1 in a "silent deployment" [18] approach, ie, an individual patient's P-CARDIAC risk score will not be communicated to cardiologists. Cardiologists will be invited to rate if they think the patient is at high or low risk according to existing risk scores and the proposed treatment management and target based on treatment guidelines based on their experience. Each patient profile will be reviewed by at least 2 cardiologists and discussed to reach a consensus. We will then evaluate the sensitivity of P-CARDIAC risk score with clinician's judgment. Prior to the commencement of Part 2, researchers will collect insights from cardiologists regarding the interpretability of performance metrics and acceptable threshold of model performance for clinical practice. In addition, through this exercise, HKU team will assist cardiologists to draft a guideline for follow up for patients at different risk level.

A total of 10% of the recruitment patients will be included in Part 2. Clinicians are invited to review the medical records of the sub-cohort at recruitment to determine their CVD risk using existing risk scores such PCE and Thrombolysis in Myocardial Infarction (TIMI) Risk Score for Secondary Prevention (TRS-2P). The P-CARDIAC score will not be provided to ensure objective assessment. Clinicians will then be asked to map to relevant clinical management plan according to local or international guidelines such as AHA/ACC and ESC/EAS. A Delphi study will then be conducted to align the clinical judgement with the P-CARDIAC scores to establish the risk thresholds.

3.2 Recruitment

Inclusion and exclusion

Individuals aged 18 – 80 years old who visit the medical or cardiology SOPC clinics at the Queen Mary Hospital (QMH), and experienced prior CVD event (such as peripheral artery diseases, coronary heart diseases, myocardial infarction, stroke, revascularization) before the scheduled SOPC clinic visit date (i.e. index date) are eligible to participate in this study. Participants will be also able to comprehend English or Traditional Chinese as well. People with history of dementia, psychiatric diseases or physical impairment will be excluded in this study as they are less likely to benefit from the medication education and counselling service. Part 1 and 2 will share the same inclusion criteria as a random subset of Part 2 participants will be selected from Part 1.

We will adopt a 2-steps approach to obtain informed consent. The research team will, in collaboration with the patient support groups such as Care For Your Heart and the Hong Kong Stroke Association, to disseminate an study invitation pack to their members. The study invitation pack will include a flyer with the above inclusion and exclusion criteria

clearly outlined, a study consent form, an information sheet and a stamped to be returned to the research team. A QR code link to the study registration site will also be included in the invitation pack for participants to register them intend to participate in this study. In this first step of recruitment, we will obtain their consent to collect their contact information and the next SOPC visit for scheduling.

The research team will also prepare a flyer, which may also serve as an advertisement in online platform, e.g. Facebook to recruit participants.

The research team will recruit also use the flyer to recruit at the medicine SOPC or cardiac clinics of the QMH between 1 February 2024 to 31st January 2025 (up to 12 months) with a possibility to extend the recruitment period depending on recruitment progress. The follow-up period will be up to 2 years after recruitment (31st January 2027). During their SOPC visit, our research personnel will perform the 2nd step of recruitment which is to obtain participants' written consent to collect their personal identifier to access their eHR from the Clinical Management System and Hospital Authority Clinical Data Analysis and Reporting System (CDARS) and inputted to the P-CARDIAC model for recurrent CVD risk calculation. An interview on their symptoms will be collected through questionnaire. We will assess CDARS to conduct pragmatic prospective follow-up of the recruitment patients.

3.3 Sample size consideration

Considering the services allocation and the traffic of SOPC appointment in practice, we aim to recruit no more than 800 participants at Queen Mary Hospital for Part 1. 10% of the participants, i.e. no more than 80 patients will be included in Part 2.

3.4 Informed Consent

Informed consent will be obtained in written form at recruitment and before the start of each part of the study. Informed consent can be obtained by either clinician or research assistant if necessarily. Consent form will be available in English or Traditional Chinese with detailed information of the study including study aim, procedures and usage of collected data. Participants will be given time to read through and understand the purpose of the study before giving their consent.

All research personnel will be provided relevant training on the study objectives, questionnaire, information sheet and consent form before commencement of duty.

3.5 Data collection and outcome

With the use of their personalized identifier (such as Hospitalization Number, Accidence and Emergency Number, HKID or Out-Patient appointment reference), we will retrieve their eHR from CDARS. Health parameters from eHR will be extracted for the P-CARDIAC risk calculation. All participants will be followed up to 2 years in Part 1 and 2. Data will be retrieved from CDARS every 3 to 6 months until end of follow-up period. The outcomes will include the recurrent CVD event between the follow-up period, including a composite of coronary heart disease, ischemic or hemorrhagic stroke, peripheral artery disease, and revascularization which will be compliance with P-CARDIAC outcomes by ICD 9 codes. The ICD 9 code confirmation will be considered as the standard and confirmed by the independent clinicians for adjudication to ascertain the date of diagnosis for conflicting records. Any cause-specific (cardiovascular-related or other) death will be confirmed by death certificate in the health information system. During the followed-up period, other risk parameters, such as hospitalization, healthcare service utilization and medication records, will also be extracted from CDARS. Other potential outcome may include but limited to parameters included in P-CARDIAC such as blood pressure, blood cholesterol, blood glucose, medications, and healthcare utilization (such as A&E admission, hospitalization, SOPD visit not limited to medical / cardiology clinics). Patient reported symptoms will also be collected in Part 1.

3.6 Data analysis

Continuous and categorical baseline characteristics of participants will be summarized by mean (standard deviation) or numbers (percentage), respectively. Student's t-test or one-way ANOVA was used to compare baseline clinical characteristics. Considering differences in health status and survival for participants with and without missing data, our main analysis will be conducted on the dataset by imputation to reduce selection bias. We used the multiple imputation by chained equations (MICE) method as the development of P-CARDIAC to replace missing values.

For each participant, observed recurrent CVD risk will be calculated using the Kaplan-Meier method with 95% confidence interval (CI) and predicted recurrent CVD risk will be calculated using original model among the whole study population. Model performance will be assessed by Harrell’s C-statistics for discrimination and Hosmer-Lemeshow chi-square for calibration. Calibration was also evaluated using calibration slope, a shrinkage factor that closed to 1 for well-calibrated equations and visualized by plotting the predicted risk against the observed risk for each decile of the predicted risk.

In addition, to assess the robustness of the results, we excluded participants with missing data of any predictors included in model and then evaluated the performance of P-CARDIAC in sensitivity analyses. We further restricted the analyses to participants without any predictors in mandatory field at baseline to evaluate their effects on the discordance between observed and predicted risk [23].

All analyses will be conducted using R (version 4.1.2) or statistical software deemed appropriate with a statistical significance level of $P < 0.05$.

3.7 Data handling and record keeping

Given the personal data will be collected, protective measures will be done to prevent any data leakage. To protect all the collected data, data will be stored in password-protected server located in a locked server room within the HKU premise and identifiable information will be stored separately from medical records to minimize risk of confidentiality breach. Only authorized researchers have the access to the datasets. The collected data is confidential for this study research use only. The data will be kept for five years after the completion of study in case there might have a need of analyzing the collected data in the future.

3.8 Ethical Consideration

This study protocol will seek the ethics approval by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster. Participant's confidentiality and data leakage will be the major risk and concern of the study. Although there are personal data will be collected, protective measures will be done to prevent any data leakage. Besides, informed consent process will be done before the start of the study in compliance of the Declaration of Helsinki. To protect the participant, all the data collected in this study will be kept confidential throughout the whole study period. All the participants are voluntary to join and able to drop out any time during the study without any prejudice. All data will be destroyed according to HA guidelines five years after the completion of study.

3.9 Risk management

We believe the imposed risk to patients will be minimal for Part 1 and 2 as no intervention will be given to patients.

3.10 Funding

This study received no grant by any funding agencies.

4. Limitation

Limitations should be acknowledged and include the limited length of follow-up in the validation. The long-term incidence of CVD would have been the optimal primary outcome measure for our study. However, time and budgetary restrictions prevented us from using this 10-years endpoint. The study period was 2 years prospective followed-up cohort, we understand the loss-to followed-up rate is inevitable. Nevertheless, as for all registry and multi-center studies, patient inclusion is always subject to bias. We will try to review hospital records and laboratory results carefully to ascertain the diagnosis standard for our collaborated clinicians but potential confounding still cannot be excluded.

Reference

1. Heron, M., *Deaths: leading causes for 2008.* National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 2012. **60**(6): p. 1-94.
2. Roth, G.A., et al., *Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017.* The Lancet, 2018. **392**(10159): p. 1736-1788.
3. Zhou, M., et al., *Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017.* The Lancet, 2019. **394**(10204): p. 1145-1158.
4. *Public urged to lead healthy lifestyle to guard against cardiovascular diseases*. 2022.
5. *Cardiovascular diseases: Avoiding heart attacks and strokes*. 2015.
6. Grant, S.W., G.S. Collins, and S.A. Nashef, *Statistical Primer: developing and validating a risk prediction model.* European Journal of Cardio-Thoracic Surgery, 2018. **54**(2): p. 203-208.
7. Yang, X., et al., *Predicting the 10-year risks of atherosclerotic cardiovascular disease in Chinese population: the China-PAR project (prediction for ASCVD risk in China).* Circulation, 2016. **134**(19): p. 1430-1440.
8. Bohula, E.A., et al., *Atherothrombotic risk stratification and the efficacy and safety of vorapaxar in patients with stable ischemic heart disease and previous myocardial infarction.* Circulation, 2016. **134**(4): p. 304-313.
9. Hageman, S.H., et al., *Estimation of recurrent atherosclerotic cardiovascular event risk in patients with established cardiovascular disease: the updated SMART2 algorithm.* European Heart Journal, 2022. **43**(18): p. 1715-1727.
10. Zhiting, G., et al., *Cardiovascular disease risk prediction models in the Chinese population-a systematic review and meta-analysis.* BMC Public Health, 2022. **22**(1): p. 1-16.
11. Lin, F.-J., et al., *Residual risk factors to predict major adverse cardiovascular events in atherosclerotic cardiovascular disease patients with and without diabetes mellitus.* Scientific reports, 2017. **7**(1): p. 9179.
12. *Cardiovascular diseases (CVDs)*. 2021; Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
13. Stone, N.J., et al., *2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines.* Circulation, 2014. **129**(25_suppl_2): p. S1-S45.
14. Damman, P., et al., *2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: comments from the Dutch ACS working group.* Netherlands Heart Journal, 2017. **25**: p. 181-185.
15. Members, W.C., et al., *2015 ACC/AHA focused update of secondary prevention lipid performance measures: a report of the American College of*

Cardiology/American Heart Association Task Force on Performance Measures. Circulation: Cardiovascular Quality and Outcomes, 2016. **9**(1): p. 68-95.
16. Huang, D., et al., *TIMI risk score for secondary prevention of recurrent cardiovascular events in a real-world cohort of post-non-ST-elevation myocardial infarction patients.* Postgraduate Medical Journal, 2019. **95**(1125): p. 372-377.
17. Barrett, D. and R. Heale, *What are Delphi studies?* Evidence-based nursing, 2020. **23**(3): p. 68-69.
18. Antoniou, T. and M. Mamdani, *Evaluation of machine learning solutions in medicine.* Cmaj, 2021. **193**(36): p. E1425-E1429.
19. Leung, J.T., et al., *Non-inferiority trials in cardiology: what clinicians need to know.* Heart, 2020. **106**(2): p. 99-104.
20. Xu, H., et al., *Impacts of clinical pharmacist intervention on the secondary prevention of coronary heart disease: a randomized controlled clinical study.* Frontiers in pharmacology, 2019. **10**: p. 1112.
21. Henderson, A.R., *Evidence-Based Medicine—How to Practice and Teach EBM. DL Sackett, WS Richardson, W. Rosenberg, and RB Haynes. New York: Churchill Livingstone, 1997, 250 pp. Paperback, $24.99. ISBN 0-443-05686-2*. 1997, Oxford University Press.
22. Zhong, B., *How to calculate sample size in randomized controlled trial?* Journal of thoracic disease, 2009. **1**(1): p. 51.
23. Muntner, P., et al., *Validation of the atherosclerotic cardiovascular disease Pooled Cohort risk equations.* Jama, 2014. **311**(14): p. 1406-1415.