

BMJ Open Predictors on outcomes of cardiovascular disease of male patients in Malaysia using Bayesian network analysis

Nurliyana Juhan ¹, Yong Zulina Zubairi ², Ahmad Syadi Mahmood Zuhdi ³, Zarina Mohd Khalid ⁴

To cite: Juhan N, Zubairi YZ, Mahmood Zuhdi AS, *et al*. Predictors on outcomes of cardiovascular disease of male patients in Malaysia using Bayesian network analysis. *BMJ Open* 2023;**13**:e066748. doi:10.1136/bmjopen-2022-066748

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-066748>).

Received 19 July 2022
Accepted 30 August 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia

²Institute for Advanced Studies, University of Malaya, Kuala Lumpur, Malaysia

³Department of Medicine, University of Malaya Medical Centre, Kuala Lumpur, Malaysia

⁴Department of Mathematical Sciences, Universiti Teknologi Malaysia, Skudai, Malaysia

Correspondence to

Professor Yong Zulina Zubairi; yzulina@um.edu.my

ABSTRACT

Objectives Despite extensive advances in medical and surgical treatment, cardiovascular disease (CVD) remains the leading cause of mortality worldwide. Identifying the significant predictors will help clinicians with the prognosis of the disease and patient management. This study aims to identify and interpret the dependence structure between the predictors and health outcomes of ST-elevation myocardial infarction (STEMI) male patients in Malaysian setting.

Design Retrospective study.

Setting Malaysian National Cardiovascular Disease Database–Acute Coronary Syndrome (NCVD–ACS) registry years 2006–2013, which consists of 18 hospitals across the country.

Participants 7180 male patients diagnosed with STEMI from the NCVD–ACS registry.

Primary and secondary outcome measures A graphical model based on the Bayesian network (BN) approach has been considered. A bootstrap resampling approach was integrated into the structural learning algorithm to estimate probabilistic relations between the studied features that have the strongest influence and support.

Results The relationships between 16 features in the domain of CVD were visualised. From the bootstrap resampling approach, out of 250, only 25 arcs are significant (strength value ≥ 0.85 and the direction value ≥ 0.50). Age group, Killip class and renal disease were classified as the key predictors in the BN model for male patients as they were the most influential variables directly connected to the outcome, which is the patient status. Widespread probabilistic associations between the key predictors and the remaining variables were observed in the network structure. High likelihood values are observed for patient status variable stated alive (93.8%), Killip class I on presentation (66.8%), patient younger than 65 (81.1%), smoker patient (77.2%) and ethnic Malay (59.2%). The BN model has been shown to have good predictive performance.

Conclusions The data visualisation analysis can be a powerful tool to understand the relationships between the CVD prognostic variables and can be useful to clinicians.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Data from 18 major hospitals in Malaysia were used to conduct this study, which involved a sizable number of ST-elevation myocardial infarction male patients.
- ⇒ The patients represented a variety of ethnicities, including Chinese, Indian, and Malay, three of Asia's major ethnic groups.
- ⇒ The relationships between the cardiovascular disease prognostic variables in Malaysia were analysed and visually presented.
- ⇒ Only in-hospital mortality was analysed as follow-up data is not available.

INTRODUCTION

Despite extensive advances in medical and surgical treatment as well as biomedical innovation, cardiovascular disease (CVD) remains the leading cause of mortality worldwide.^{1 2} With an alarmingly high figure, CVD affects roughly 422 million people and is responsible for approximately 18 million fatalities yearly.³ While CVD morbidity and mortality have stabilised in high-income countries, CVD deaths and risk factors or as they are commonly referred to in this setting, predictors burden are increasing in low-income and middle-income countries.⁴ Clearly, there is a persistent need for interventions targeting CVD predictors, particularly cost-effective interventions that can be effectively scaled up for population-wide distribution.

An increasingly used method in the medical field, especially cardiology known as Bayesian network (BN), was a perfect solution.^{5–9} BN was first introduced by Judea Pearl in 1985 and it was known as a powerful method for knowledge representation under the influence of uncertainty.¹⁰ It was a combination of two different mathematical areas namely graph theory and probability theory.¹¹ Both categorical and numerical data were involved in these

previous studies. However, all of these studies have been based on worldwide CVD samples such as China, USA, United Kingdom, Egypt and Canada and BN is relatively less explored in Malaysian CVD dataset. Also, most of these BN studies do not contain indication on the degree of confidence or strength of each arc in the network and BN model validation was not performed.^{5 6 8 12} Moreover, no specific details are given on the BN algorithm used in their study.^{8 12} In order to address this limitation, a graphical model based on BN will be applied to discover the dependence structure between variables in Malaysian CVD dataset and the conditional independence of a variable or groups of variables from a given variable or variables. This is the first study to describe the relationships between the CVD prognostic variables in Malaysia using visualisation analysis and has great potential to be applied in other medical studies.

In this study, BN model is used for building clinical decision support system to help in identify the prognostic variables. The BN approach has both the elements of uncertainty and causality in which it becomes both expert knowledge and data-driven modelling.¹³ The analysis is performed within a graphical probabilistic model framework. There are two main steps to obtain a BN model from the data. The first step is known as structural learning where the graph contains the conditional independencies of the data is determined followed by a second step known as parameter learning which estimates and updates of the parameters of the local distributions.¹⁴ In this study, score-based algorithm was chosen to be applied in the structural learning in the motivation of Scutari *et al* in 2018 which stated that by using both simulated and real-world data, score-based algorithms appeared to be more efficient or more sensitive to errors than constraint-based algorithms while hybrid algorithms are not much faster or more accurate than constraint-based algorithms.¹⁵ Hill-climbing algorithm of the score-based algorithm was applied in the structural learning.

The motivation of the study is to provide a visualisation of data-driven analysis. It is hoped that the findings are useful to primary care clinicians in fields such as family practice, internal medicine, nurse practitioners and physician assistants, who may benefit from data-driven evidence to identify CVD risk factors, or as they are commonly referred to in this setting, predictors, are most effectively diagnosed and managed. Specialists may benefit, as a specialist in one part of preventive cardiology may lack expertise in other fundamental components.¹⁶ Furthermore, most studies have shown that the majority of patients with CVD have several CVD predictors, necessitating a multimodal.^{17 18} Patients with or at risk of CVD will benefit from a holistic approach to CVD risk reduction, with sufficient attention paid to all relevant CVD predictors.¹⁹ It may therefore be beneficial for physicians to have a general understanding of the fundamental concepts that apply to the various CVD predictors that frequently coexist in the same patient who has or is at risk for CVD.

Comprehensive and systematic analysis of long-term trends and patterns in worldwide CVD is critical for guiding public policy and providing decision makers with benchmarks.^{1 16} Understanding how significant each predictor is will help to improve assessment processes, clinical decisions and trial designs.¹ Therefore, the main aim of the study is to identify and interpret the dependence structure between the predictors and health outcomes of ST-elevation myocardial infarction (STEMI) male patients using BN.

MATERIALS AND METHODS

Source of data

From the year 2006 to 2013, a total of 7180 anonymised patient data for patients diagnosed with STEMI from 18 major hospitals in Malaysia were obtained from the National Cardiovascular Disease Database-Acute Coronary Syndrome (NCVD-ACS) registry. STEMI was chosen since it is the deadliest kind of acute coronary syndrome. STEMI in this context is defined as persistent ST segment elevation exceeding 1 mm in two consecutive electrocardiographic leads or the occurrence of a new left bundle branch block in the presence of positive cardiac markers. From the time a patient with STEMI was admitted to the hospital until he or she was discharged, data was recorded. Variables on demographics, risk factors, comorbidities, clinical presentation, in-hospital treatment and clinical outcome of the patients were all recorded. The clinical presentation, called the Killip class, was classified into four subclasses. The Killip classification predicts how likely it is that people with an acute myocardial infarction will live for 30 days, with Killip IV being the most likely to die.²⁰ This study used only the cases in a dataset for which there were no missing values on any of the variables which is also known as complete case analysis. Furthermore, only in-hospital mortality was analysed as insufficient follow-up data prevented long-term analysis.

Statistical methods

For BN analysis, the data were divided into learning and test datasets. The model was built using the learning dataset. Test dataset was used for model validation and to check the performances of model. The learning and test partitioning used in this study were based on the typical ratio of 70:30.^{21 22} BN is a directed acyclic graph (DAG) which consists of nodes and directed line segments or arcs between pairs of the nodes. Each node represents a variable, which has a different probability distribution, and each arc represents the joint or inference dependency between two nodes. [Table 1](#) shows the description of the variables in the study.

When new information is added, the BN is used to make inferences, that is, to determine new probabilities. Therefore, it is important to understand the basic concepts for inference flow. Two variables A and B in a BN are d-separated if, for every possible path between them, there is an intermediate variable C such that the connection is

Table 1 Description of male patients' dataset variables

Characteristic	Variable (abbreviation)	Category
Demographic	Ethnicity (ethnicity)	Malay
		Chinese
		Indian
		Others
	Age group (age)	<65
		≥65
Risk factor	Diabetes mellitus (diabetes)	No
		Yes
	Hypertension (hypertension)	No
		Yes
	Smoking status (smoking)	Never
		Active/former
Dyslipidaemia (dyslipidaemia)	No	
	Yes	
Family history of CVD (FmlyHist)	No	
	Yes	
Comorbidities	MI history (MI)	No
		Yes
	Chronic lung disease (chronic lung)	No
		Yes
	Cerebrovascular disease (cerebro)	No
		Yes
Peripheral vascular disease (peripheral)	No	
	Yes	
Renal disease (renal)	No	
	Yes	
Clinical presentation	Killip class (Killip class)	Class I
		Class II
		Class III
		Class IV
Treatment	Percutaneous coronary intervention (PCI)	No
		Yes
	Cardiac catheterisation (CCath)	No
		Yes
Outcome	Patient status (patient status)	Alive
		Dead

either serial ($A \rightarrow C \rightarrow B$ or $A \leftarrow C \leftarrow B$) or diverging ($A \leftarrow C \rightarrow B$) and C is instantiated or (ii) the connection is converging ($A \rightarrow C \leftarrow B$) and neither C nor any of C's descendants have received evidence.²³ To analyse the BN, it is crucial to know when influence flows from a node A to a node B via a node C. When this happens, the A node is said to be active.²³

In order to obtain the DAG, the structure of the data needs to be determined. All variables are considered to be

included in the BN model. The Bayesian model relies on prior knowledge of the phenomenon we are modelling to decide which arcs are present in the graph and which are not. Also, as a rule on the model selection process, arcs that point from a later to an earlier node were blacklisted. A score-based algorithm known as hill-climbing is additionally applied in this study.²⁴ The hill-climbing algorithm began with an empty DAG (without any arc), and it then continuously added, deleted or reversed one edge at a time until the Bayesian Dirichlet equivalent uniform score can no longer be improved.²⁵

A method known as model averaging was implemented on top of the structural learning in order to obtain the best structure and a robust model.²⁶ In this study, the method used to perform model averaging is using bootstrap resampling. The basic idea of bootstrap is to randomly draw datasets with replacement from the training dataset D , with each sample the same size as the original set, that is, N . This is performed B times, producing B bootstrap replicates.²⁷ Then the averaged network structure was built using the arcs present in at least 85% of the networks which provides a measure of the strength of each arc and establishes its significance given a threshold of 85%. By learning the network structure, this study is able to identify which variables have a direct effect on the risk of mortality of CVD and which variables at the same time have their effect on the tendency intermediated by the earlier variables.

After learning the BN's structure from the data, the task of estimating and updating the parameters of the global distribution, also known as the joint probability distribution, is easily done by breaking it down into local distributions, each of which only involves a node and its parents.²⁶ For the discrete BN in this study, the parameters to estimate are the conditional probabilities in the local distributions. Suppose X_i is the i th random variable containing r_i discrete states and X_{ik} is the k th state of the random variable X_i . Consider a BN consisting a set of nodes or random variables $X = \{X_1, X_2, \dots, X_N\}$, a pair of (G, Θ) where G denotes DAG and Θ is the set of conditional probabilities with $\Theta = (\theta_{ijk} : \forall ijk)$, where θ_{ijk} is the joint probability $\theta_{ijk} = P(X_i = X_{ik} | \text{parent node}(X_i) = p_{ij}^k)$ where p_{ij}^k is the parent node of X_i from G and p_{ij} is the j th combination of parents nodes of X_i .

Model validity and performance are evaluated through measures of discrimination and calibration. Discrimination is the ability of the model to distinguish and segregate between the dead and alive patients. Calibration is used to assess the agreement between predicted and observed risk of mortality obtained from a model. All the algorithms involved were performed using R software, while the advanced graphical structures were plotted using Netica (Norsys Software Corp.).^{28 29}

Patient and public involvement

There was no patient or public participation in the development of this study's research question and outcome.

All data were obtained retrospectively from the Malaysian NCVD-ACS. Result of this study will not be disseminated to study participants.

RESULTS

By performing the bootstrap resampling approach in the model averaging, strength and direction of each arc in the averaged network structure or known as the final DAG for male patients are obtained and given in online supplemental table. Arc is considered to be significant if the strength value greater than or equal to 0.85 and the direction value appears to be greater than or equal to 0.50.²⁶ Out of 240 possible arcs, only 25 arcs are significant in the final model for male patients as in online supplemental table. Since all the direction values are greater than 0.5, it can be concluded that the direction of the arcs is well established.

Then, the final DAG for male patients which consisted of 25 arcs as in online supplemental table is obtained and shown in online supplemental figure. The final DAG reflected the causal order and dependencies of each variable. Quite a few interesting features can be observed. From online supplemental figure, the variables age, Killip class and renal disease can be classified as the key predictors as they are the most influential variables directly associated with the outcome which is the patient status. The network analysis also showed widespread probabilistic associations between the key predictors and the remaining variables. Age is directly associated with chronic lung disease, family history of CVD and hypertension, whereas Killip class are associated with MI and renal disease. Renal disease appeared to have further association with cerebrovascular disease and has a direct effect from hypertension.

Dyslipidaemia, smoking and age are associated to patient status through hypertension, diabetes mellitus and renal disease. In this situation, dyslipidaemia, smoking and age are said to have common effect on three variables namely hypertension, diabetes mellitus and renal disease; and once any of these three variables is removed, the connection thru this trail is broken. However, there is also other possible trails such as dyslipidaemia → MI → Killip class → patient status. On the other hand, the variables ethnicity, dyslipidaemia and hypertension have direct connection on diabetes. It is worthwhile noted that MI has caused percutaneous coronary intervention (PCI) and chronic lung disease has caused cardiac catheterisation which also led to PCI. Ethnicity and age are known as root variables as they have no parents.

After the final DAG is obtained, the BN for male patients is then plotted using Netica as shown in figure 1. The software transformed the final DAG structure as shown in online supplemental figure in a probabilistic inference model which determined beliefs for the values of the rest of the variables in the structure. High likelihood values are observed for patient status variable stated alive (93.8%), Killip class I on presentation (66.8%), patient

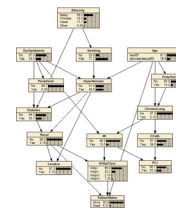


Figure 1 Bayesian network for the study of features relationships among male patients.

with age less than 65-year-old age group (81.1%), smoker patient (77.2%) and ethnic Malay (59.2%). On another note, there are low likelihoods for patients having renal disease (3.2%), cerebrovascular disease (2.7%), chronic lung disease (2.1%) and peripheral disease (0.3%). It also shows an almost similar likelihood in patients with and without hypertension.

As for parameter learning, for each node, a conditional probability distribution is obtained. Table 2 showed an example of conditional probability distribution that involved the three key predictors for the BN model for male patients namely renal disease, Killip class and age group. Table 2 suggested that the conditional probability of death given that the patient has the renal disease is higher than those without it for all the Killip class presentations and for both age groups.

The BN model's validity and performance are assessed using another 2154 male patient datasets. Receiver operating characteristics (ROC) curve is used, which generated four statistical measures, namely the AUC, sensitivity, specificity and accuracy, as shown in table 3. Figure 2 illustrated the ROC curve of BN model for male patients. From table 3, it can be seen that the AUC with the sensitivity and specificity values for the male BN model are highly statistically significant. The sensitivity of the BN model reached 100%. This suggests that the BN model has good predictive performance and has correctly classified the different features. Additionally, this model demonstrated a high degree of overall accuracy.

DISCUSSION

This study demonstrates the potential of using BN in epidemiological studies, specifically in CVD. The BN has set up the relationships between features using dependency and conditional independence.²³ In the structural learning, unique and interesting network structures have been obtained for BN model for male patients. In the BN model, the variables ethnicity, dyslipidaemia and smoking are d-separated of patient status feature by hypertension feature. The 'd' in d-separation stands for directional.²⁶ Therefore, if two variables are d-separated relative to a set of variables C in a DAG, then they are conditionally independent on C in all probability distributions such a graph can represent. In general, two variables A and B are conditionally independent on C if knowledge about A gives no extra information about B once knowledge of C is known. Considering the local Markov property of a node which

Table 2 Expected values of probabilities for patient status feature conditional on combinations of its parent values, renal disease, Killip class and age features

Renal	Killip class	Age	Patient status=Dead	Patient status=Alive
No	I	<65	0.0167	0.9833
Yes	I	<65	0.1277	0.8723
No	II	<65	0.0455	0.9545
Yes	II	<65	0.1080	0.8920
No	III	<65	0.1430	0.8570
Yes	III	<65	0.2740	0.7260
No	IV	<65	0.2836	0.7164
Yes	IV	<65	0.2864	0.7136
No	I	≥65	0.0530	0.9470
Yes	I	≥65	0.1887	0.8113
No	II	≥65	0.0934	0.9066
Yes	II	≥65	0.2241	0.7759
No	III	≥65	0.3417	0.6583
Yes	III	≥65	0.5000	0.5000
No	IV	≥65	0.4559	0.5441
Yes	IV	≥65	0.6988	0.3012

is each node is conditionally independent of its non-descendants,²⁴ in this study for example given the parent of smoking feature, which is composed of ethnicity and taking into account the local Markov condition, smoking feature remains independent of its non-descendants. In other words, smoking feature is independent of all other variables except for hypertension, diabetes, cerebrovascular disease, renal disease, Killip class and patient status features, respectively.

Based on the structure of a BN, a set of features with the strongest influence can be established by using the global Markov property which will be composed by the Markov blanket of this specific feature. The Markov blanket term was created by Judea Pearl in 1988 where the Markov blanket of a node A is the set consisting of the parents of A, the children of A and all the other nodes sharing a child with A sometimes known as spouses.¹⁰ All these nodes shield node A from the rest of the network.³⁰ In this study, the focus is mainly on the patient status feature. Nevertheless, the BN gave a characterisation of the whole set of variables; for example, in the BN model for male patients, the Markov blanket for Killip class feature is

Table 3 Bayesian network model validation and performance measures

Performance measure	BN model
AUC	0.9730
Sensitivity	1.0000
Specificity	0.8630
Accuracy	0.9583

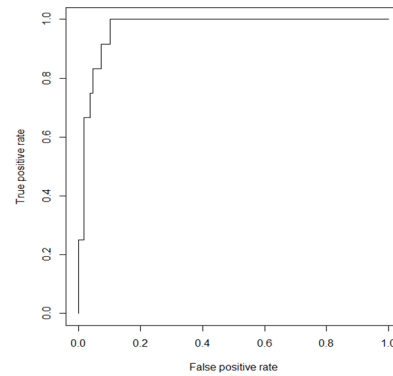


Figure 2 Receiver operating characteristics curve of Bayesian network model for male patients.

given by MI history (parent), renal disease (parent), age (spouse) and patient status (child); with these four features, Killip class feature is independent of the remaining ones. Also, it could be used to find the combination of states which maximise or minimise a specific state of Killip class feature. Age is directly associated with chronic lung disease, family history of CVD and hypertension, whereas Killip class is associated with MI and renal disease. It is also noted that renal disease has further association with cerebrovascular disease and has a direct effect from hypertension.

Dyslipidaemia, smoking and age are associated to patient status through hypertension, diabetes mellitus and renal disease. In this situation, dyslipidaemia, smoking and age are said to have common effect on these three variables namely hypertension, diabetes mellitus and renal disease; once any of these three variables is removed, the connection thru this trail is broken. However, there is also other possible trail such as dyslipidaemia → MI → Killip class → patient status. On the other hand, ethnicity, dyslipidaemia and hypertension have direct connection on diabetes. MI has caused PCI and chronic lung disease has caused cardiac catheterisation which also led to PCI. Ethnicity and age are known as root variables as they have no parents. A few studies among the Japanese population state that renal disease and Killip class appear to be related, where acute renal diseases are associated with older age and Killip class II and higher.^{30–32} This trend is similarly observed in Malaysian STEMI male patients in this study. Also, another study suggests that renal disease occurred more frequently with increasing Killip class, which was then associated with increased in-hospital mortality.^{33 34}

Additionally, other studies have stated that for Chinese and Danish populations, patients with renal insufficiency are older and more likely to have hypertension and have presented with a higher Killip class.^{35 36} In the Indian, Turkish and Arab populations, patients with a higher Killip class at the late age are more likely to have diabetes mellitus, hypertension and renal failure.^{37–40} This means, these previous studies are all in agreement with the results of this study, where four variables, namely age, hypertension, renal disease and Killip class, are highly related



to each other. Also, a few studies in Korea, Spain and Malaysia have suggested that the strongest determinants of mortality among CVD patients appear to be Killip class and age of the patient.^{41–43} The findings of these previous studies are in agreement with the proposed BN model for male patients in this study where Killip class, renal disease and age group are found to be the most influencing variables and can be classified as the key outcome predictors in the BN model.

The difference of this approach with regard to other BN CVD studies in the literature is that we employed bootstrap resampling approach which has helped to determine those features with great influence in each of the diagnostic features. As mentioned before, the bootstrap resampling approach is adopted in the model averaging. The advantage of adopting bootstrap approach on top of the structural learning is that it induces better structures from the data and produced a robust Markov blanket of a given node, thus confirming the stability of the averaged network. Considering the local Markov property and the Markov blanket property, some features are optimised. The results of this study are in accordance with other previous studies where model averaging is known to produce better predictive performance as opposed to a single, high-scoring BN.^{44 45} In addition, the bootstrap resampling approach applied in the model averaging is able to estimate the level of confidence or the strength of each arc. Although not shown, remarkably, in model averaging the results of the analysis did not change even though the threshold was lowered from 85% to 50%. This indicates that the results are not sensitive to the threshold value. The main advantage of BN in this study is the ability to represent the causal relationships. The graphical model obtained is capable of displaying cause-effect relationships clearly and intuitively. The approach leverages on visualisation to illustrate findings in data-driven analytics. Moreover, BN can be used for answering clinical questions based on unobserved evidence since the probability distributions can be automatically updated when new patient information is added in an appealing way.⁴⁶ However, it is worthwhile to note that this study used only complete cases. Further studies can be carried out to take into account of missing value. Another limitation of this study is it only includes in-hospital mortality.

CONCLUSION

In this study, BN analysis has been performed in order to produce an intuitive, interesting, graphical representation of the probabilistic causal dependencies and conditional independencies between predictors and health outcomes of STEMI patients. As a result, the relationships between 16 features in the domain of CVD are obtained and illustrated in visual form. Age group, Killip class and renal disease are classified as the key outcome predictors in BN model for male patients as they are the most influencing features and has shown to have direct effect on the outcome of patient status. In conclusion, BN is

a graphical-based structure of a joint multivariate probability distribution which provides guidance to expert in making prognosis of patients by exploiting the relationships between the variables.

Acknowledgements The authors would like to thank all medical and non-medical staffs who participated in the Malaysian National Cardiovascular Disease Database-Acute Coronary Syndrome registry. The authors would also like to acknowledge the financial assistance provided by Universiti Malaysia Sabah (Grant No. SLB2009).

Contributors Conception and design: YZZ, NJ and ASMZ. Data collection: ASMZ. Statistical analysis: NJ, YZZ, ZMK. Interpretation of the results: All authors. Drafting the manuscript: NJ. Critical revision of the manuscript and final approval of the version to be published: YZZ, ASMZ and NJ. Guarantor: NJ.

Funding This research was funded by Universiti Malaysia Sabah (Grant No. SLB2009).

Competing interests NJ has received funding from Universiti Malaysia Sabah (Grant No. SLB2009). None of the other authors has any conflicts of interest to disclose.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The NCVD registry study was approved by the Medical Review & Ethics Committee (MREC), Ministry of Health (MOH) Malaysia in 2007 (Approval Code: NMRR-07-20-250). MREC waived informed consent for NCVD.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Nurliyana Juhan <http://orcid.org/0000-0002-1557-9145>

Yong Zulina Zubairi <http://orcid.org/0000-0002-6174-7285>

Ahmad Syadi Mahmood Zuhdi <http://orcid.org/0000-0002-5349-0301>

Zarina Mohd Khalid <http://orcid.org/0000-0003-2473-3412>

REFERENCES

- 1 Roth GA, Mensah GA, Johnson CO, *et al*. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. *J Am Coll Cardiol* 2020;76:2982–3021.
- 2 Bergman BP, Mackay DF, Pell JP. Trends in cardiovascular disease in Scottish military veterans: a retrospective cohort study. *BMJ Open* 2021;11:e044465.
- 3 Cajita MI, Zheng Y, Kariuki JK, *et al*. mHealth technology and CVD risk reduction. *Curr Atheroscler Rep* 2021;23:36.
- 4 Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. *J Am Coll Cardiol* 2019;74:2529–32.
- 5 Tian S, Bi M, Bi Y, *et al*. A Bayesian network analysis of the probabilistic relationships between various obesity phenotypes and cardiovascular disease risk in Chinese adults: Chinese population-based observational study. *JMIR Med Inform* 2022;10:e33026.
- 6 Li Y, Xu J, Wang Y, *et al*. A novel machine learning algorithm, Bayesian networks model, to predict the high-risk patients with

- cardiac surgery-associated acute kidney injury. *Clin Cardiol* 2020;43:752–61.
- 7 Sayed A, Abdelfattah OM, Addison D. "Reply to comment on "long-term effectiveness of Empiric Cardio-protection in patients receiving cardiotoxic Chemotherapies: A systematic review and Bayesian network meta-analysis"" *Eur J Cancer* 2022;174:323–4.
 - 8 Badawi A, Di Giuseppe G, Gupta A, et al. Bayesian network Modelling study to identify factors influencing the risk of cardiovascular disease in Canadian adults with hepatitis C virus infection. *BMJ Open* 2020;10:e035867.
 - 9 Van den Eynde J, Bomhals K, Noé D, et al. Revascularization strategies in patients with Multivessel coronary artery disease: a Bayesian network meta-analysis. *Interact Cardiovasc Thorac Surg* 2022;34:947–57.
 - 10 Pearl J. *Probabilistic reasoning in intelligent systems: networks of plausible inference. Revised Se.* San Francisco, California: Morgan Kaufmann Publishers, 2014: 552.
 - 11 Meng QQ, Zhu QX, Gao HH, et al. A novel scoring function based on family transfer entropy for Bayesian networks learning and its application to industrial alarm systems. *Journal of Process Control* 2019;76:122–32.
 - 12 Kyrimi E, Neves MR, McLachlan S, et al. Medical idioms for clinical Bayesian network development. *J Biomed Inform* 2020;108:103495.
 - 13 Butcher B, Huang VS, Robinson C, et al. Causal Datasheet for Datasets: an evaluation guide for real-world data analysis and data collection design using Bayesian networks. *Front Artif Intell* 2021;4:612551:612551..
 - 14 Scutari M. Learning Bayesian networks with the Bnlearn R package. 2009. Available: <http://arxiv.org/abs/0908.3817>
 - 15 Scutari M, Graafland CE, Gutiérrez JM. Who learns better Bayesian network structures: constraint-based, score-based or hybrid Algorithms. Proceedings of Machine Learning Research 72, 2018 International Conference on Probabilistic Graphical Models; Prague, Czech Republic, 2018:1–12
 - 16 Bays HE, Taub PR, Epstein E, et al. Ten things to know about ten cardiovascular disease risk factors. *Am J Prev Cardiol* 2021;5:100149.
 - 17 Collard D, Nurmohamed NS, Kaiser Y, et al. Cardiovascular risk factors and COVID-19 outcomes in hospitalised patients: a prospective cohort study. *BMJ Open* 2021;11:e045482.
 - 18 Al Hayek AA, Robert AA, Matar AB, et al. Risk factors for hospital admission among COVID-19 patients with diabetes: A study from Saudi Arabia. *Saudi Med J* 2020;41:1090–7.
 - 19 Feigin VL, Brainin M, Norrving B, et al. What is the best mix of population-wide and high-risk targeted strategies of primary stroke and cardiovascular disease prevention *J Am Heart Assoc* 2020;9:e014494.
 - 20 Killip T, Kimball JT. Treatment of myocardial infarction in a coronary care unit. A two-year experience with 250 patients. *Am J Cardiol* 1967;20:457–64.
 - 21 Juhan N, Zubairi YZ, Zuhdi ASM, et al. Priors comparison in Bayesian Models of risk factor of Malaysian coronary artery disease male patients. *J Phys Conf Ser* 2021;1988:1.
 - 22 Liao Y, Xu B, Wang J, et al. n.d. A new method for assessing the risk of infectious disease outbreak. *Sci Rep*;7:1–12.
 - 23 Fuster-Parra P, Tauler P, Bannasar-Veny M, et al. Bayesian network modeling: A case study of an epidemiologic system analysis of cardiovascular risk. *Comput Methods Programs Biomed* 2016;126:128–42.
 - 24 Russel S, Norvig P. *Artificial Intelligence a Modern Approach* Third Edition. New Jersey: Pearson Education, Inc, 2016: 1151.
 - 25 Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data. *Mach Learn* 1995;20:197–243.
 - 26 Scutari M, Denis J-B. Bayesian networks. In: *Bayesian Networks: With Examples in R.* Second edition. Boca Raton: Chapman and Hall/CRC, 2022.
 - 27 Singh K KXieM2010;7.
 - 28 R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 2022.
 - 29 Netica. Norsys software corporation. copyright 1995–2012. 1995.
 - 30 Hatori M, Sakakura K, Taniguchi Y, et al. Factors associated with in-hospital death in patients with Killip class 3 acute myocardial infarction. *Int Heart J* 2021;62:756–63.
 - 31 Suzuki M, Nishihira K, Takegami M, et al. Clinical profiles and outcomes in the treatment of acute myocardial infarction in Japan of aging society. *Heart Vessels* 2020;35:1681–8.
 - 32 Fu Y, Sun H, Guo Z, et al. A risk score model to predict in-hospital mortality of patients with end-stage renal disease and acute myocardial infarction. *Intern Emerg Med* 2021;16:905–12.
 - 33 Kuji S, Kosuge M, Kimura K, et al. The Identifi-impact of acute kidney injury on in-hospital outcomes of patients with acute myocardial infarction-results from the Japanese Registry of acute myocardial infarction diagnosed by universal definition (J-MINUET) Substudy. *Circ J* 2017;81:733–9.
 - 34 Zhang Q, Zhou L, Cai H-L, et al. Relationship of the ORBIT and HAS-BLED scores with Killip class 3-4 in patients with ST-segment elevation myocardial infarction. *Medicine (Baltimore)* 2019;98:e14578.
 - 35 Sabroe JE, Thayssen P, Antonsen L, et al. Impact of renal insufficiency on mortality in patients with ST-segment elevation myocardial infarction treated with primary percutaneous coronary intervention. *BMC Cardiovasc Disord* 2014;14:15:1–11..
 - 36 Chaudhary AK, Pathak V, Kunal S, et al. Cha2Ds2-Vasc score as a novel Predictor for contrast-induced nephropathy after percutaneous coronary intervention in acute coronary syndrome. *Indian Heart J* 2019;71:303–8.
 - 37 Karabağ Y, Çağdaş M, Rencuzogullari I, et al. The C-reactive protein to albumin ratio predicts acute kidney injury in patients with ST-segment elevation myocardial infarction undergoing primary percutaneous coronary intervention. *Heart, Lung and Circulation* 2019;28:1638–45.
 - 38 Sakr H, Azazy AS, Hillani A, et al. Clinical profiles and outcomes of acute ST-segment elevation myocardial infarction in young adults in a tertiary care center in Saudi Arabia Autoimmunity view project. *SMJ* 2021;42:1201–8.
 - 39 Ljuca F, Hadžiefendić B, Jahić E, et al. Pentraxin 3 might be better Prognostic serum marker than IL-6, IL-10, and high-sensitivity C-reactive protein for major adverse cardiovascular events in patients with ST-elevation myocardial infarction after bare-metal Stent implantation. *Saudi Med J* 2019;40:1202–8.
 - 40 Yang HY, Ahn MJ, Jeong MH, et al. Predictors of in-hospital mortality in Korean patients with acute myocardial infarction. *Chonnam Med J* 2019;55:40–6.
 - 41 Sanchis J, Ruiz V, Bonanad C, et al. Growth differentiation factor 15 And geriatric conditions in acute coronary syndrome. *Int J Cardiol* 2019;290:15–20.
 - 42 Aziz F, Malek S, Ibrahim KS, et al. Short- and long-term mortality prediction after an acute ST-elevation myocardial infarction (STEMI) in Asians: A machine learning approach. *PLoS One* 2021;16:e0254894.
 - 43 Juhan N, Zubairi YZ, Zuhdi ASM, et al. Risk factors of mortality among male patients with cardiovascular disease in malaysia using bayesian analysis. *Iran J Public Health* 2020;49:1642–9.
 - 44 D'Urso P, Vitale V. Bayesian networks model averaging for BES indicators. *Soc Indic Res* 2020;151:897–919.
 - 45 Luo Y, McShan D, Ray D, et al. Development of a fully cross-validated Bayesian network approach for local control prediction in lung cancer. *IEEE Trans Radiat Plasma Med Sci* 2019;3:232–41.
 - 46 Cypko MA, Stoehr M. Digital patient models based on Bayesian networks for clinical treatment decision support. *Minimally Invasive Therapy & Allied Technologies* 2019;28:105–19.