Original Article

OPEN ACCESS

MODESTUM

Prediction of mortality in young adults with cardiovascular disease using artificial intelligence

Muayyad Ahmad [1]* (iD), Salam Bani Hani [2] (iD), Mahmoud Abu-Abeeleh [3] (iD), Ibrahim Aljarah [4] (iD)

[1] School of Nursing, University of Jordan, Amman, JORDAN
[2] Irbid National University, Irbid, JORDAN
[3] School of Medicine, University of Jordan, Amman, JORDAN
[4] Department of Information Technology, King Abdullah II School of Information Technology, Amman, JORDAN
*Corresponding Author: mma4@ju.edu.jo

ARTICLE INFO	ABSTRACT
	Background: Young mortality is prevalent among patients with cardiovascular disease (CVD). To develop prediction models for CVD mortality in young adults, it is crucial to assess CVD risks. Early detection of cardiac disorders using machine learning algorithms, a branch of artificial intelligence (AI) is crucial for preventing more damage to coronary arteries and saving lives. **Aims:** To predict mortality versus a life outcome among young adults (18-45 years) with CVD using AI technique known as Chi-squared automatic interaction detector (CHAID) algorithms. **Methods:** A large-scale dataset was extracted from the electronic health records of 809 young adult patients diagnosed with CVD using a retrospective design. Information was retrieved regarding young adults from Jordan who were admitted to public health institutions between 2015 and the end of 2021. **Results:** CHAID algorithms were chosen among seven prediction models based on accuracy and area under curve to predict mortality vs life in young individuals (18-45 years old) with CVD. The mortality prediction algorithms started with pulse pressure, then diastolic blood pressure, then ischemic heart disease, and last geographical location. **Conclusions:** CHAID model used in our study indicated how the death rate was classified and distributed among a variety of parameters. As a result, we may argue that AI model could provide additional information on how many aspects are articulated in connection to CVD patient fatality situations. **Keywords:** heart diseases, death, mortality, artificial intelligence, algorithms, young

INTRODUCTION

Cardiovascular disease (CVD) is one of the world's major causes of death [1, 2]. Increased CVD risks were estimated, identified during childhood, and noted in young adults [3]. According to the most recent WHO report from 2018, heart disease death in Jordan reached 7,615 cases, representing 22.97% of all deaths. With a death rate of 172.28 per 100,000 people, Jordan is ranked 52 globally in terms of the prevalence of heart disease [4].

Many patients with CVD continue to have a high risk of mortality at a young age [5]. Pulse pressure (PP) is regarded as an indicator of arterial stiffness and/or atherosclerosis. PP predicts the onset of CVD [6]. Furthermore, the assessment of the primary risk factors for the mortality from CVD was the subject of a sizable number of studies [7]. The main risk factors for the mortality associated with CVD were tested using a variety of tools with various sample sizes. The framingham tool revealed that age, sex, systolic blood pressure (SBP), diabetes, smoking status, total cholesterol, and high-density lipoprotein were the main antecedents of CVD [8]. Therefore, assessing the

risks of CVD is essential in building prediction models to estimate the magnitude of the mortality of CVD at the population level and in specific subgroups such as young adults [9, 10].

Early detection of heart diseases through utilizing machine learning algorithms (MLAs) as sub-fields of artificial intelligence (AI) is an essential step to prevent cardiac patients from developing further coronary artery damage and saving their lives [11]. MLAs and other large-scale data analytical techniques must take place as an effective and factual method to predict heart diseases such as CVD, ischemic heart disease (IHD), and CHF [12]. MLAs that can create a variety of prediction models to accurately assess the presence or absence of risk factors that contribute to the development of CVD can be used to manage these enormous data sets, also known as big data or large-scale data [13, 14]. Based on various populations and subgroups, several prediction models were created using the well-established risk factors and antecedents of CVD [9].

The literature claims that prediction models are lacking in Jordan and the Arab world as a whole, particularly among young people with CVDs. Additionally, this study focused on patients who were younger adults, whereas the majority of

heart disease research focuses on elderly patients. The creation of a trustworthy model for classifying each population's risk of developing CVD has consequently elevated to the top of the priority list for researchers and organizations involved in this field. To develop a prediction model for young adults (18-45) with IHD versus CHD, with an emphasis on the outcome of death versus live status, is the goal of the current study.

METHODS

Study Design

The study employed a retrospective design to collect data from electronic health records (EHRs) system called "Hakeem" for young adult Jordanian patients with CVD. Data were gathered from Jordanian young people who were hospitalized in public health facilities since 2015.

Sample & Variables of Study

The dataset was acquired through the retrospective analysis of EHRs pertaining to 809 young adult patients who had received a diagnosis of CVD. The sample of the study was for Jordanians who were hospitalized in public health facilities between 2015 and 2021.

EHRs contain various levels of patient variables, such as demographic information, a diagnosis, laboratory results, medication, and a history of issues, all of which allow for the meaningful use of patient information [15]. Patient's identification (ID) numbers, sex, death status, governorate, medical diagnosis based on international classification of disease (ICD-10), laboratory results including high-density lipoprotein (HDL), lactate dehydrogenase (LDH), cholesterol level, fasting blood sugar (FBS), SBP, and diastolic blood pressure (DBP) were obtained from the HDA department for the admitted patients from the period of 2015 to the end of 2021. Data were downloaded as EXCEL sheets in many files from the HDA. Age, gender, and geographic location have no missing data. Data were merged into a single file using SPSS program [16].

Data Analysis

Using frequency analysis and outlier detection, the data were examined for noise, inconsistency, and missing values. A huge number of redundant data was removed after sorting, cleaning, and organizing the extracted data. For example, there was numerous missing data (more than 80.00%) for variables like HDL, LDH, and FBS. Thus, these variables were excluded from the analysis.

Chi-squared automatic interaction detector (CHAID) is a subtype of AI that is created for decision trees. It is a very effective statistical method for segmentation, also known as tree growing [17]. CHAID assesses each value in a potential predictor field using the significance of a statistical test as a criterion [18].

Transformation of data

The researchers chose the most relevant attributes of CVDs using data visualization. Besides, international business machine statistical package for social science (SPSS) modeler (version 18.0) was used for manipulating, analyzing, and visualizing the data, which provides the features of presenting the data powerfully for statistical analysis and data

Table 1. Characteristics of sample (n=809)

Item	n (%)
Gender	
Male	385 (48.8)
Female	414 (51.2)
Age group	
18-24	90 (11.1)
25-30	127 (15.7)
31-35	125 (15.5)
36-40	160 (19.8)
41-45	307 (37.9)
Governorate	
Irbid	157 (19.4)
Ajloun, Jerash, Mafraq, & Balqa'	78 (9.6)
Amman	424 (52.4)
Zarqa	85 (10.5)
Ma'daba, Karak, Tafelah, Ma'an, & Aqaba	65 (8.0)
Medical diagnosis	
IHD	588 (72.7)
CHF	221 (27.3)
Death status	
Dead	14 (1.70)
Alive	795 (98.3)

Note. n: Number & %: Frequency

management for descriptive and predictive modelling [19]. Descriptive modelling was used to identify the main risk factors of CVD, visualization of data, and rank the attributes of CVD from the most occurrences attributes to the least one. Besides, CHAID as a decision tree technique keeps all other values that are heterogeneous and merges values that are deemed to be statistically homogeneous concerning the target variable. The best predictor is then chosen to create the decision tree's first branch, with each child node consisting of a collection of uniform values from the chosen field. Recursively, this process goes on until the tree is fully developed. The statistical test that is applied depends on the target field's level of measurement. A chi-squared test is applied if the target field is categorical.

Ethical Consideration

Patient records were handled anonymously by employing an ID as the distinguishing characteristic of each record. The data that were extracted were stored in a separate file that was locked up and stored in a secure location within the researcher's office. For the purposes of the study, the researchers were the only ones who could access the data.

RESULTS

Sample Characteristics

The health records of 809 young adults (18-45 years old) were extracted. Females were slightly more (51.20%) in the sample, with the highest age group of 41-45 years (n=307, 37.90%). About three-quarters of CVD had been diagnosed with IHD (n=588, 72.70%). The sample represented all the governorates in Jordan, with the majority was from Amman (n=424, 52.40%) (**Table 1**).

The blood pressure readings, which include SBP, DBP, PP, and mean arterial pressure (MAP) are presented in **Table 2**. Unfortunately, not all patient health records had complete workup results. Each patient had many readings for the BP, thus, the first reading on admission of each patient was taken.

Table 2. Outcomes of clinical work-up for patients (n=809)

Work-up results	Minimum	Maximum	M (SD)
Pulse pressure	19	137	54.4 (15.8)
Mean arterial pressure	66	186	107.1 (16.9)
Systolic blood pressure	80	240	135.6 (22.4)
Diastolic blood pressure	47	165	81.2 (14.4)

Note. M: Mean & SD: Standard deviation

Table 3. Seven research data models

Model	AUC	Overall accuracy (%)
CHAID	0.894	98.269
Discriminant	0.844	79.604
Logistic regression	0.893	98.146
Neural net	0.866	98.269
C5	0.500	98.269
Quest	0.500	98.269
C&R Tree	0.500	98.269

Predictive Models

Seven models were built, but one of them was adopted in this study. Evaluation was based on the overall accuracy and the area under the curve (AUC). Therefore, CHAID model was utilized, which achieved the highest level of accuracy 98.27 and AUC of 0.89 among the built models (**Table 3**).

CHAID algorithms, an AI technology, were found to be the best effective predictive model for forecasting death versus a life among young adults (18-45 years old) who had CVD. The SPSS modeler was used to generate the seven Nodes model. During the model training process, about 30.00% of the data is used to check whether the model produces valid results.

Exhaustive CHAID is a CHAID modification created to address some of the method's shortcomings. Since CHAID stops merging categories as soon as it determines that all the remaining categories are statistically different, it is possible that it occasionally does not find the best split for a variable. This is fixed by exhaustive CHAID, which keeps merging the predictor variable's categories until there are only two super categories left. Once the series of merges for the predictor has been examined, the set of categories with the strongest correlation to target variable is identified.

An adjusted p-value is then calculated for that correlation. By comparing the adjusted p-values, exhaustive CHAID can determine the best split for each predictor and then decide which predictor to split on.

The decision tree in **Figure 1** begins with the root node, which displays the death status as an outcome field (node 0). Based on the statistical significance of the predictor of the strongest relationship with the target field, it was PP, the data are split into three nodes. Patients with PP<59 showed two deaths (rate of 0.368%), while those with PP between 59 and 73 showed five deaths (rate of 2.717%), and those with PP>.73 showed death of seven patients with the highest rate (8.642%), (Chi-square=29.748, p<.001).

Diastolic blood press was the next significant predictor in the model that stemmed from the cases with PP≤59 (node 4 and node 5). Patients with DBP≤90 mmhg had zero death, while those with DBP>90 mmhg had two cases of death (2.35%). Node 6 and node 7 stemmed from node 3 (PP>73). IHD in node 6 had two cases of death (3.64%), while node 7, which represents CHF cases had five cases of death (19.23%) (Chi-square=5.438, p=.020).

Two nodes related to governorates emerged from node 6. Node 8 includes 11 governorates with zero death from IHD, while node 9 refers to Zarqa governorate alone with two cases of death (22.22%) (Chi-square=10.610, p=.017).

Furthermore, to confirm the validity of the selected model (CHAID) in the analysis and results, the second model with the highest overall accuracy (98.269) and AUC (0.866) was neural Net, thus the predictors importance is plotted in **Figure 2**. The result was consistent and supported CHAID model findings.

DISCUSSION

To create a model for predicting death versus a life status in this large group of young adults who have heart disease, retrospective data from the Jordanian health information system were used. In contrast to most of the previous research on heart disease, this particular study focused on younger adults rather than the elderly, specifically those between the ages of 18 and 45. In the sample of this study, all of the patients were diagnosed with either IHD or CHF.

Based on the regression model that incorporates demographics and blood pressure readings as predictors for death cases [14]. In our research, data-driven modeling that makes use of MLAs can improve risk prediction. One example of this is CHAID, which assumes splitting the data and visualizing the most important risk factors in the model while taking into consideration the interaction between predictors. This can be accomplished even with a limited number of attributes and a limited number of algorithms [20]. Neural net model has supporting outcomes of the adopted CHAID model, which was based on the highest AUC and accuracy prediction percentage out of seven models.

PP was the first emerging predictor for death status in our study. patients with PP higher than 73 had the highest chance of death (8.64%). This finding is consistent with a Korean study, which found that high PP in patients aged 40 to 69 was a substantial predictor of cardiovascular mortality [21]. It is simple to compute PP, which enables the prognosis of cardiovascular death in individuals with advanced heart failure.

The third predictor, which emerged from node 2 (PP<.59) was DBP with a cut point of 90 mmhg. In this study we found that patients with DBP>90 mmgh had a death rate of 2.35%. In another study, patients with CVDs who have DBP>80 had death rate of 18.00% [22]. The reason for the difference in death rate between our study and the others could be due the AI method of analysis.

The fourth attributable risk factor was medical diagnosis, with CVD classified as IHD or non-IHD, which emerged only from patients with PP higher than 73. According to the findings of this study, the death rate of IHD was only 3.64%, while the death rate of CHF is 19.23%. These findings are consistent with the findings of a study in [23] that examined CVD trends among young adults (25-44 years) between 1999 and 2018, finding that the death rate was 13.40%. These trends in CVD death rates among this age group were unclear, which could be attributed to low socioeconomic status, a lack of information on comorbidity conditions, and the presence of risk factors such as smoking and substance abuse [24, 25].
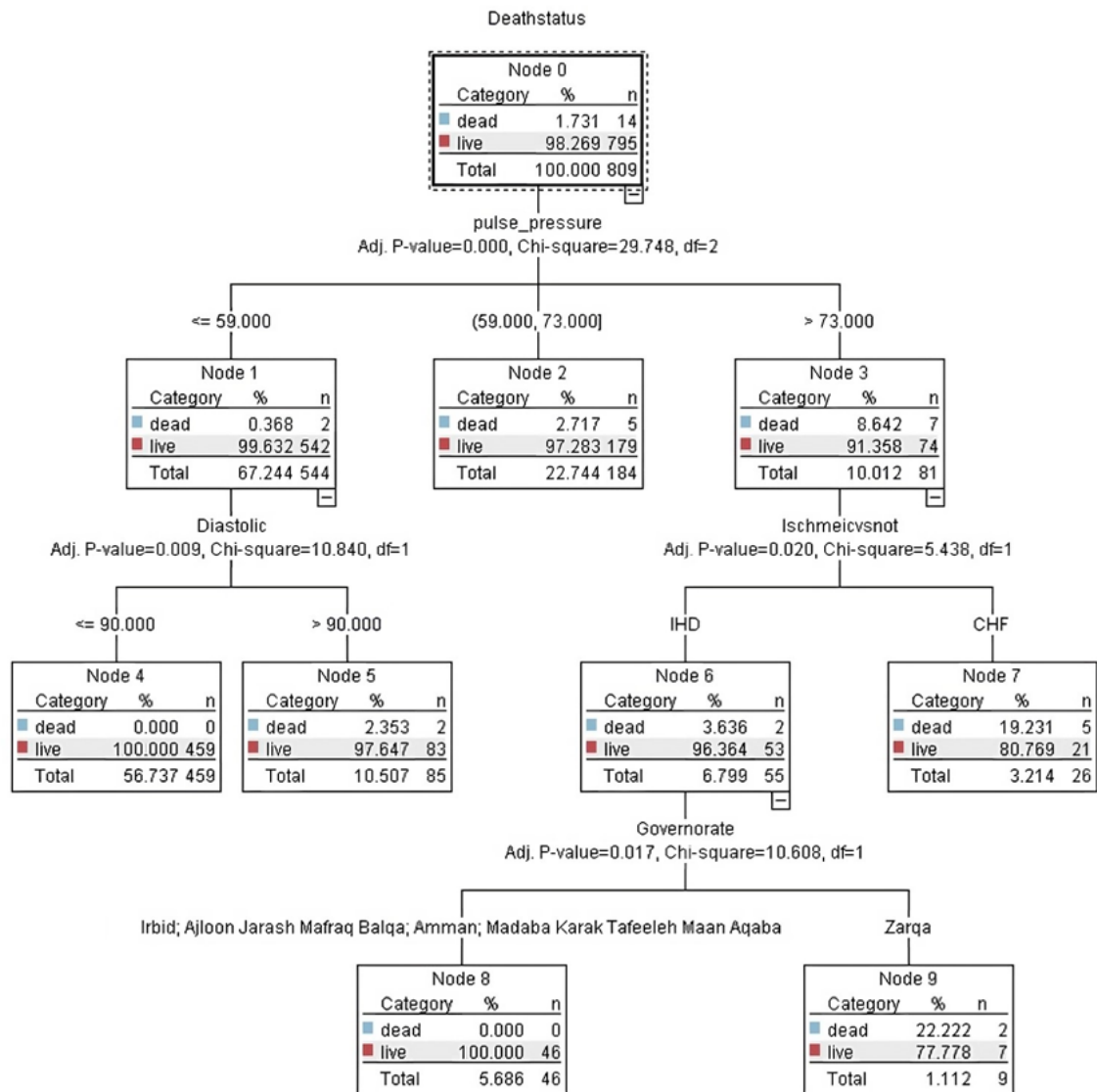
Figure 1. Study predictive model of death versus a life outcome among young adults (Source: Authors' own elaboration)
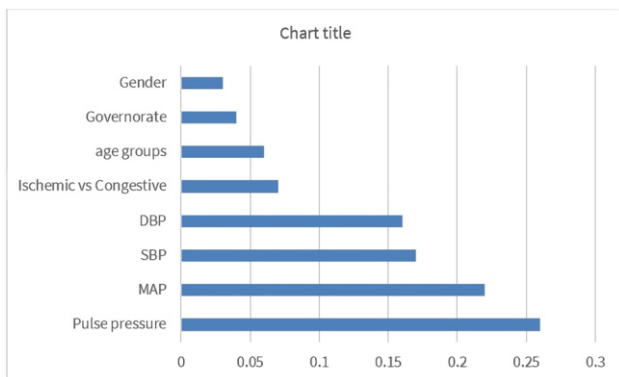

Figure 2. Predictors importance in neural net model (target: dead=0 & alive=1) (Source: Authors' own elaboration)

The residency location was the last identified attribute for death versus life status among young Jordanian adults in our study.

Out of 55 participants who had IHD, nine of them were from Zarqa governorate with a mortality rate of 22.22%.

All other governorates did not show mortality cases for patients with IHD. Research from many nations has documented relationship between area-level socioeconomic disadvantage and spatial disparities in the distribution of CVDs at various sizes [25-28].

According to Jordanian MOH statistics from 2017, circulatory system diseases are the leading cause of death among all ages in Jordan (42.36%). However, no studies were found to compare our results in relation to geographical area and IHD together. Therefore, taking into consideration the findings of our study, which is one of the few to use an AI model among young adults, there is a need for additional research to be conducted in a variety of countries to validate our findings.

CONCLUSIONS

In summary, we can claim that CHAID model of our study correctly classified and distributed the mortality rate among numerous factors. Therefore, we could assert that the AI model added more detail regarding how various factors are articulated in relation to CVD patient death cases. Although the sample size was relatively large, it was a limitation in this study that prevented us from having a large sample of the quality of the entered data in EHR. The patient's medications, laboratory results, radiological examination findings and body mass index

had several potential flaws, including incomplete documentation and missing data.

REFERENCES

1. Rehman S, Rehman E, Ikram M, Jianglin Z. Cardiovascular disease (CVD): Assessment, prediction and policy implications. BMC Public Health. 2021;21:1299. https://doi.org/10.1186/s12889-021-11334-2

2. Mensah GA, Fuster V, Murray CJL, Roth GA, Global Burden of Cardiovascular Diseases and Risks Collaborators. Global burden of cardiovascular diseases and risks, 1990-2022. J Am Coll Cardiol. 2023;82(25):2350-473. https://doi.org/10.1016/j.jacc.2023.11.007 PMid:38092509

3. Königstein K, Wagner J, Frei M, et al. Endothelial function of healthy adults from 20 to 91 years of age: Prediction of cardiovascular risk by vasoactive range. J Hypertens. 2021;39(7):1361-9. https://doi.org/10.1097/HJH.0000000000002798 PMid:33470736

4. WHO. Jordan. WHO; 2023. Available at: https://www.who.int/nmh/countries/jor_en.pdf (Accessed: 31 July 2023).

5. Benjamin EJ, Munther P, Alonso A, et al. Heart disease and stroke statistics–2019 update: A report from the American Heart Association. Circulation. 2019;139(10):e56-528.

6. Vasan RS, Song RJ, Xanthakis V, et al. Hypertension-mediated organ damage: Prevalence, correlates, and prognosis in the community. Hypertension. 2022;79(3):505-15. https://doi.org/10.1161/HYPERTENSIONAHA.121.18502 PMid:35138872 PMCid:PMC8849561

7. McGavock JM, Anderson TJ, Lewanczuk RZ. Sedentary lifestyle and antecedents of cardiovascular disease in young adults. Am J Hypertens. 2006;19(7):701-7. https://doi.org/10.1016/j.amjhyper.2006.01.013 PMid:16814124

8. Shetty DK, Rodrigues LLR, Shetty AK, Nair G. Machine learning based predictors of cardiovascular disease among young adults. Eng Sci. 2022;17:292-302. https://doi.org/10.30919/es8d627

9. Farzadfar F. Cardiovascular disease risk prediction models: Challenges and perspectives. Lancet Global Health. 2019;7(10):e1288-9. https://doi.org/10.1016/S2214-109X(19)30365-1 PMid:31488388

10. Bani Hani S, Ahmad M. Effective prediction of mortality by heart disease among women in Jordan using the Chi-squared automatic interaction detection model: Retrospective validation study. JMIR Cardio. 2023;7:e48795. https://doi.org/10.2196/48795 PMid:37471126 PMCid:PMC10401188

11. Gamboa-Cruzado J, Crisostomo-Castro R, Vilabuleje J, López-Goycochea J, Valenzuela JN. Heart attack prediction using machine learning: A comprehensive systematic review and bibliometric analysis. J Theor Appl Inf Technol. 2024;102(5).

12. Gavhane A, Kokkula G, Pandya I, Devadkar K. Prediction of heart disease using machine learning. In: Proceedings of the 2018 2nd International Conference on Electronics, Communication and Aerospace Technology. IEEE; 2018. p. 1275-8. https://doi.org/10.1109/ICECA.2018.8474922

13. Bergamini M, Iora PH, Rocha TAH, et al. Mapping risk of ischemic heart disease using machine learning in a Brazilian state. PLoS ONE. 2020;15(12):e0243558. https://doi.org/10.1371/journal.pone.0243558 PMid:33301451 PMCid:PMC7728276

14. Bani Hani S, Ahmad M. Machine-learning algorithms for ischemic heart disease prediction: A systematic review. Curr Cardiol Rev. 2023;19(1):e090622205797. https://doi.org/10.2174/1573403X18666220609123053 PMid:35692135 PMCid:PMC10201879

15. Gabriel PE. "Meaningful use" means process redesign. Chest. 2010;138(3):472-4. https://doi.org/10.1378/chest.10-1202 PMid:20822984

16. IBM. IBM SPSS statistics for windows, version 25.0. Armonk (NY): IBM Corp, 2017.

17. Kass GV. An exploratory technique for investigating large quantities of category data. Appl Stat. 1980;29:119-27. https://doi.org/10.2307/2986296

18. Kass IS, Wang CC, Walrond JP, Stretton AO. Avermectin B1a, a paralyzing anthelmintic that affects interneurons and inhibitory motoneurons in Ascaris. Proc Natl Acad Sci USA. 1980;77(10):6211-5. https://doi.org/10.1073/pnas.77.10.6211 PMid:6255481 PMCid:PMC350245

19. Wendler T, Gröttrup S. Data mining with SPSS modeler: Theory, exercises and solutions. New York City (NY): Springer; 2016. https://doi.org/10.1007/978-3-319-28709-6

20. Tsamardinos I, Charonyktakis P, Papoutsoglou G, et al. Just add data: Automated predictive modeling for knowledge discovery and feature selection. NPJ Precis Oncol. 2022;6(1):38. https://doi.org/10.1038/s41698-022-00274-8 PMid:35710826 PMCid:PMC9203777

21. Lee J-W, Shin S-J, Kim J, Kang H-T. Higher pulse pressure is associated with increased risk of cardio-cerebrovascular disease and all-cause mortality: A Korean national cohort study. Am J Hypertens. 2022;35(7):647-55. https://doi.org/10.1093/ajh/hpac043 PMid:35363861

22. Tang Y, Liu S, Shi Y, et al. Association of blood pressure in the first-week of hospitalization and long-term mortality in patients with acute left ventricular myocardial infarction. Int J Cardiol. 2022;349:18-26. https://doi.org/10.1016/j.ijcard.2021.11.045 PMid:34838680

23. Khan SU, Bashir ZS, Khan MZ, et al. Trends in cardiovascular deaths among young adults in the United States, 1999 to 2018. Am J Cardiol. 2020;128:216-7. https://doi.org/10.1016/j.amjcard.2020.05.014 PMid:32534735 PMCid:PMC7457428

24. German CA, Baum SJ, Ferdinand KC, et al. Defining preventive cardiology: A clinical practice statement from the American Society for Preventive Cardiology. Am J Prev Cardiol. 2022;12:100432. https://doi.org/10.1016/j.ajpc.2022.100432 PMid:36425534 PMCid:PMC9679464

25. Naghavi M, Ong KL, Aali A, et al. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: A systematic analysis for the global burden of disease study 2021. Lancet. 2024 Apr 19:S0140-6736(24)00824-9. https://doi.org/10.1016/S0140-6736(24)00367-2 PMID: 38582094

26. Jain V, Minhas AMK, Morris AA, et al. Demographic and regional trends of heart failure-related mortality in young adults in the US, 1999-2019. JAMA Cardiol. 2022;7(9):900-4. https://doi.org/10.1001/jamacardio.2022.2213 PMid: 35895048 PMCid:PMC9330269

27. Montinari MR, Minelli P, Russo A, Gianicolo E. Patterns of coronary heart disease mortality in Italy from 1931 to 2015 and a focus on a region with highly industrialized areas. Int J Cardiol. 2022;354:56-62. https://doi.org/10.1016/j.ijcard.2022.03.006 PMid:35278577

28. Schumacher, Kyu HH, Aali A, et al. Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. Lancet. 2024 Mar 8:S0140-6736(24)00476-8. https://doi.org/10.1016/s0140-6736(24)00476-8 PMid: 38484753