

Research on the Influence Factors of Heart Disease

Feng Guo^{1,*}

¹School of Mathematics, China University of Mining and Technology, Xuzhou, 221116, China

*Corresponding author: 10193768@cumt.edu.cn

Abstract:

Heart disease accounts for almost 30% of all deaths worldwide and about 18 million deaths each year, making it the primary cause of death and disability worldwide. Recent research highlights variations in incidence rates and causative factors among different populations and regions, underscoring the need for detailed studies to develop effective prevention and treatment strategies. This article explores the various factors that influence heart disease. Although many studies have been done to show that the incidence of heart disease is related to dietary habits, genes, and other factors, there are still many details that should be investigated. This paper focuses on determining which factors have effects on the incidence of heart disease, such as age, gender, and some diseases. Binary logistic regression is used to statistically analyse the data and found that certain factors, such as gender, chest pain, thallium stress, exercise-induced ST depression and major fluoroscopy vessels were associated with the development of heart disease. The use of binary logistic regression modelling enabled the study to quantify the impact of these factors and provided a useful tool for predicting heart disease risk. The study's findings serve as a valuable resource for the formulation of heart disease prevention and treatment plans.

Keywords: Heart Disease; Pathogenic Factors; Binary Logistic Regression.

1. Introduction

Heart disease is considered one of the main causes of death and disability globally. The World Health Organization (WHO) claims that cardiovascular disease causes about 18 million deaths per year, which account for around 30% of global deaths. Wei (2022) pointed out that the mortality rate of heart disease in urban and rural populations in China fluctuated over time from 2004 to 2019 [1]. In addition to having a serious negative effect on a patient's personal health, heart disease places a substantial financial strain on families and the community.

In recent times, medical and public health research has extensively explored the influencing factors of heart disease. These factors include genetic background, age, gender, environmental factors, psychological stress, and lifestyle related risk factors. However, there are differences in the incidence rate and pathogenic factors of heart disease among different populations and regions, which requires more detailed and specific research in order to develop effective prevention and treatment strategies. Identifying the influencing factors of heart disease and analysing the correlation with certain factors can provide empirical evidence and reference basis for preventing and treating heart disease.

In the field of modern medicine, research on the psycho-

logical state, and other influencing factors is increasingly receiving attention. Some articles have studied similar factor analyses. Wang (2024) found that being angry within 8 minutes may significantly increase the risk of heart attack [2]. Anger has the potential to set off a stress response that raises blood pressure and heart rate, raising the extent of getting heart attack. This indicated that emotional management is vital in preventing heart disease and the significance of mental health on heart disease had received widespread attention. Qi et al. (2023) studied the association between depression and anxiety symptoms and heart disease in elderly people in China, and found that depression and anxiety symptoms significantly increased the risk of heart disease in elderly people [3]. Similarly, some studies found that sorrow and insomnia increased the risk of heart disease respectively [4, 5]. These studies indicated that emotional management was crucial in preventing heart attack.

This disease can be influenced via social and environmental elements. Xu (2024) found that heart disease and stroke risk may rise as a result of microplastic exposure [6]. Specifically, in addition to changes in individual health behaviors, improvements in environmental protection and social support systems were also crucial. Lifestyle played an important role in preventing heart disease. Cheng et al. (2024) interpreted a study, and emphasized the key role of

regular physical activity and high-level cardiopulmonary fitness in preventing atherosclerotic cardiovascular disease [7]. This indicated that developing scientific exercise prescriptions and improving lifestyle were important strategies for reducing the risk of this disease. The influencing factors of disease varied ranging from young to elderly individuals. The incidence rate of heart attacks increased considerably with a rise in blood pressure level, according to research by Li et al. (2023) on the relationship between young people under 40 years old’s baseline blood pressure level and premature heart disease [8]. These studies suggested that preventive measures should be designed specifically based on risk factors in different age groups. In the Jiaozuo City liberated area, Song et al. (2019) examined 260 middle-aged and elderly individuals suffering from chronic heart failure (CHF). The research conclusion pointed out that there was a correlation between the increase of age and the increase of the incidence rate of CHF, and more men than women [9]. Clinical preventive measures should be taken against these influencing factors.

In addition, Zhao et al. (2024) found that congenital heart disease (CHD) accounted for a large proportion of birth defects. Identifying tobacco exposure during pregnancy was an important risk factor for CHD, and CHD had a high incidence in newborns and children, and its influencing factors involved heredity, environment, social economy and other aspects [10]. The relationship between metabolic diseases such as diabetes and heart disease cannot be ignored. Huang et al. (2023) studied the influencing factors of angina pectoris with coronary atherosclerotic heart disease (CHDAP) in elderly patients with diabetes, and found that age, body mass index (BMI), diabetes, hypertension, as well as other factors were independent risk factors of CHDAP [11]. These results indicated that controlling metabolic diseases and their related indicators was of great significance for preventing heart disease. The existence and management of these chronic diseases may affect the overall health status and diagnostic effectiveness of patients.

The goal of this research is to identify each factor’s mode of action in the course of heart disease. The aims of this article to identify and analyse some relevant factors that may affect the prevalence of heart disease in light of the

notable variations in the causative factors of heart disease in different populations and other related diseases and gain insight into whether these factors will accurately affect the prevalence of heart disease, so as to fill the gap in existing research and provide new scientific evidence. Multivariate analysis is used to demonstrate the impact of these factors on the disease, which provides a basis for better prevention and treatment strategies.

This study has important public health and medical implications. By identifying and analyzing the key influencing factors of heart disease, research can provide an empirical basis for developing effective prevention and treatment strategies to reduce the incidence and mortality of heart disease. Furthermore, policymakers can use the findings as experience when developing focused health policies that lessen the financial and social costs associated with heart disease and enhance overall health.

The following sections of this article focusing on the research topic are as follows. In the methods, we used 303 sets of data from Kaggle and employed a binary logistic regression model based on the data type. Subsequently, descriptive analysis was conducted on the data in result and discussion, and correlation analysis was conducted using SPSS software to identify factors that own a certain correlation between the incidence of heart attack. Data was further optimized, and corresponding regression equations were derived based on the optimized data analysis.

2. Methods

2.1 Data Sources and the Selection of Variables

The information used was gathered by FARZAD NEK-OUEI from the Kaggle website and updated in 2023. The data used in this article includes 303 people whose age ranges from 29 to 77 years old. People are divided into two categories: those who suffer from heart disease and those who are healthy. About 96 patients are male and 207 patients are female. This dataset includes some variables including age, gender, chest pain, serum cholesterol, blood pressure in rest, blood sugar level, resting electrocardiographic, which are all listed in Table 1.

Table 1. Elements and the range

Symbol	Variable	Range
age	age of people	29-77
sex	gender	0 = male, 1 = female

cpt	type in chest pain	0 = typical angina; 1 = atypical 2 = non; 3 = asymptomatic
restbps	blood pressure in resting	94-200
cho	blood cholesterol levels	126-564
bs	blood sugar above 120 mg/dl	0 = false; 1 = true
recg	resting electrocardiographic	0 = normal; 1 = abnormal ST-T wave 2 = left ventricular hypertrophy
phrs	the peak of heart rate under stress	71-202
eang	exercise-induced angina	0 = no; 1 = yes
opeak	ST depression caused by exercise	0-6.2
slope	slope of ST segment's peak	0 = upsloping; 1 = flat; 2 = down sloping
nflu	number of major fluoroscopy vessels	0-4
thalt	thallium stress test	0 = normal; 1 = fixed defect 2 = reversible defect; 3 = not described
target	status of heart attack	0 = no; 1 = presence

Age, restbps, cho, phrs, opeak and nflu are numeric, and others are categorical

2.2 Research Protocol

The logistic linear regression function is the most widely utilized

$$\text{logit}p = \ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

From above equation, we can obtain

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)} \quad (2)$$

The right-hand side of equation (1) ($\frac{p}{1-p}$) is known as relative risk (or advantage ratio), it is the ratio of the likelihood that the noteworthy event would occur to the likelihood that it may not happen, β_i call Logistic regression coefficient. If $p > 0.5$, it can be predicted that the event will occur, otherwise it will not occur. In this way, it is possible to classify samples of an unknown category, solving similar problems in discriminant analysis.

The logistic regression model is actually an extension of the ordinary multiple linear regression model, but its error term follows a Bernoulli distribution rather than a normal distribution. Therefore, the maximum likelihood method is used to estimate the regression coefficients. This paper uses SPSS to analyze the relationship between influencing factors and heart disease.

3. Result and Discussion

3.1 Descriptive Statistics

The characteristic statistics results indicate that there are a total of 303 samples. In Table 2 and Fig.1, descriptive statistics of continuous variables and categorical variables can be shown clearly. The feature of these factors can be easily used to analyze the correlation. The mean values for the variables are as follows, age (54.37), resting blood pressure (131.62), cholesterol (246.26), maximum heart rate (149.65), opeak (1.04), and nflu (0.73). The standard deviations indicate the variability within the sample, with cholesterol and maximum heart rate showing relatively high standard deviations of 51.831 and 22.905, respectively, suggesting significant variation in these values among the subjects. The type of each categorical variable is easy to find out, and each of them has different proportion, such as the variable cpt. It has four types, and the number of each type is easily to check.

Table 2. Descriptive statistics of continuous variables

elements	N	Min	Max	Mean	Standard deviation	Variance
age	303	29	77.0	54.37	9.082	82.485
restbps	303	94	200.0	131.62	17.538	307.586
cho	303	126	564.0	246.26	51.831	2686.427
phrs	303	71	202.0	149.65	22.905	524.646
opeak	303	0	6.2	1.04	1.161	1.348
nflu	303	0	4.0	0.73	1.023	1.046

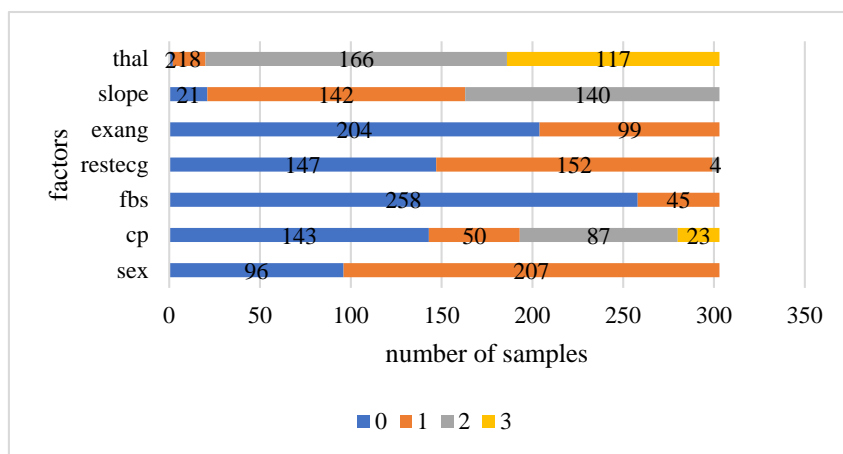


Fig. 1 Descriptive statistics of categorical variables

3.2 Results

Table 3. Hosmer-Lemeshow good of fit test

χ^2	df	p
6.535	8	0.587

The test of HL fit is employed to evaluate the model’s level of fit. Table 3 demonstrates that: The initial premise for this model testing is that there is a consistent level of agreement between the fitted and observed values of the

model; the p-value is over 0.05 ($\chi^2 = 6.535$, $p=0.587$), illustrating acceptance of the null hypothesis. This indicates that the framework fit well and passed the HL test.

Table 4. Model results

elements	regression coefficient	standard error	p-value	exp (β)	95% CI	
					low limit	upper limit
age	-0.001	0.024	0.981	0.999	0.954	1.047
sex (1)	-1.515	0.521	0.004	0.220	0.079	0.611
cpt			<0.001			
cpt (1)	0.983	0.564	0.081	2.673	0.885	8.075
cpt (2)	1.945	0.477	<0.001	6.995	2.745	17.823
cpt (3)	2.016	0.651	0.002	7.508	2.097	26.873

restbps	-0.017	0.011	0.111	0.983	0.963	1.004
cho	-0.004	0.004	0.265	0.996	0.988	1.003
bs (1)	0.176	0.566	0.755	1.193	0.393	3.619
recg			0.304			
recg (1)	0.570	0.375	0.128	1.769	0.849	3.685
recg (2)	-0.277	2.267	0.903	0.758	0.009	64.517
phrs	0.017	0.011	0.111	1.017	0.996	1.039
eang (1)	-0.763	0.426	0.073	0.466	0.202	1.075
opeak	-0.489	0.226	0.030	0.613	0.394	0.954
slope			0.106			
slope (1)	-0.720	0.863	0.404	0.487	0.090	2.643
slope (2)	0.202	0.938	0.830	1.223	0.194	7.694
nflu	-0.833	0.204	<0.001	0.435	0.291	0.649
thalt			0.005			
thalt (1)	1.815	2.379	0.446	6.139	0.058	649.809
thalt (2)	1.853	2.290	0.418	6.381	0.072	568.263
thal (3)	0.473	2.301	0.837	1.605	0.018	146.031
constant	1.101	3.365	0.744	3.006		
Dependent variable: target, Cox & Snell R ² : 0.510, Nagelkerke R ² : 0.681						

As showing in Table 4, it can be seen that the age, sex, cpt, restbps, cho, bs, detect, phrs, eang, opeak, slope, nflu, thalt are independent variables, whereas the dependent

variable in a binary Logit regression analysis is the target. The model formula derived from the above table is

$$\ln \frac{p}{1-p} = 1.435 - 1.394 \times \text{sex}(1) + 1.751 \times \text{cpt}(1) + \dots - 0.868 \times \text{opeak} \quad (3)$$

We can use equation (3) to obtain the specific calculation formula (4) for p .

$$p = \frac{\exp(1.435 - 1.394 \times \text{sex}(1) + 1.751 \times \text{cpt}(1) + \dots - 0.868 \times \text{opeak})}{1 + \exp(1.435 - 1.394 \times \text{sex}(1) + 1.751 \times \text{cpt}(1) + \dots - 0.868 \times \text{opeak})} \quad (4)$$

Where p is the likelihood that target 1 will occur. The model derives its conclusion from Table 4 by determining whether the p -value is greater than 0.05 and extrapolating potential heart disease risk factors. The final specific analysis shows that age, $\text{cpt}(1)$, restbps, cho, bs, recg, phrs, eang, slope, thalt do not have an impact on the target because of their p -values are all greater than 0.05, but sex, $\text{cpt}(2)$, $\text{cpt}(3)$, opeak and nflu will have an impact on the target. In specific, the regression coefficient value of sex(1), opeak and nflu are both under 0, which means that both of them will have a negative impact on the target. The figure of $\text{cpt}(2)$ and $\text{cpt}(3)$ are all greater than 0, which means they will have a positive impact. In addition, the impact on the target can be analyzed by judging the regression coefficients and OR. On the one hand, the probability of women getting heart disease is higher than men. When opeak and nflu increases by one unit separately, the decreasing amplitude of the target is 61.3% and 43.5%.

On the other hand, the probability of people with non-anginal pain and asymptomatic get heart disease is 6.995 and 7.508 times higher than people with typical angina separately.

However, when discussing about the influence of age, there are some differences between the contrast of traditional thoughts. The older people are, the higher possibility of getting heart disease would be. Song et al. (2019) pointed this trend in their article [9], whereas it considered another factor altogether. But the data used in this paper shows differences in different stage of age, hence the error in data distribution will not have a significant impact on the results. The different result of this paper about the influence of age may owing to the influence of other factors. This article should not only consider the issue of age, but also consider other influencing factors comprehensively, which may be the main reason for different results. Zhao and his team found that the results of the age-period-co-

hort (APC) model show that the death rate and disability adjusted life years (DALY) rate of ischemic heart disease (IHD) attributed to a high salt diet in China displayed a trend that increased with age from 1990 to 2019. [12]. With the passage of time, the risk of male mortality and DALY showed a declining trend from 1990 - 1994 to 1995 - 1999, and an rising trend from 1995 - 1999 to 2010 - 2014, hence this research did not consider a sole effect of age, and the probability of getting sick does not increase with age, which may help to explain the irrelevance of age to heart disease in this paper.

As to another traditional factors, gender, there is a considerable body of research indicating that the risk and manifestation of heart disease vary depending on a person's gender. The data about gender in this article is unevenly distributed, so this will cause some small errors in the results of this article. However, males are typically more probably than women to suffer from CHD during their young time, according to a number of studies. In other way, after menopause, women's risk of CHD increases and becomes more comparable to that of men.

For instance, the American College of Cardiology pointed out that women were more easily to suffer from non-obstructive coronary artery disease and microvascular disease and coronary artery spasm, which were less com-

mon in men [13]. Additionally, women with myocardial infarction often presented with different symptoms than men and tended to delay seeking medical care, which can contribute to worse outcomes in some cases. Furthermore, a study published in the BMJ Global Health indicated that while CHD mortality rates had generally declined over recent decades, the patterns differed between men and women across various countries. In the majority of countries, the impact of aging on CHD death rates has diminished over time, but this attenuation has been more significant in men than in women [14]. In the Netherlands Heart Journal, it is noted that menopause and hormone replacement therapy significantly influence women's cardiovascular risk factors, such as plasma lipids and lipoproteins [15]. Thus, it can be said that there is a relationship between gender and the incidence of heart disease based on the research mentioned above as well as the findings of this article, but there is no definite indicator of who is more susceptible to the disease, whether it is female or male. This correlation is influenced by other factors within the body and cannot be considered separately. All the other relevant factors in this article, such as chest pain, ST depression and number of major fluoroscopy vessels are all related to heart disease and conform to general rules.

Table 5. Binary Logit Regression Prediction Accuracy Summary

0		predicted value		forecast accuracy	Prediction error rate
		1			
true value	0	111	27	80.4%	19.6%
	1	14	151	91.5%	8.5%
summary				86.5%	13.5%

From Table 5, the accuracy of this model is displayed, and the overall prediction accuracy has reached 86.5%, so it's acceptable that the model fits. More specifically, the forecast accuracy is 91.5% when the true value is 1.

3.3 Further simplification

Further simplification of the model may be done possibly which is shown below. As can be shown in Table 6, the HL test is employed to evaluate the simplified model's goodness of fit.

Table 6. Hosmer-Lemeshow good of fit test

χ^2	df	p
15.462	8	0.051

Table 7. Simplified model results

elements	regression coefficient	standard error	p-value	exp (β)	95% CI	
					low limit	upper limit
sex (1)	-1.394	0.377	<0.001	0.248	0.118	0.519
cpt			<0.001			
cpt (1)	1.751	0.464	<0.001	5.763	2.319	14.322
cpt (2)	2.418	0.407	<0.001	11.228	5.059	24.923
cpt (3)	2.319	0.584	0.002	10.170	3.240	31.926
nflu	-0.736	0.164	0.030	0.479	0.347	0.660
opeak	-0.868	0.175	<0.001	0.420	0.298	0.592
constant	1.435	0.388	<0.001	4.198		
Dependent variable: target, Cox & Snell R ² : 0.426, Nagelkerke R ² : 0.570						

From Table 8, the overall prediction accuracy has reached 82.5%, hence the model fitting is acceptable. Specifically, when the true value is 1, the forecast accuracy reached 86.7%.

Table 8. Binary Logit Regression Prediction Accuracy Summary

0		predicted value		forecast accuracy	Prediction error rate
		1			
true value	0	107	31	77.5%	22.5%
	1	22	143	86.7%	13.3%
summary				82.5%	17.5%

In contrast to earlier research, scientists typically employ single-factor analyses, such as environment, congenital diseases, and physical and mental factors. However, this analysis has certain limitations and does not comprehensively consider the influence of other factors. In contrast, the factors covered in this paper have a certain degree of comprehensiveness, which can not only effectively combine multiple variables to identify errors caused by single factors, but also expand the scope of forthcoming research, benefiting people, especially in high-risk populations and medical staff, to provide more specific detection and prevention directions, and to prevent and detect heart disease earlier, so as to take timely response strategies.

4. Conclusion

This paper selected different data focusing on influential factors that may be associated with heart disease. The study found that heart disease may be associated with some factors, like gender, type of chest pain, ST-depression caused by exercise and number of fluoroscopy vessels, most of which have not been previously noticed. Although the accuracy of this article may be affected due to the limited amount of data, the research method of this

article still has some certain advantages. This article did not use single-factor analysis method which was widely used in previous studies, but instead used a multi-factor model to comprehensively analyze the impact of relevant factors on the result. Without ignoring the mutual influence between factors makes the construction of regression equation more accurate. Additionally, there is a positive impact of this article’s findings on heart disease prevention and treatment. As to the impact of chronic diseases, whether factors such as chest pain really affect heart disease requires more data and more comprehensive statistical methods for further study. Analyzing the relevant influencing factors that may cause heart disease can minimize the number of people suffering from this as much as possible, and more accurately prevent and treat this disease, improving people’s quality of life and the cure rate of this.

References

[1] Wei Qin, Shi Weiwei, Gao Jinchai, Ma Xiaoyan, Cao Yajing, Sun Jixin&Cui Ze Joinpoint regression analysis of heart disease mortality trends in urban and rural China from 2004 to 2019 Chinese Journal of Cardiology (04), 371-376.

- [2] Wang Fang (2024-05-06). Being angry for 8 minutes may increase the risk of heart attack. *China Science Daily*, 002.
- [3] Qi Ran, Sheng Lily, Zhang Xinyu, Yang Hongxi&Wang Yaogang (2023). Association Study of Depression and Anxiety Symptoms with Heart Disease and Chronic Non communicable Respiratory System Diseases in Elderly People in China. *Prevention and Control of Chronic Diseases in China* (12), 881-886.
- [4] Fang Liumin (2023). Sadness increases the risk of heart disease. *Family Medicine* (12), 33.
- [5] Yang Guangping (2023). Insomnia increases the risk of heart disease. *Family Medicine* (12), 33.
- [6] Xu Rui (2024-03-11). Microplastics increase the risk of heart disease and stroke. *China Science Daily*, 002.
- [7] Cheng Jingwei, Qiao Junjun, Yin Zhen, Hu Junpeng, Shao Xiayan, Liu Yangqing&Yang Jun (2024). Interpretation of “Physical Activity, Cardiopulmonary Fitness, and Cardiovascular Health: Clinical Practice Statement of the American Society for the Prevention of Cardiology” *Chinese Journal of Evidence Based Medicine* (02), 220-227.
- [8] Li Wenjuan, Wan Yongchao, Guo Lu, Xu Wenqi, Gu Ping, Chen Shuohua& Xing Aijun (2023). The impact of different baseline blood pressure levels on early onset heart disease in young people under 40 years old. *Chinese Journal of Circulation* (12), 1267-1273.
- [9] Song Fang, Li Yunxia&Liu Suzing (2019). Epidemiological characteristics and influencing factors analysis of 260 middle-aged and elderly patients with chronic heart failure in the liberated area of Jiaozuo City. *Journal of Qiqihar Medical College* (10), 1269-1270.
- [10] Zhao Huajun, Pei Leilei, Gao Xianshun&Yang Xiao (2024). Monitoring of birth defects in perinatal infants and analysis of factors affecting congenital heart disease in Ankang City from 2017 to 2021. *China Maternal and Child Health Research* (05), 36-41.
- [11] Huang Jinfa&Huang Shicong (2023). Analysis of influencing factors of angina pectoris in elderly patients with type 2 diabetes and coronary atherosclerotic heart disease. *Zhongwai Medical* (33), 39-42.
- [12] Zhao Mingri&Yu Chuanhua (2024). Analysis of the trend of ischemic heart disease burden attributed to high salt diet in China from 1990 to 2019 based on age period cohort model *Public Health and Preventive Medicine* (03), 1-5.
- [13] The EUGenMed, Cardiovascular Clinical Study Group, Vera Regitz-Zagrosek, Sabine Oertelt-Prigione, Eva Prescott, Flavia Franconi, Eva Gerdts, Anna Foryst-Ludwig, Angela H.E.M. Maas, Alexandra Kautzky-Willer, Dorit Knappe-Wegner, Ulrich Kintscher, Karl Heinz Ladwig, Karin Schenck-Gustafsson, Verena Stangl, Gender in cardiovascular diseases: impact on clinical manifestations, management, and outcomes, *European Heart Journal*, Volume 37, Issue 1, 1 January 2016, Pages 24–34.
- [14] Bots SH, Peters SAE, Woodward M Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010 *BMJ Global Health* 2017;2:e000298.
- [15] Maas, A., Appelman, Y. Gender differences in coronary heart disease. *Neth Heart J* 18, 598–603 (2010).