ANALYSIS OF NON-CLINICAL RISK FACTORS FOR HEART DISEASE AND STROKE USING STATISTICAL METHOD

Sri Sumarlinda[1], Dr. Azizah Binti Rahmat[2], Assoc Prof Zalizah Binti Awang Long[3]
Malaysia Institute of Information Technology (MIIT)[1], Malaysia Institute of Information Technology (MIIT)[2], Malaysia Institute of Information Technology (MIIT)[3]
Email: sumarlinda.sri@s.unikl.edu.My[1], azizah@unikl.edu.my[2], zalizah@unikl.edu.my[3]

ARTICLE INFO
Received:
Revised:
Approved:

ABSTRACT

Heart disease and stroke are the main contributions to health disorder. Two factors influence the susceptibility of the disease, namely clinical risk factors and non-clinical factors. This study aims to analyze the effect of non-clinical risk factors on the susceptibility to heart disease and stroke. The non-clinical risk factors are stress management, age, obesity, genetics, smoking, gender, lifestyle (nutrition), lifestyle (timing rest), and physical activity. Analysis of the influence of non-clinical risk factors on susceptibility using statistical methods, namely descriptive statistics, normalization tests with the Kolmogorov-Smirnov test, validity tests, reliability tests, and correlation tests. The descriptive statistical tests show that the risk factors for stress management, obesity, and smoking have a more significant influence than others. While Gender and physical activity have more negligible effect than others. Testing with one sample Kolmogorov-Smirnov Test shows that the data is normally distributed. Validity testing produces 100% valid data. The reliability test using Cronbach's alpha of 9 non-clinical risk factor items resulted in a value of 0.684, which means reliable. Correlation test between 9 items of non-clinical risk factor shows significant between items.

KEYWORDS
non-clinical factor risk, heart disease and stroke, statistical method, normalization test, validity and reliability test

INTRODUCTION

Heart disease and Stroke are the fatal diseases that contribute to being the cause of death in the world. For Specific words in the medical field, heart disease is called cardiovascular disease and stroke is called cerebrovascular disease. More than 868,000 Americans die of heart disease, stroke, or other cardiovascular diseases every year—that's one-third of all US deaths. From the National Centre for Chronic Disease Prevention and

Health Promotion (NCCDPHP), these diseases also take an economic toll, costing $214 billion a year to our health care system and causing $138 billion in lost productivity from premature death alone. To reduce the causes of heart disease and stroke, it is important to know the factors that cause these diseases. There are two types of factors that cause heart disease and stroke, namely clinical and non-clinical factors. Cardiovascular disease also increased the risk to get the infectious disease infections such as covid 19 [1][2].

Clinical data consists of blood pressure, blood sugar, cholesterol and uric acid. For clinical data, measurement and data acquisition can be done easily and objectively. Non-clinical data consisted of obesity, smoking, age, genetics, lifestyle and physical activity. The addition of age after middle age, gender, lifestyle (smoking, physical activities and lifestyles) will increase the susceptibility to heart disease and stroke [3],[4],[5],[6], [7].

The current era and technological developments leads to new habits and lifestyles. There are several new lifestyles that have a negative effect or increase susceptibility to the disease. Nowadays, people tend to have less physical activities, high stress and pressure and followed by unhealthy food intake such fast food. As a result, non-clinical risk factors such as obesity, smoking, age, genetics, gender, lifestyle and physical activity emerged. Each of these risk factors are interrelated.

Analysis and identification of non-clinical disease risk factors is more difficult than clinical disease risk factors, this is because objectivity and data acquisition are more difficult. Of the non-clinical risk factors discussed, the risk factors for age and obesity are the easiest to analyze. As you age and obesity increases, your susceptibility increases. Genetic risk factors are statistically very influential, but problems in identification and measurement are associated with heart disease and stroke susceptibility. Gender affects the susceptibility of cardiovascular disease. Physical activity and smoking are risk factors that depend on the frequency of the activity. These two factors clearly affect the susceptibility to heart disease and stroke disease. Measurement standards for these two risk factors need to be determined, so that it will lead to the category of these risk factors. Lifestyle risk factors are related to many things, for example alcohol consumption, unhealthy food, lack of rest and others.

The Statistical methods have been widely used for data analysis and processing in the health sector. The Descriptive and inferences Statistic are used to provide a literature source of data for the training of medical personnel [8]. Statistical methods can be used to estimate mortality, incidence and mortality to incidence ratio due to cardiovascular diseases [9]. The statistical method used for analytical methods in the health sector is related to observation, reflection and interpretation [10].

In quantitative research, statistics is the main method used to generalize the research results. To draw conclusions from the influence of non-clinical risk factors that cause heart disease and stroke, this statistical method is needed. In addition, statistical analysis is very useful in generalizing the susceptibility to heart disease and stroke. To draw conclusions from these events, statistical measures can describe well. The influenced of non-clinical risk factors for heart diseases and stroke susceptibility more difficult to be measured than clinical risk factors.

The data collection is through online questionnaire. Statistical methods are used to support quantitative validation claims. Due to the problem of objectivity and validation of the measurement of non-clinical risk factors for cardiovascular disease susceptibility, an analytical method is needed. This study proposes a statistical method for analysis, validity and increasing objectivity of the effect of cardiovascular disease susceptibility. The statistical analysis method used is descriptive statistics, normalization, validity and reliability tests.

RESEARCH METHOD

This study aims to analyze the effect of non-clinical risk factors on susceptibility to heart disease and stroke using statistical methods.

The recognized non-clinical risk factors are obesity, smoking, age, gender, genetics, lifestyle and physical activity. This study uses a descriptive research approach.

For more details, the stages of this research can be seen in Fig. 1:

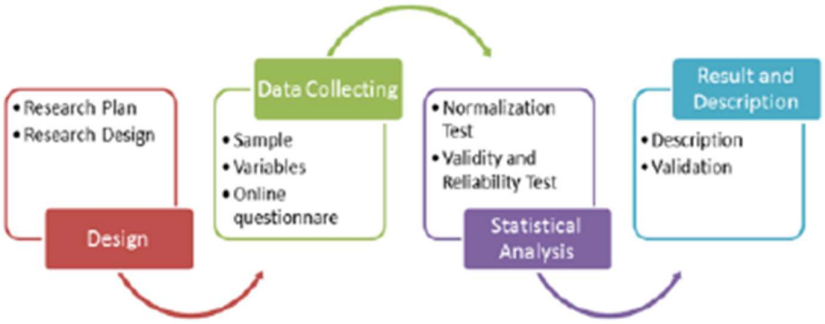

Fig 1. Research Stage

A. Sample and Variables of Research

The sample of this research is paramedic staff in Surakarta and surroundings area. The number of samples is 54 medical personnel. Because it is still during the COVID-19 pandemic, data collection is done by means of an online questionnaire. By applying to several clinics and hospitals for medical personnel to fill out the questionnaire.

The variable consists of several questions. Calculation of variables using a scale linker, namely strong agree (5), agree (4), neutral (3), disagree (2). strongly disagree (1). The list of variables can be seen in Table I:

Table I. List of variables

NO	Variable
1.	Will someone who cannot manage stress well affect the susceptibility to heart disease and stroke?
2..	Does increasing a person's age affect the susceptibility to heart disease and stroke?
3.	Will obesity affect susceptibility to heart disease and stroke?
4.	Do genetic factors affect susceptibility to heart disease and stroke?
5.	Can smoking affect susceptibility to heart disease and stroke?
6.	Can exercise habits affect susceptibility to heart disease and stroke?
7.	Will someone who has an unhealthy diet affect heart disease and stroke?
8.	Will inadequate and irregular rest and sleep patterns affect heart disease and stroke?
9.	Will gender affect susceptibility to heart disease and stroke?

B. Statistical Method

• Normalization Test

The Kolmogorov Smirnov (KS) test is a statistical test tool used to determine whether a sample comes from a population that has a certain distribution of data or follows a certain statistical distribution. The statistical distribution that is often tested using the Kolmogorov Smirnov (KS) test is the Normal Distribution. The Kolmogorov Smirnov (KS) test is based on the empirical distribution function (EDF) [8] [9][10] N sorted data X1, X2, X3, …Xn then the empirical distribution function is defined:

$EN=niN$ (1)

the variable ni is the number of points that are less than Xi, where the value of Xi is data that has been sorted from the smallest value to the largest value. the empirical distribution function (EDF) is a function that increases by 1/N at each data point.

The Normality test is a test carried out with the aim of assessing the distribution of data in a group of data or variables, whether the distribution of the data is normally distributed or not.no. Based on the empirical experience of several statistics, the data whose number is more than 30 numbers (> 30), then experts can already have a normal distribution. Commonly said as large sample. The Kolmogorov-Smirnov method has steps. Completion and use of the same formula, but with different significance. Significance The Kolmogorov-Smirnov method uses the Kolmogorov-Smirnov comparison table.

• Validity and Reliability Test

The data validity test is to measure how much the research questionnaire questions really measure the concept/variable being measured while the reliability test is to measure

the level of data consistency. The researcher followed the suggestion from to use the Average Variant Extracted (AVE) technique to measure validity and use construct reliability instead of Cronbach's alpha coefficient for reliability. The construct validity test includes convergent validity and discriminant validity. Samples of data and research instruments are said to be of high quality if they can meet the rule of thumb [11] [8] [10] . The validity used AVE, its formula shown bellow:

$AVE=\{\Sigma\gamma ini=1\}2\{\Sigma\gamma i\}ni=12+\{\Sigma\ Var(\in i)\}ni=1$ (2)

For the reliability test used construct reliability with formula:

CR= $r=nn-1[1-\Sigma si2ni=1st2]$ (3)

where :

r = realibility coeficient

n = number of item/questions

$si2$ = item score variant of ith item

$st2$ = total score variant

The validity test is a condition that describes whether the instrument that we use is able to measure what we are going to measure. The results obtained from the validity test are a valid or valid instrument. A high level of validity is best. On the other hand, an instrument that has low validity is an instrument that is not good or is not recommended and should even be removed from the indicator group.

The reliability is a series of measurements or a series of measuring instruments that have consistency if the measurements made with the measuring instrument are repeated. Test reliability is how much the degree of the test consistently measures the measured target. Reliability is expressed as a number, usually as a coefficient. High coefficient means high reliability. To calculate the reliability test of the description form test, it can be done using the Cronbach-Alph . formula.

RESULT AND DISCUSSION

descripted statistic, normalization test using Kolmogorov-Smirnov Test Validate using correlation and reliability using Cronbach's Alpha.

A. Sample Demografi & Desciptive Statistic

This study involved 54 respondents who are the medical personnel from several clinics and hospitals.. Of all the respondents, the age of the respondent is under 40 years, while for gender, 83.3% are female. From these respondents, data were obtained from 9 items in the online questionnaire.

The sample age & gender distribution is shown at Fig. 2 and Fig. 3 respectively:

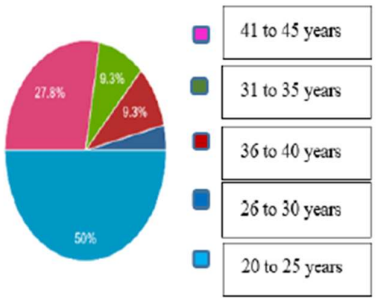

Fig 2. Sample Age Distribution

From the age distribution, it can be seen that the sample is productive age or under 40 years.

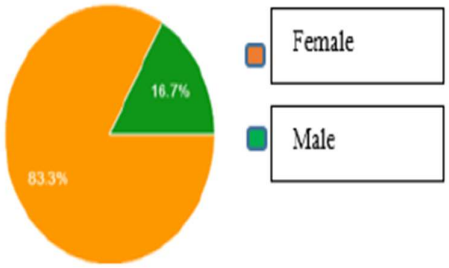

Fig 3. Sample Gender Distribution

Based on Fig. 3, it can be seen that the majority of the sample is female.

From descriptive statistical analysis, the distribution of input data was obtained as shown in Table II:

Table II. Analysis of Descriptive Statistic

	N	Range	Minimum	Maximum	Mean	Std Deviation
Q1	54	2	3	5	4.203	.483
Q2	54	3	2	5	3.593	.931
Q3	54	3	2	5	4.148	.764
Q4	54	3	2	5	3.815	1.027
Q5	54	4	1	5	3.926	1.160
Q6	54	4	1	5	2.833	1.174
Q7	54	4	1	5	3.870	1.072
Q8	54	3	2	5	3.833	.751
Q9	54	4	1	5	2.630	1.178

According to mean value from Q1 to Q9 are more than 2.5 so all non-clinical risk factor variable significant in cardiovascular disease susceptibility. The big tree of mean value areQ1 (4.203), Q3 (4.148) and Q5 (3.926). Stress Management, obesity, and smoking have strong effect to susceptibility cardiovascular disease. Q9 (2.630) and Q6 (2.388) have mean value less than 3. Gender and physical activity have less effect then others.

B. Normalization Test

The data was tested for normality with the one sample Kolmogorov-Smirnov Test. The complete results of the normalization test are as shown in Table III:

Table III. Normalization test

One-Sample Kolmogorov-Smirnov Test										
			Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
N			54	54	54	54	54	54	54	54	54
Normal Parameters [a,b]	Mean		4,20	3,59	4,15	3,81	3,93	2,83	3,87	3,83	2,63
	Std.Deviation		,491	,962	,787	,992	1,163	1,095	1,010	,666	1,121
Most Extreme Differences	Absolute		,420	,331	,259	,278	,266	,314	,310	,358	,250
	Positive		,420	,206	,223	,167	,178	,314	,190	,290	,250
	Negative		-,302	-,331	-,259	-,278	-,266	-,190	-,310	-358	-,167
Test Statistic			,420	,331	,259	,278	,266	,314	,310	,358	,250
Asymp. Carlo sig. (2-tailed)[c]			<,001	<,001	<,001	<,001	<,001	<,001	<,001	<,001	<,001
Monte Carlo Sig.(2-tailed)[d]	Sig.		˙000	˙000	˙000	˙000	˙000	˙000	˙000	˙000	˙000
	99% Confidence Interval	Lower Bound	˙000	˙000	˙000	˙000	˙000	˙000	˙000	˙000	˙000
		Upper Bound	˙000	˙000	˙000	˙000	˙000	˙000	˙000	˙000	˙000

From the table III the value of one sample Kolmogorov-Smirnov Test (Normalization test) Q1 0.420, Q2 0.331, Q3 0.359, Q4 0.278, Q5 0.266, Q6 0.314, Q7 0.310, Q8 0.358, Q9 0.250. From the result so the data is normal.

C. Validiti and Realibility Test

Testing the validity and reliability of the data using SPSS software with the results in table IV below:

Table IV Validity Test

RELIABILITY TEST			
Case Processing Summary			
		N	%
Cases	Valid	54	100,0
	Excluded[a]	0	,0
	Total	54	100,0
a. Listwise deletionbased on all variables in the procedure			

Based on Table IV, 54 respondents are valid 100%, so the influence of 9 non-clinical risk factors on cardiovascular disease susceptibility is valid.

Reliability testing using Cronbach's alpha obtained a value of 0.684 so that 9 items which are non-clinical risk factors are reliable as shown at Table V.

Table V. Reliability Test

Reliabiity Statistics	
Cronbach's Alpha	N of items
,684	9

The correlation between items Q1 to Q9 is used to determine the significance between items. The results of the correlation test are as shown in Table VI:

Table VI. Correlation Validity Test

VALIDATION TEST										
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Q1	Person Correlation	1	,339*	,262	,234	,060	,275*	,130	,221	,346*
	Sig.(2-tailed)		,012	,055	,089	,666	.044	,347	,108	,010
	N	54	54	54	54	54	54	54	54	54
Q2	Person Correlation	,399*	1	,580**	,078	,394**	-,030	,294*	,393**	,067
	Sig.(2-tailed)	,012		<,001	,577	,003	,830	,031	,003	,628
	N	54	54	54	54	54	54	54	54	54
Q3	Person Correlation	,262	,580**	1	-,013	589**	-,037	,286*	,372**	,235
	Sig.(2-tailed)	,055	<,001		,928	<,001	,793	,036	,006	,235
	N	54	54	54	54	54	54	54	54	54
Q4	Person Correlation	,234	,078	-,013	1	-,306*	,319*	-,024	-,105	,277*
	Sig.(2-tailed)	,089	,577	,928		,024	,019	,861	,451	,043
	N	54	54	54	54	54	54	54	54	54
Q5	Person Correlation	,060	,394**	,589**	-,306*	1	-,291*	,184	,398**	,167
	Sig.(2-tailed)	,666	,003	<,001	,024		,032	,182	,003	,228

	N	54	54	54	54	54	54	54	54	54
Q6	Person Correlation	,275*	-,030	-,037	,319*	-,291*	1	,202	,194	,533**
	Sig.(2-tailed)	,044	,830	,793	,019	,032		,143	,159	<001
	N	54	54	54	54	54	54	54	54	54
Q7	Person Correlation	,130	,294*	,286*	-,024	,184	,202	1	,640**	,323*
	Sig.(2-tailed)	,347	,031	,036	,661	,182	,143		<,001	,017
	N	54	54	54	54	54	54	54	54	54
Q8	Person Correlation	,221	,292**	,372**	-,105	,398**	194	,640**	1	,396**
	Sig.(2-tailed)	,108	,003	,006	,451	,003	,159	<,001		,003
	N	54	54	54	54	54	54	54	54	54
Q9	Person Correlation	,346*	,067	,235	,277*	,167	,523*	,323*	,396**	1
	Sig.(2-tailed)	,346*	,628	,088	,043	,228	<,001	,017	,003	
	N	54	54	54	54	54	54	54	54	54

Table VI indicated that the correlation for each item Q1 to Q9 is significant

CONCLUSION

The non-clinical risk factor which consists of Management stress, age, obesity, genetic, smoking, life style (nutrition), life style (timing rest) and physical activity. From the statistical descriptive analysis, stress management, obesity, and smoking have a greater influence than other risk factors. While Gender and physical activity have less effect then others. By using the Kolmogorov-Smirnov Test, it results that the data used is normally distributed

Testing the validity of 54 samples is 100% valid. Reliability testing of 9 items with Cronbach's alpha is reliable. Correlation testing shows that there is a significant correlation between items. So, stress management, age, obesity, genetics, smoking, gender, life style (nutrition), life style (timing rest) and physical activity statistically affect the susceptibility of heart diseases and stroke.

REFERENCES

[1] F. Mai, R. Del Pinto, and C. Ferri, "COVID-19 and cardiovascular diseases," J. Cardiol., vol. 76, no. 5, pp. 453–458, 2020, doi: 10.1016/j.jjcc.2020.07.013.

[2] J. Liu, C. Li, J. Xu, and H. Wu, "A patient-oriented clinical decision support system for CRC risk assessment and preventative care," BMC Med. Inform. Decis. Mak., vol. 18, no. Suppl 5, 2018, doi: 10.1186/s12911-018-0691-x.

[3] C. Chang, C. Huang, Y. Chou, J. Chang, and C. Sun, "Association Between Age‐Related Macular Degeneration and Risk of Heart Failure: A Population‐Based Nested Case‐Control Study," J. Am. Heart Assoc., pp. 1‐8, 2021, doi: 10.1161/jaha.120.020071.

[4] M. El‐Chouli et al., "Time Trends in Simple Congenital Heart Disease Over 39 Years: A Danish Nationwide Study," J. Am. Heart Assoc., vol. 10, no. 14, 2021, doi: 10.1161/jaha.120.020375.

[5] J. Siddique et al., "Relative‐Intensity Physical Activity and Its Association With Cardiometabolic Disease," J. Am. Heart Assoc., vol. 10, no. 14, pp. 1‐3, 2021, doi: 10.1161/jaha.120.019174.

[6] K. S. Zachrison et al., "Strategy for reliable identification of ischaemic stroke, thrombolytics and thrombectomy in large administrative databases," Stroke Vasc. Neurol., vol. 6, no. 2, pp. 194–200, 2021, doi: 10.1136/svn-2020-000533.

[7] J. Yang, L. Xiao, and K. Li, "Modelling clinical experience data as an evidence for patient-oriented decision support," BMC Med. Inform. Decis. Mak., vol. 20, no. Suppl 3, pp. 1–11, 2020, doi: 10.1186/s12911-020-1121-4.

[8] M. J. Hayat, A. Powell, T. Johnson, and B. L. Cadwell, "Statistical methods used in the public health literature and implications for training of public health professionals," PLoS One, vol. 12, no. 6, pp. 1–10, 2017, doi: 10.1371/journal.pone.0179032.

[9] M. Amini, F. Zayeri, and M. Salehi, "Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017," BMC Public Health, vol. 21, no. 1, pp. 1–12, 2021, doi: 10.1186/s12889-021-10429-0.

[10] R. Ocaña-riola, "The Use of Statistics in Health Sciences : Situation Analysis and Perspective," no. October 2016, 2018, doi: 10.1007/s12561-015-9138-4.

[11] L. Moyé, "Statistical methods for cardiovascular researchers," Circ. Res., vol. 118, no. 3, pp. 439 453, 2016, doi: 10.1161/CIRCRESAHA.115.306305.

[12] D. C. Yadav and S. Pal, "Prediction of heart disease using feature selection and random forest ensemble method," Int. J. Pharm. Res., vol. 12, no. 4, pp. 56–66, 2020, doi: 10.31838/ijpr/2020.12.04.013.

[13] D. Normawati and D. P. Ismi, "K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining," Signal Image Process. Lett., vol. 1, no. 2, pp. 23–35, 2019, doi: 10.31763/simple.v1i2.3.

[14] C. Series, "Comparison of coronary heart disease stratification using the Jakarta cardiovascular score between main office and site office workers Comparison of coronary heart disease stratification using the Jakarta cardiovascular score between main office and site ," 2018.

[15] S. G. Kwak and S.-H. Park, "Normality Test in Clinical Research," J. Rheum. Dis., vol. 26, no. 1, p. 5, 2019, doi: 10.4078/jrd.2019.26.1.5.

[16] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: A guide for non-statisticians," Int. J. Endocrinol. Metab., vol. 10, no. 2, pp. 486–489, 2012, doi: 10.5812/ijem.3505.

[17] H. Taherdoost, "Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research," SSRN Electron. J., no. September, 2018, doi: 10.2139/ssrn.3205040.–