Study of factors caused diabetes mellitus and stroke diseases toward gender

Sudarno, T Widiharh
Department of Statistics, Faculty Sciences and Mathematics, Diponegoro University, Semarang, Indonesia
Corresponding author: dsghani@gmail.com

Abstract. Gender consists of female and male. Diabetes mellitus is a chronic disease characterized by high levels of sugar in the blood. Glucose is controlled by the hormone insulin which is produced by the pancreas. In diabetics, the pancreas is unable to produce insulin according to the body's needs. Blood sugar tests can be done to find out someone has diabetes. This method by a time blood sugar test and fasting blood sugar test. Stroke is a condition that occurs when the blood supply to the brain is interrupted or reduced due to a blockage (ischemic stroke) or rupture of a blood vessel (hemorrhagic stroke). We can use the proportional hazards regression and the stratified proportional hazards model to determine the factors causing diabetes mellitus and stroke. The significant covariate could be seen as a factor causing disease. Significant covariate also yields hazard ratio. If the event time assumed has loglogistics distribution. Some of the functions discussed in this paper, namely the cumulative distribution function, survival function, hazard rate function, and cumulative hazard function. These functions are useful for prediction tools provided the event time is known. The purpose of this study was to find the factors that cause diabetes mellitus and stroke in the gender strata and making functions of random variables with the logistic distribution. The results obtained from this study are the factors that influence a person at risk of stroke namely the type of stroke, hypertension, systolic blood pressure, diastolic blood pressure, and diabetes mellitus. The factors that affect a person at risk of stroke according to gender are the type of stroke, hypertension, systolic blood pressure. If the sample includes all women who have had a stroke, it is primarily affected by the type of stroke. If the sample consists of all men who have had a stroke, it is concluded that the significant covariates are type of stroke, hypertension, systolic blood pressure, and diabetes mellitus. Men have a higher rate of stroke than women.

1. Introduction
The people has a characteristic gender (sex). Gender consists of female and male. A person can get diabetes mellitus or stroke. Diabetes mellitus is a chronic disease characterized by high levels of sugar (glucose) in the blood. Blood glucose is controlled by the hormone insulin, which is produced by the pancreas. If the pancreas is unable to produce insulin according to the body's needs, a person will have diabetics. Some of the common symptoms of diabetes are frequent thirst, frequent urination, especially at night, often feeling very hungry, losing weight for no apparent reason, weakness, blurred vision, wounds that are difficult to heal, frequent infections, ketones in the urine [1]. Ketones are a waste product from the breakdown of muscle and fat because the body cannot use sugar as an energy source. A blood sugar test is an absolute test to diagnose diabetes. The results of the glucose measurement will
show whether a person has diabetes or not. Blood sugar test methods that can be undertaken by patients include a temporary blood sugar test and a fasting blood sugar test [1].

Stroke is a condition that occurs when the blood supply to the brain is interrupted or reduced due to a blockage (ischemic stroke) or rupture of a blood vessel (hemorrhagic stroke). The brain will not get oxygen and nutrients if no blood, so cells in some areas of the brain will die. The parts of the body controlled by the damaged areas of the brain cannot function properly if some areas of the brain die. Ischemic stroke has a more frequent incidence than hemorrhagic stroke, but hemorrhagic stroke causes death relatively quickly than ischemic stroke. The most common symptoms of stroke are leg numbness, speech becomes difficult and the face looks asymmetrical [1,6].

Diabetes mellitus and stroke can caused by several factors. the factors causing diabetes mellitus and stroke were searched using a proportional hazards regression. The proportional hazards regression is assumed that the covariates are not time dependent. Comparison between hazard function is a constant. The significant covariate shows the factor that cause disease. There is a nonproportional hazards model, namely the stratified proportional hazards. It is a proportional hazards model but there is stratified covariate and not significant. This covariate is used as a stratum. It is an insignificant covariate and does not depend on time. The covariate that can be used as strata is gender. The event time assumed has logistics distribution. Some of the functions discussed in this paper are the cumulative distribution function, survival function, hazard rate function and cumulative hazard function [10,11,12]. The functions are useful for prediction tools provided the event time is known.

Several papers related to this research, namely [2] use Cox regression model with nonproportional hazards applied to lung cancer survival data shows that cigarette consumption, tumor size and tumor stage are found to be important risk factors which affect the failure. In [5] made a comparison of exact, Efron and Breslow parameter approach method on hazard ratio and stratified Cox regression models about lung cancer patients. The result shows that the exact method of stratified Cox regression model is better than other. The stroke factor was influenced by body mass index, smoking, and diabetes mellitus. Body mass index is positive effect, while smoking and diabetes mellitus are negative effect [13]. The purpose of this study was to find the factors that cause diabetes mellitus and stroke, using gender as strata, according to female and male sex, and making functions of random variables with logistolistic distribution. We can prevent that diseases by knowing the factors that cause diseases. The distribution functions of random variables with logistic distribution could be used to predict possible events for another time.

2. Literature review

2.1. Proportional hazards model
The hazard function of the different patient are assumed to be proportional (constant) and independent of time. This means that the ratio of the risk of an event of two patients is the same no matter how long them survive. It is assumed that there is only one cause of failure, that is, failure is only allowed to occur once per patient, and there is no correlation between the times of failure of different patient. The hazard function given a set of covariates $x = (x_1, x_2, ..., x_p)$ can be written as a function of an underlying hazard function $h_0$ and a function $g(x)$ of only the covariates, that is

$$h(t \mid x) = h_0(t) \cdot g(x), \text{ where } t \text{ is the event time.} \tag{1}$$

The underlying hazard function $h_0(t)$ represent the risk changes with time and $g(x)$ represent the effect of covariates. The function $h_0(t)$ can be interpreted as the hazard function when all covariates are ignored that is when $g(x) = 1$, so it is called the baseline hazard function. The proportional hazards model assumes that $g(x)$ is an exponential function of covariates [3]. Therefore, this function is
\[ g(x) = \exp\left( \sum_{j=1}^{p} b_j x_j \right) = \exp(b'x), \]  

where \( b = (b_1, b_2, \ldots, b_p) \) denotes the coefficients of covariates.

The hazard function or the proportional hazards model becomes

\[ h(t \mid x) = h_0(t) \exp(b'x) \]  

(3)

The coefficients of covariates can be estimated from the observed data and they indicate the magnitude of the effect of the hazard ratio \([3,10,11,12]\).

**The hazard proportion assumption**

The assumed hazard proportion is met by using the concept of correlation between two variables. If the two variables are correlated, the hazard proportion assumption does not meet. But if the two variables are uncorrelated, the hazard proportion assumption is fulfilled \([10,11,12]\). The test steps are

**Hypotheses:**

\( H_0 \): there is no correlation between the rank variable and the residual Schoenfeld value each variable.

\( H_1 \): there is a correlation between the rank variable and the residual Schoenfeld value each variable.

**Level of Significance** is \( \alpha \)

**Test Statistics:**

\[ r_h = \frac{\sum_{i=1}^{n} (R_{ji} - \bar{R}_{ji})(RT_i - \bar{RT}_i)}{\sqrt{\sum_{i=1}^{n} (R_{ji} - \bar{R}_{ji})^2} \sqrt{\sum_{i=1}^{n} (RT_i - \bar{RT}_i)^2}} \]  

(4)

where \( r_h \) : hazard correlation coefficient; \( R_{ji} \) : residual Schoenfeld for each covariate \( j \)th, with \( j = 1, 2, \ldots, p \) for individual \( i \)th with \( i = 1, 2, \ldots, n \); \( RT_i \) : individual life time \( i \)th; \( \bar{RT}_i \) : the mean of individual life time \( i \)th.

**Decision:** \( H_0 \) rejected if \( \left| r_h \right| > r_{n-2;\frac{\alpha}{2}} \) or \( p \)-value < \( \alpha \).

**Wald test**

The Wald test was used to determine the significance of each covariate on survival time \([10,11,12]\).

**Hypotheses:**

\( H_0 \) : covariate \( j \)th has no effect on survival time.

\( H_1 \) : covariate \( j \)th effect on survival time, \( j = 1, 2, \ldots, p \).

**Level of Significance** is \( \alpha \)

**Test Statistics:**

\[ Z^2 = \left( \frac{b_j}{se(b_j)} \right)^2, \]  

where \( se(b_j) = \sqrt{\var(b_j)} \).

**Decision:** \( H_0 \) rejected if \( Z^2 > \chi^2_{1,\alpha} \) or \( p \)-value < \( \alpha \).

**Hazards ratio**

The hazards ratio is the ratio of the failure of one group of patients to the failure of another group of individuals. The two groups of patients being compared are distinguished by their covariate values, but their regression coefficients are the same \([10,11,12]\). The hazards ratio is stated by formula

\[ HR = \exp\left( \sum_{j=1}^{p} b_j (x_j^* - x_j') \right) \]  

(5)
where \( b_j \): coefficient of regression covariate \( j \)th, with \( j = 1, 2, \ldots, p \); \( x_j^* \): covariate value \( j \)th, the first patient; \( x_j \): covariate value \( j \)th, the other patient.

2.2. Stratified Proportional Hazards Model

The stratified proportional hazards model is part of the nonproportional hazards model. The nonproportional hazards model is the proportional hazards model generalized using stratification concept [10]. The data will be stratified by gender. We consider \( m \) strata, equation (3) becomes [3]

\[
h_k(t | x_i) = h_{0k}(t)\exp\left(\sum_{j=1}^{p} b_j x_j \right) = h_{0k}(t)\exp(b^T x_k),
\]

where \( k = 1, 2, \ldots, m \) for the \( m \) strata.

The underlying hazard function \( h_{0k}(t) \) is assumed to be different for the \( m \) strata, but the regression coefficients are the same for all strata. We assume that the hazard functions for patients may be proportional within every stratum. The hazard functions are not proportional among different strata. The marginal likelihood function for all observations from the \( m \) strata is defined by

\[
L(b) = \prod_{k=1}^{m} L_k(b), \text{ where } L_k(b) \text{ is the marginal likelihood function for the } k\text{th stratum.}
\]

The regression coefficient \( b \) will be estimated by the Newton-Raphson method [10,11,12].

2.3. Loglogistic distribution function

A random variable \( T \) has a loglogistic distribution if \( \ln t \) has logistic distribution. If random variable \( T \) has a loglogistic distribution, it has, respectively[8,9,12]

- The probability density function \( f(t; \alpha, \gamma) = \left(\alpha \gamma t^{-1}\right) \left(1 + \alpha t^\gamma\right)^{-2}, \; 0 < t < \infty \) (8)
- The cumulative distribution function \( F(t; \alpha, \gamma) = \left(\alpha \gamma \right) \left(1 + \alpha t^\gamma\right)^{-1} \) (9)
- The survival function \( S(t; \alpha, \gamma) = \left(1 + \alpha t^\gamma\right)^{-1} \) (10)
- The hazard rate function \( h(t; \alpha, \gamma) = \left(\alpha \gamma t^{-1}\right) \left(1 + \alpha t^\gamma\right)^{-1} \) (11)
- The cumulative hazard function \( H(t; \alpha, \gamma) = \ln\left(1 + \alpha t^\gamma\right), \; t \geq 0, \; \alpha > 0, \; \gamma > 0 \) (12)

2.4. Diabetes mellitus and stroke diseases

2.4.1. Diabetes mellitus. Diabetes mellitus is a chronic disease characterized by high blood sugar (glucose) levels. Glucose is the main energy source for human body cells. Glucose that accumulates in the blood due to various organ disorders. If diabetes is not well controlled, various complications can arise that can endanger the sufferer life. Blood sugar levels are controlled by the hormone insulin, which is produced by the pancreas. If the pancreas is unable to produce insulin according to the body needs a person become diabetic. The body cells cannot absorb and process glucose into energy without insulin. Diabetes is divided into two types, namely type 1 and type 2. Type 1 occurs if the patient's immune system attacks and destroys the pancreatic cells. This results in an increase in blood sugar levels, resulting in damage to body organs. Type 2 is caused by the body cells are less sensitive to insulin, so the insulin produced cannot be used properly [1].

Some of the common symptoms of diabetes are frequent thirst, frequent urination, especially at night, often feeling very hungry, losing weight for no apparent reason, weakness, blurred vision, wounds that are difficult to heal, frequent infections, ketones in the urine. Blood sugar test is an absolute test to diagnose this type of diabetes. The results of measuring blood sugar will show whether
a person has diabetes or not. If the current blood sugar test shows the sugar level is 200 mg/dL or more, the patient can be diagnosed with diabetes. Alternative test by fasting blood sugar test aims to measure blood sugar levels when the patient is fasting. The patient will be asked to fast for 8 hours first. The results of a fasting blood sugar test that showed a fasting blood sugar level of 126 mg/dL or more indicated that the patient had diabetes [1].

2.4.2. Stroke. Stroke is a condition that occurs when the blood supply to the brain is interrupted or reduced due to a blockage (ischemic stroke) or rupture of a blood vessel (hemorrhagic stroke). The brain will not get oxygen and nutrients without blood, so cells in some areas of the brain will die. When some areas of the brain die, the parts of the body controlled by the damaged areas of the brain cannot function properly. Prompt treatment can minimize brain damage and the possibility of coagulation. Ischemic stroke has a more frequent incidence than hemorrhagic stroke, but hemorrhagic stroke is more causes death than ischemic stroke. The most common symptoms of stroke are leg numbness, speech becomes difficult and the face looks asymmetrical. Stroke can be prevented by adopting a healthy diet, exercising regularly, avoiding smoking and consuming alcoholic beverages [1,6].

3. Result and discussion
The application of this research in the health sector is to analyze a person has diabetes mellitus and stroke diseases. The purpose of this study was to find factors that influence diabetes mellitus and stroke based on gender. The type of data used is secondary data. The data was obtained from the medical records section of patients with stroke at the Tugurejo Regional General Hospital, Semarang City, for the period January 2018 to December 2018. Data processing used software R [4,7,12].

The variables used in this study were stroke event time T, censor cens, gender X1, age X2, type of stroke X3, hypertension X4, systolic blood pressure X5, diastolic blood pressure X6, diabetes mellitus X7 and body mass index X8. The response variable is stroke event time with patient status as indicator variable cens, while covariate variables are gender, age, sex, type of stroke, hypertension, systolic blood pressure, diastolic blood pressure, diabetes mellitus, and body mass index.

3.1. Factors caused diabetes mellitus and stroke
We want to determine the factors affecting patient with diabetes mellitus and stroke. First, data were processed using proportional hazards regression. The results are summarized in the table below

| Covariate | Coefficient | z   | Pr(>|z|)  | Sign |
|-----------|-------------|-----|----------|------|
| X1        | -0.1448     | -0.44 | 0.6568  | -    |
| X2        | 0.0610      | 0.20 | 0.8392  | -    |
| X3        | 1.9784      | 5.88 | 4.11e-09 | ***  |
| X4        | -1.4174     | -2.72 | 0.0065  | **   |
| X5        | 0.0207      | 2.47 | 0.0135  | *    |
| X6        | -0.0223     | -1.90 | 0.0581  | .    |
| X7        | 0.0018      | 1.74 | 0.0812  | .    |
| X8        | 0.3044      | 0.93 | 0.3537  | -    |

If we take the significance level $\alpha = 10\%$, and based on table 1. The influential covariates that affect the event time of patients with stroke are $X_3, X_4, X_5, X_6$ and $X_7$. The sign are $p$-value less than $10\%$, asteric or point. Meanwhile, the covariates did not affect the event time of patients with stroke were covariates $X_1, X_2$, and $X_8$. The assumption for the proportional hazards model that all covariates are time independent. To find out the fulfillment of this assumption using the correlation method. If
the covariates are not correlated, it means that the two covariates are considered independent. If all
covariates are considered independent, it means that the covariates are considered independent of time.
Testing the proportional hazards assumption using the Schoenfeld residuals with the hypothesis test
below.
Hypotheses:
H₀ : ρₜ = 0 (the proportional hazard assumption is met)
H₁ : ρₜ ≠ 0 (the proportional hazard is not met)
If the level of significance α = 10% is taken, the correlation coefficient is obtained between the residual Schoenfeld value and the event time be presented in table 2 as follow.

| Covariate | rₜ | p-value | Decision |
|-----------|----|---------|----------|
| X₁        | 0.1317 | 0.321 | Accept H₀ |
| X₂        | 0.0156 | 0.916 | Accept H₀ |
| X₃        | 0.0309 | 0.812 | Accept H₀ |
| X₄        | 0.0303 | 0.824 | Accept H₀ |
| X₅        | -0.0082 | 0.949 | Accept H₀ |
| X₆        | 0.0936 | 0.487 | Accept H₀ |
| X₇        | -0.0785 | 0.712 | Accept H₀ |
| X₈        | -0.1597 | 0.261 | Accept H₀ |

Based on the product moment correlation coefficient with a two-way test, it is obtained that the value r = 0.195. Then the decision hypothesis of all covariates is H₀ is accepted. So the residual Schoenfeld value and the event time are not correlated. This means that all covariates meet the proportional hazards assumption. The next discussion is to find the best model using the backward method. The best regression model is a model where all covariates are significant. The computational results to get the best model are presented in the following table.

| Covariate | Regression Coefficient | Z     | Pr(>|z|) | Sign |
|-----------|------------------------|-------|---------|------|
| X₃        | 1.9851                 | 5.926 | 3.1e-09 | ***  |
| X₄        | -1.3377                | -2.624| 0.0087  | **   |
| X₅        | 0.0204                 | 2.566 | 0.0103  | *    |
| X₆        | -0.0227                | -2.076| 0.0380  | *    |
| X₇        | 0.0020                 | 1.969 | 0.0489  | *    |

Next, the Wald test is used to test the best model. The hypothesis test with the following steps.
Hypotheses:
H₀ : covariate jth has no effect on survival time.
H₁ : covariate jth has effect on survival time, j = 3,4,5,6,7.

If the significance level is α = 5%, the statistical values are obtained as shown in table 4.
Table 4. Wald test result of the best covariate significance

| Covariate | Regression Coefficient | Hazard Ratio | p-value       | Sign     | Decision |
|-----------|------------------------|--------------|---------------|----------|----------|
| $X_3$     | 1.9851                 | 7.28         | 3.1e-09       | ***      | Reject $H_0$ |
| $X_4$     | -1.3377                | 0.26         | 0.0087        | **       | Reject $H_0$ |
| $X_5$     | 0.0204                 | 1.02         | 0.0103        | *        | Reject $H_0$ |
| $X_6$     | -0.0227                | 0.98         | 0.0380        | *        | Reject $H_0$ |
| $X_7$     | 0.0020                 | 1.00         | 0.0489        | *        | Reject $H_0$ |

Because all p-values are smaller than the level of significance or there is a asteric sign, all $H_0$ are rejected. This means that all covariates affect the event time. Therefore, the factors that affect a patient at risk of stroke are the type of stroke, hypertension, systolic blood pressure, diastolic blood pressure, and diabetes mellitus. The best model can be written as [3]

$$h(t \mid X) = h_0(t) \exp(b \cdot X)$$  \hspace{1cm} (13)

$$b = (1.9851 \hspace{0.1cm} -1.3377 \hspace{0.1cm} 0.0204 \hspace{0.1cm} -0.0227 \hspace{0.1cm} 0.0020)^\top; \hspace{0.1cm} X = (X_3 \hspace{0.1cm} X_4 \hspace{0.1cm} X_5 \hspace{0.1cm} X_6 \hspace{0.1cm} X_7)^\top$$  \hspace{1cm} (14)

3.2. Factors caused diseases by strata

In the previous discussion, gender covariate was an insignificant covariate. Therefore, it can be used as a strata. The strata variable is a variable that has no effect but is still included in the process with the hope of being useful. A discussion was carried out with covariates consisting of all significant covariates plus the gender strata variable. The response variable is the event time with the indicator variable is cens. The results obtained are in the table 5.

Table 5. Regression coefficient and hazard ratio

| Covariate | Regression Coefficient | Hazard Ratio | Z         | Pr($>\mid z \mid$) | Sign     |
|-----------|------------------------|--------------|-----------|---------------------|----------|
| $X_3$     | 1.9418                 | 6.97         | 5.82      | 5.95e-09            | ***      |
| $X_4$     | -1.3932                | 0.25         | -2.72     | 0.0066              | **       |
| $X_5$     | 0.0203                 | 1.00         | 2.47      | 0.0134              | *        |

Based on table 5, it means that for gender strata consisting of female and male, significant covariates for stroke and diabetes sufferers are the type of stroke, hypertension, systolic blood pressure. Based on the hazard ratio value, it can be said that

a) The risk of a person having ischemic stroke is 5.97 times greater than hemorrhagic stroke.

b) The risk of a person having a stroke who has hypertension failing to survive is 3 times higher than that who does not have hypertension.

c) Increasing systolic blood pressure results in a stroke 0.02 times higher than a person with normal systolic blood pressure.

The regression model for gender strata is[3,12]

$$h(t \mid X) = h_0(t) \exp(b \cdot X)$$  \hspace{1cm} (15)

with $b = (1.9418 \hspace{0.1cm} -1.3932 \hspace{0.1cm} 0.0203)^\top$ and $X = (X_3 \hspace{0.1cm} X_4 \hspace{0.1cm} X_5)^\top$.

3.3. Factors caused disease toward gender
This following, data processing was carried out by dividing disease sufferers based on sex, namely female and male. There are two kinds of samples, namely a group of stroke patients with female and male sex. These two samples are assumed to be independent to each other.

**Sample for female**
After data processing with all covariates was complete, only the covariate type of stroke was significant. Therefore, a woman who has a stroke is primarily affected by the stroke type covariate. If it be informed the result as Table 6.

| Covariate | Regression Coefficient | Hazard Ratio | Z   | Pr(>|z|)         | Sign |
|-----------|------------------------|--------------|-----|-----------------|------|
| X₃        | 1.84                   | 6.3          | 3.9 | 9.37e-05        | ***  |

Based on table 6, it can be explained that the risk of a woman having ischemic stroke is 5.3 times greater than hemorrhagic stroke.

**Sample for male**
After processing the data with all the complete covariates, the significant covariates were type of stroke, hypertension, systolic blood pressure, and diabetes mellitus. Therefore, a man who had a stroke was affected by these four covariates. The result is shown in the following table.

| Covariate | Regression Coefficient | Hazard Ratio | Z    | Pr(>|z|)         | Sign |
|-----------|------------------------|--------------|------|-----------------|------|
| X₃        | 2.08                   | 8.0          | 4.51 | 6.63e-06        | ***  |
| X₄        | -1.65                  | 0.19         | -2.49| 0.01            | *    |
| X₅        | 0.03                   | 1.03         | 2.51 | 0.01            | *    |
| X₇        | 0.01                   | 1.01         | 2.52 | 0.01            | *    |

Based on table 7, it can be taken decision that
a) The risk of a man having ischemic stroke is 7 times greater than having hemorrhagic stroke.

b) The risk of a man having a stroke who has hypertension failing to survive is 4.26 times higher than who does not have hypertension.

c) With the increase in systolic blood pressure resulted in a man having a stroke 0.03 times higher than a person with normal systolic blood pressure.

d) The risk of a man getting strokes who have diabetes mellitus is 1% higher than a man who does not have diabetes mellitus.

### 3.4. Distribution functions of event time
The event time $T$ is the length of time a stroke patient is hospitalized (days). The random variable $T$ assumed has a loglogistic distribution. Based on data processing, the Intercept value is 3.904 and the Scale value is 0.776 so their parameters are [12]

\[
\alpha = \exp(- \text{Intercept}) (\text{Scale}^{-1}) = \exp(- 3.904/0.776) = 0.0065, \quad (17)
\]

\[
\gamma = (\text{Scale})^{-1} = (0.776)^{-1} = 1.29 \quad (18)
\]

Therefore, the cumulative distribution function as

\[
F(t) = \left(0.0065t^{1.29}\right) \left(1 + 0.0065t^{1.29}\right)^{-1} \quad (19)
\]
The survival function is
\[ S(t) = \left(1 + 0.065t^{1.29}\right)^{-1} \tag{20} \]

If we illustrate the relationship between the cumulative distribution and the survival functions toward the event time. It be presented by figure 1 below[4,7]

![Cumulative distribution and survival functions](image)

**Figure 1.** Cumulative distribution and survival functions.

In figure 1, the cumulative distribution function is an increasing function, while the survival function is a decreasing function. The cumulative distribution function states the chance that a person has a stroke to survive before the event time, while the survival function states the chance that a person has a stroke to survive after the event time.

The hazard rate function as
\[ h(t) = (0.0065) (1.29)^{0.29} \left(1 + 0.0065t^{1.29}\right)^{-1} \tag{21} \]

The cumulative hazard function as
\[ H(t) = \ln\left(1 + 0.0065t^{1.29}\right), \quad t \geq 0. \tag{22} \]

The following shows the relationship between the event time both the hazard function and the cumulative hazard function as in figure 2[4,7].
Figure 2. Hazard rate and cumulative hazard functions.

Figure 2 be presented hazard $h(t)$ and cumulative hazard $H(t)$ functions. The hazard rate function states the risk of failure per unit time during the living process. The hazard rate function is an increase function in this case. The cumulative hazard function expresses the cumulative hazard function over the event time. This function is also an increase function. The cumulative hazard function increases faster than the hazard rate function. If the data simulation is carried out for the selected time, it is presented in table 8 below.

| $t$  | $F(t)$ | $S(t)$ | $h(t)$ | $H(t)$ |
|-----|--------|--------|--------|--------|
| 0   | 0.000  | 1.000  | 0      | 0      |
| 3   | 0.028  | 0.972  | 0.012  | 0.029  |
| 5   | 0.054  | 0.946  | 0.014  | 0.055  |
| 8   | 0.095  | 0.905  | 0.015  | 0.099  |
| 10  | 0.123  | 0.877  | 0.016  | 0.131  |

This table is useful for predicting the chance or risk of developing disease using the function selected based on the desired time.

4. Conclusion
Conclusions that can be generated from this discussion are factors that affect a person at risk of stroke are the type of stroke, hypertension, systolic blood pressure, diastolic blood pressure, and diabetes mellitus. If we are taken gender as strata consisting of female and male, the significant covariates for stroke and diabetes sufferers are types of stroke, hypertension, systolic blood pressure. If it be taken sample of all women who have had a stroke, it is mainly affected by the type of stroke. If it be taken sample with all male members who have had a stroke. It can be said that the significant factors that cause stroke are type of stroke, hypertension, systolic blood pressure, and diabetes mellitus. Men have a higher rate of stroke than women, in general.

References
[1] Alodokter 2019 Diabetes Kementerian Kesehatan Republik Indonesia
[2] Ata N and Sozer T 2007 Hacettepe J. Math. Stat. 36 2 157-167
[3] Cox D R 1972 J. Royal Stat. Soc. 34 187-220
[4] Crawley M J 2007 *The R Book* (UK: John Wiley & Sons Ltd.)
[5] Fatekurohman M, Nurmla N and Anggraeni D 2017 *J. Phys: Conf. Ser.* **1008**
[6] Fisher M, Moores L, Alsharif M N and Paganini A 2016 *JAMA Neurology* **73** 2 186-189
[7] Fox J and Weisberg S 2018 *Cox Proportional Hazard Regression for Survival Data in R* Third Edition (CA: Thousand Oaks)
[8] Hanagal DD 2011 *Modeling Survival Data using Frailty Models* (New York: Taylor and Francis Group LLC)
[9] Hosmer D W, Lemeshow S and May S 2008 *Applied Survival Analysis: Regression Modelling of Time to Event Data* Second Edition (New Jersey: Jhon Wiley & Sons, Inc.)
[10] Kalbfleisch S D and Prentice R L 2002 *The Statistical Analysis of Failure Time Data* 2nd ed. (New York: John Wiley & Sons Inc.)
[11] Kleinbaum D G and Klein M 2012 *Survival Analysis A Self-Learning Text* Third Edition (London: Springer)
[12] Lee E T and Wang J W 2013 *Statistical Methods for Survival Data Analysis* Fourth Edition (Canada: JohnWiley & Sons Inc.)
[13] Sudarno and Prahutama A 2018 *J. Phys.: Conf. Ser.* **1217**