International Journal of Scientific Research in Physics and Applied Sciences
Vol.**12**, Issue.**4**, pp.**01-08**, August **2024**
E-ISSN: 2348-3423
Available online at: www.isroset.org

Research Article

Application of Weibull Regression to Recurrence Data Kidney Infection Patients

Andrea Tri Rian Dani[1*], Qonita Qurrota A'yun[2], Fachrian Bimantoro Putra[3], Meirinda Fauziyah[4]

[1,3,4]Statistics Study Program, Department of Mathematics, Mulawarman University, Samarinda, Indonesia
[2]Mathematics Study Program, Department of Mathematics, Mulawarman University, Samarinda, Indonesia

**Corresponding Author: andreatririandani@fmipa.unmul.ac.id*

Received: 07/Jun/2024; **Accepted**: 09/Jul/2024; **Published**: 31/Aug/2024

Abstract— The kidneys are the main organ for removing metabolic waste products that are not needed by the body. Kidney infection or pyelonephritis is an infection in the bladder tract that attacks the kidneys and enters through the lower or external urinary tract, spreads to the bladder, to the ureters (upper urinary tract), then finally to the kidneys. This disease is often not detected so it can cause complications of kidney failure. This research discusses survival analysis in kidney infection patients using Weibull survival regression by carrying out descriptive statistical analysis, estimating Weibull regression model parameters, testing hypothesis of regression model parameters simultaneously and partially. The aim of this study was to determine the Weibull survival regression model and determine the factors that influence the length of time for kidney infection patients to relapse. The results of this research obtained the best model is weibill regression , from which it can be seen that frailty, gender, and the patient's medical history are factors that significantly influence the recovery rate of kidney infection patients.

Keywords— Kidneys, Kidney Infection, Weibull Survival Regression

1. Introduction

The kidneys are the main organs for removing metabolic waste products that are not needed by the body. The main function of the kidneys is to get rid of normal metabolic waste, excrete xenobiotics and their metabolites and non-excretory functions. The kidneys are very vital organs of the human body because the kidneys are one of the urinary organs. Kidney disease can increase the risk of death for sufferers and can also be a trigger for the onset of other diseases. If kidney disease can be detected early, other diseases that cause death can be prevented immediately [1].

Kidney infection or pyelonephritis is an infection that occurs in the bladder tract and can propagate to the kidneys. Infection of the kidneys is associated with urinary tract infections because generally germs or bacteria that attack the kidneys enter through the lower or outer urinary tract, propagate to the bladder, to the ureters (upper urinary tract), then finally to the kidneys. Germs can also infect the kidneys through the bloodstream, but this is rare. Some of the causes of the higher rate of infection sensitivity in patients with kidney damage in patients due to decreased immunoglobulin levels, protein deficiency, phagocytosis disorders against bacteria and due to immunosuppressive treatment. The disease often goes undetected so it can lead to complications of kidney failure. For this reason, it is necessary to analyze which factors can predominantly affect kidney function [2].

In the health sector, statistical analysis that is often used in analyzing the rate of cure is survival analysis. Survival analysis is a statistical procedure for data analysis with the results of the variable being considered, namely time. In survival analysis, there are two main functions, namely the hazard function and the survival function. The hazard function is the failure rate of an event in question, which is when an individual experiences an event in the time interval t provided that he has survived until that time. The survival function is the probability of survival until a certain time which can be estimated using parametric methods [3]. Parametric methods can explain the individuals in the group, survival time, and the variables that affect them or can explain the variables that affect individuals in the group on their survival time.

One parametric method that is often used is Weibull Regression because the Weibull distribution has more flexible properties when compared to other models. This is because the Weibull distribution has shape β parameters that make the hazard curve more flexible. Weibull regression modeling emphasizes changes in survival and hazard functions after

being directly influenced by covariates. In the application of Weibull regression to time data generally describes the chances of survival and the determination of the rate at which an individual experiences an event which refers to death or incidence of disease [4].

2. Related Work

Some previous studies, one of which is research by Safitri, et al. (2013) [5] and Safitri &Helma (2016) on factors affecting kidney failure using Cox regression [6]. In addition, there are several studies comparing Weibull regression and Cox Proportional Hazard regression to conclude that the Weibull regression model is better than Cox Proportional Hazard regression because the AIC value is small [7] [8] [9]. Research and development on the use of Weibull regression has also been carried out by Hasan, et al [10], namely survival analysis using Weibull regression on the cure rate of pulmonary tuberculosis patients at Aloei Saboe Hospital Gorontalo City; Hasmiati, et al [11], namely Weibull regression model on hospitalization time data of patients with coronary heart disease with death events at Abdul Wahab Sjahranie Hospital Samarinda; Panduwinata, et al [12], namely Weibull regression model on classified continuous data; Solehah &; Fatekurohman [4] , namely survival analysis of lung cancer patients using Weibull regression; Suyitno [13], which is about parameter estimation and hypothesis testing of univariate Weibull regression models; and Suyitno, et al [14], namely about the multivariate Weibull regression model. Based on the description above, this study focuses on the latest application of data and the application of Weibull regression to find out what factors affect the length of time for relapse of kidney infection patients. Factors affecting hospitalization time data for kidney infection patients can be determined based on Weibull regression models.

3. Theory/Calculation

3.1 Weibull Distribution

The Weibull distribution is a distribution that plays an important role in survival data analysis [15]. This distribution is widely used in survival analysis modeling. The opportunity density function of the scale-form version of the Weibull distribution is [16]:

$$f(t)=\frac{\gamma}{\lambda}\left(\frac{t}{\lambda}\right)^{\gamma-1} exp\left(-\left(\frac{t}{\lambda}\right)^{\gamma}\right), \lambda>0, \gamma>0, t\geq 0 \quad (1)$$

The cumulative functions of the Weibull distribution are as follows:

$$F(t)=1-exp\left(-\left(\frac{t}{\lambda}\right)^{\gamma}\right) \quad (2)$$

The survival function of the scale-shape version of the Weibull distribution can be written by the following equation:

$$S(t)=1-F(t)=exp\left[-\left(\frac{t}{\lambda}\right)^{\gamma}\right] \quad (3)$$

The equation of the hazard function of the Weibull distribution is as follows:

$$h(t)=\frac{f(t)}{S(t)}=\frac{\gamma}{\lambda}\left(\frac{t}{\lambda}\right)^{\gamma-1}=\gamma\lambda^{-\gamma-1} \quad (4)$$

One method of estimating Weibull distribution parameters is Maximum likelihood Estimation (MLE), which is an estimation method by maximizing the likelihood function. The likelihood function and log-likelihood function of the weibull distribution are as follows [13]:

$$L(\boldsymbol{\theta_0}|\boldsymbol{t})=\prod_{i=1}^{n} f(\boldsymbol{\theta_0}|\boldsymbol{t_i}) = \prod_{i=1}^{n}\frac{\gamma}{\lambda}\left(\frac{t_i}{\lambda}\right)^{\gamma-1} exp\left(-\left(\frac{t_i}{\lambda}\right)^{\gamma-1}\right) \quad (5)$$

$$\ell(\boldsymbol{\theta_0}|\boldsymbol{t})=\sum_{i=1}^{n}\begin{bmatrix}\ln(\gamma)-\ln(\lambda)+(\gamma-1)\\ [ln\ t_i-\ln\lambda]-\left(\frac{t_i}{\lambda}\right)^{\gamma-1}\end{bmatrix} \quad (6)$$

3.2 Weibull Regression

The Weibull regression model is a regression model of the Weibull distribution with scale parameters expressed in the following equation:

$$\lambda = exp[\boldsymbol{\beta^T x}] \quad (7)$$

where $\boldsymbol{\beta^T}=[\beta_0 \quad \beta_1 \ \dots \ \beta_p]$ is a dimensionless regression parameter vector $p+1$ and $\boldsymbol{x}=[X_0 \quad X_1 \ \dots \ X_p]^T$ an independent variable vector with $X_0=1$ [17]. Then λ it is a scale parameter or natural parameter of the Weibull distribution, which can be expressed in terms of a function of regression parameters (covariates) [18].

Based on the above similarities, FKP Weibull becomes [13]:

$$f(t,\boldsymbol{\theta})=\gamma t^{\gamma-1}exp[-\gamma[\boldsymbol{\beta^T x}]]exp[-t^{\gamma}exp[-\gamma\boldsymbol{\beta^T x}]] \quad (8)$$

where $\boldsymbol{\theta}=[\gamma \quad \beta_0 \quad \beta_1 \dots \beta_p]^T$ is a dimensioned parameter vector $p+2$ with Weibull survival regression model and Weibull hazard regression model as follows:

$$S(t,\boldsymbol{\theta})=exp[-t^{\gamma}exp[-\gamma\boldsymbol{\beta^T x}]] \quad (9)$$

$$h(t,\boldsymbol{\theta})=\gamma t^{\gamma-1}exp[-\gamma\boldsymbol{\beta^T x}] \quad (10)$$

3.3 Weibull Regression Model Parameter Estimation

Distance estimation parameters of Weibull regression models can use the MLE method. Weibull regression parameter estimation consists of $S(t)$ survival regression model parameter estimation and Weibull hazard regression model. The likelihood function of Weibull regression based on the probability density function of Weibull regression is as follows: [13]

$$L(\boldsymbol{\theta})=\prod_{i=1}^{n}\frac{\left(\gamma t_i^{\gamma-1}exp[-\gamma\boldsymbol{\beta^t x_i}]\right)^{\delta_i}}{exp\left[-t_i^{\gamma}exp[-\gamma\boldsymbol{\beta^T x_i}]\right]} \quad (11)$$

with $\boldsymbol{x_i} = [x_0 \quad x_{i1} \quad \dots \quad x_{ip}]^T$. The ML estimator of the Weibull regression model is a vector value $\widehat{\boldsymbol{\theta}}$ that maximizes the likelihood function as well as maximizes the log-likelihood function. ML estimators are easily obtained through maximum log-likehood function. The log-likelihood function is as follows:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} [\delta_i [ln\gamma + (\gamma - 1) lnt_i - \gamma \boldsymbol{\beta^T x_i}] - t_i^{\gamma} exp[[-\gamma \boldsymbol{\beta^T x_i}]] \quad (12)$$

The exact solution of the Likelihood equation to obtain the exact estimator of Maximum Likelihood (ML) cannot be found analytically for several reasons, so an alternative method for obtaining the estimator $ML(\boldsymbol{\theta_0})$ is found, namely the Newton-Raphson iterative method. The Newton-Raphson method is one approach that is often used to estimate parameters, because the derivative of the equation is easy to calculate [19]. In this Newton-Raphson iterative method, it is necessary to calculate gradient vectors and Hessian matrices $\mathbf{H}(\boldsymbol{\theta})$ to determine ML estimators. The gradient vector is as follows:

$$\mathrm{g}(\boldsymbol{\theta}) = \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma} \quad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta^T}}\right]^T \quad (13)$$

The general form of the Hessian matrix is as follows:

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \beta_0} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \beta_p} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_0 \partial \gamma} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_0^2} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p \partial \beta_0} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p \partial \gamma} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p \partial \beta_0} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p^2} \end{bmatrix} \quad (14)$$

The Newton-Raphson algorithm is given by the following Equation:

$$\widehat{\boldsymbol{\theta}}^{(q+1)} = \widehat{\boldsymbol{\theta}}^{(q)} - \left[\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^{(q+1)}\right)\right]^{-1} \boldsymbol{g}\left(\widehat{\boldsymbol{\theta}}^{(q)}\right), q = 0,1,2,.. \quad (15)$$

the Newton-Raphson iteration process begins with determining the initial estimated value of the parameter $\widehat{\boldsymbol{\theta}}^{(0)} = \left[\hat{\gamma}^{(0)} \; \hat{\beta}_0^{(0)} \; \hat{\beta}_p^{(0)}\right]^T$ and terminated on the fourth iteration $\mathrm{q} + 1$ if convergent conditions are met, i.e. $\left\|\boldsymbol{\theta}^{(q+1)} - \widehat{\boldsymbol{\theta}}^{(q)}\right\| < \varepsilon$, where ε is a positive real number that is small enough for example 10^{-6} and $\left\|\boldsymbol{\theta}^{(q+1)} - \widehat{\boldsymbol{\theta}}^{(q)}\right\|$ expresses the norm or distance of two vectors.

3.4 Weibull Regression Model Parameter Significance Testing

Parameter significance testing of Weibull regression models consists of two, namely simultaneous and partial parameter signification testing. Testing the significance of parameters is simultaneously carried out to determine whether the estimated parameters affect the regression model together. This test is also used to confirm whether the estimated parameters provide a suitable regression mode (fit). The hypothesis of testing the significance of parameters simultaneously is: [20]

H_0 : $\beta_1 = \beta_2 = \ldots = \beta_p = 0$
(Weibull's regression model is not feasible)
H_1 : minimal ada satu $\beta_k \neq 0, k = 1,2,\ldots,p$
(Weibull's regression model is feasible)

The test statistics used are G which is determined by the likelihood ratio test method, namely:

$$\mathrm{G} = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_0)) \quad (16)$$

The critical area of this test will $\mathrm{H_0}$ be rejected at the level of significance α if the value $\mathrm{G} \geq \chi^2_{(\alpha,p)}$.
Partial significance testing is used to determine whether certain covariates individually have an effect on the regression model. The partial testing hypothesis for the specified c is $\beta_k, k = 0,1,2,\ldots,p$ as follows:
H_0 : $\beta_k = 0$
(Covariates X_k have no effect on regression models)
H_1 : $\beta_k \neq 0$
(Covariates X_k have no effect on regression models)
The test statistics used are the Wald test given by:

$$W_0 = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \sim N(0,1) \quad (17)$$

This critical area of testing will H_0 be rejected at the level of significance α if a value is obtained from $|W_0| > Z_{1-\alpha/2}$ [20].

3.5 Multicollinearity Detection

Multicollinearity is the condition that there is a linear relationship or strong correlation between covariates in a regression model. Multicollinearity can be detected by looking at the value of Variance Inflation Factor (VIF) which is formulated as follows:

$$\mathrm{VIF} = \frac{1}{1 - \mathrm{R_k^2}} \quad (18)$$

A VIF value greater than 10 indicates the presence of multicollinearity between covariates [21].

4. Experimental Method/Procedure/Design

This research was conducted with a research design, literature study, and empirical review. The type of research is non-experimental, namely research whose observations are carried out on a number of characteristics (variables) of research subjects according to the circumstances as they are, without any manipulation (intervention) of researchers. The data used is taken from the package provided by the R software, namely the survival package. The sampling technique of this study is purposive sampling. Research data analysis techniques consist of analysis of Weibull survival regression model and Weibull hazard model, because this study focuses more on the length of time it lasts until the recurrence of kidney infection disease and prioritizes obtaining factors that can

affect the length of time kidney infection patients relapse based on assumptions from the distribution used, in this case, the Weibull distribution. The stages of data analysis in the Weibull regression model are descriptive statistical analysis, estimation of Weibull regression model parameters, simultaneous and partial testing of regression model parameter hypotheses.

5. Results and Discussion

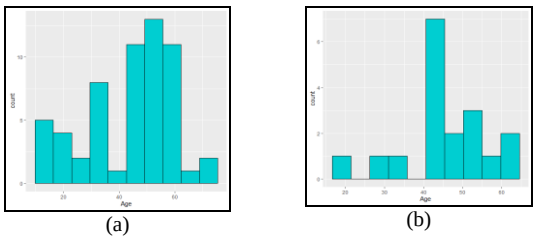
5.1 Data Exploration

In numerical type research data, data descriptions can be obtained as in the following table:

Table 1. Descriptive Statistics of Research Data

Data	Mean	Standard Deviation	Min	Max
Hospitalization Time (T)	101,6	130.91	1.00	562.00
Age (X_1)	43.70	14.74	10.00	69.00
Frail (X_2)	1.18	0.68	0.20	3.00

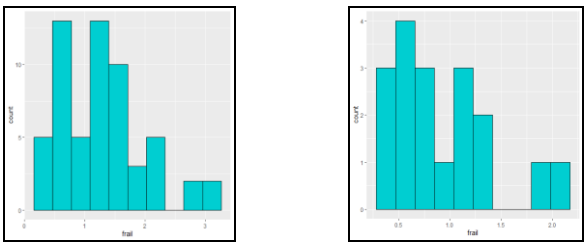
Furthermore, exploration using the histogram is shown in Figure 1.



(a) (b)

Figure 1. Descriptive Statistics of Age Variables
(a) Relapse (b) No Relapse

Based on the diagram in Figure 1, it can be seen that the highest recurrence rate of infection after catheter installation in kidney disease patients is around the age of 50 years.



(a) (b)

Figure 2. Descriptive Statistics of Frail Variables
(a) Relapse (b) No Relapse

Based on the diagram in Figure 2, it can be seen that the highest recurrence rate of infection after catheter insertion in kidney disease patients is in patients with estimated weakness of about 0.5 to 1.4.

In categorical type research data, data descriptions can be obtained as shown below:

(a) (b)

Figure 3. Descriptive Statistics
(a) Gender; (b) Types of disease

Based on Figure 3(a), the number of male patients is 20 people or 26% while female patients are 56 people or 74%. In Figure 3.3(b), the number of patients who have GN disease is 18 people or 24%, then patients who have AN disease there are 24 people or 32%, then patients who have PKD disease there are 8 people or 10% and patients who have other diseases (other) there are 26 people or 34%.
Checking the survival function of two visually different categories (treatment) using the Kaplan-Meier method on each variable of the categorical type.

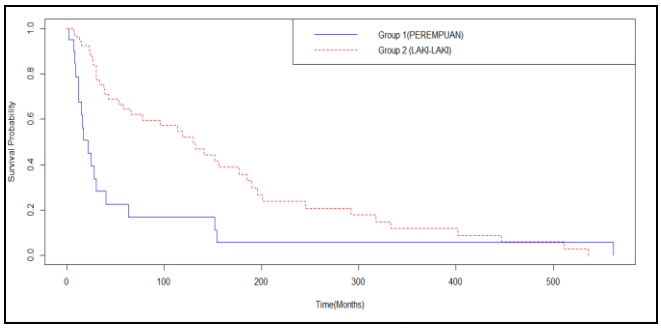


Figure 4. Kaplan-Meier curve of gender variables

Based on Figure 4, a graph of survival function using Kaplan-Meier, it is known that female patients provide greater results than male patients. It can be concluded that the chances of survival in female patients are greater than those of male patients.

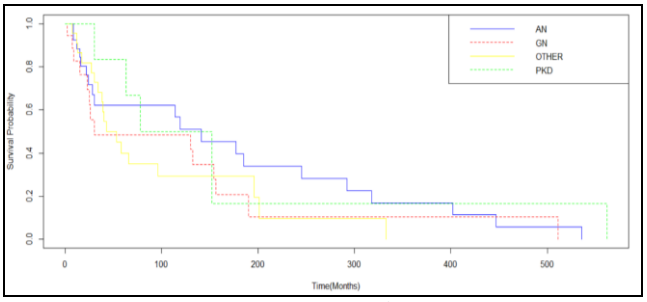


Figure 5. Kaplan-Meier curve of disease variables

Based on Figure 5, survival function graph using Kaplan-Meier, it is known if patients have types of AN, GN, PKD, and other diseases give the same results (lines close together), so it can be concluded if the chances of survival in patients with AN, GN, PKD, and other diseases do not have different chances of recurrence.

To be more accurate, hypothesis testing was carried out using the Log-Rank test and obtained results as in the following table:

Table 2. Log Rank Testing

Variable	P-value	α	Results
Gender (X_3)	0.00	0.05	Reject H_0
Disease (X_4)	0.40	0.05	Failed to Reject H_0

Based on Table 2, it is known that for the sex variable it is known that $p-value = 0{,}004 < \alpha = 0{,}05$ which means H_0 is rejected so that it can be concluded that there are differences in survival function between female patients and male patients, this is in accordance with the Kaplan-Meier graph in Figure 4. As for the disease variable, it can be known that the value $p-value = 0{,}4 > \alpha = 0{,}05$ that means H_0 failed to be rejected so that it can be concluded that there is no difference in survival function between patients with GN, AN, PKD, other (other) this is in accordance with Kaplan Meier's graph in Figure 5.

5.2 Weibull Distribution Parameters

Estimation of Weibull distribution parameters was carried out on the analysis of recurrence time data of patients with kidney infection. The parameter estimation method is MLE which is solved by the Newton-Raphson iterative method.This is because the derivative of the equation is easy to calculate [18]. The Weibull distribution is characterized by three parameters, namely the location parameter (μ), scale parameters (λ), and shape parameters (γ). The location parameter (μ) is set as the minimum value in the distribution, where in survival analysis the location parameter (μ) is zero [23]. The results of parameter estimation using R software can be seen in the following table:

Table 3. Weibull Distribution Parameter Assessment

Parameters	Estimates
Scale (λ)	0.80579494
Shape (γ)	89.12814246

Based on Table 3, the survival function estimator is obtained, namely:

$$\hat{S}(t) = P(T > t) = \exp\left[0{,}80579494 \times t^{89{,}12814246}\right] \quad (20)$$

and the estimator of the theoretical cumulative distribution function is:

$$\hat{F}(t) = P(T \leq t) = 1 - \exp\left[0{,}80579494 \times t^{89{,}12814246}\right] \quad (21)$$

5.3 Inpatient Time Data Distribution Testing

Testing the distribution of recurrence time data using the Anderson-Darling approach to determine whether the recurrence data of patients with kidney disease follow the Weibull distribution, with a cumulative distribution function. The calculation results obtained using R software can be seen in the following table.

Table 4. Weibull Distribution Test Results

P-value	α	Results
0.3421	0.05	Failed to Reject H_0

Based on Table 4, which means H_0 failed to be rejected so it can be concluded that the data on the length of time for relapse are Weibull distributed with a shape of magnitude 0.80579494 and scale of 89.1281424

5.4 Multicoliniearity Detection

One method of detecting multicollinearity is to use VIF values. A VIF value of > 10 indicates multicholinerity between independent variables.

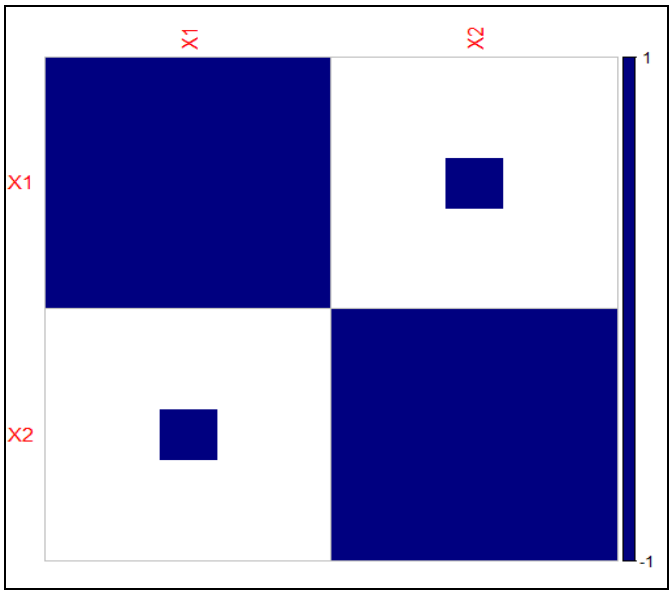


Figure 6. Heatmap Correlation

Based on Figure 6, visually it can be seen that there is no strong correlation between the Age and Frail variables, where based on the calculation of the VIF value of each independent variable using R software can be seen in the following table.

Table 5. VIF value

Variable	VIF value
Age (X_1)	1.062353
Frail (X_2)	1.106196

Based on Table 5, it can be seen that the VIF values for both variables are below 10 which indicates that there is no correlation between predictor variables, this is in accordance with the Heatmap Correlation in Figure 6.

5.5 Weibull Survival Regression Modeling

The model formed using independent variables, namely Age, Frail, Gender, and Disease is as follows:

$$\hat{Y} = \exp\left(-y^{\gamma} \exp\begin{pmatrix}\gamma(-\beta_0 - \beta_1 X_1 - \beta_2 X_2 - \\ \beta_3 D_1 - \beta_4 D_2 - \beta_5 D_3 - \beta_6 D_4\end{pmatrix}\right) \quad (22)$$

with:

y : Time Variable
γ : Weibull distribution form parameters
X_1 : Age Variable
X_2 : Frail Variable
D_1 : Gender Dummy variable (1 if female, 0 if male)

D_2 : Dummy variable GN disease (1 if GN, 0 others)
D_3 : Dummy variable AN disease (1 if AN, 0 others)
D_4 : Dummy variable PKD disease (1 if PKD, 0 others)
β_0 : Intercept
β_1 : Slope Variable Age
β_2 : Slope Variable Frail
β_3 : Slope Variable Dummy Gender
β_4 : Slope Variable Dummy GN disease
β_5 : Slope Variable Dummy Disease AN
β_6 : Slope Variable Dummy Disease PKD

Next, the parameter estimation of the Weibull survival regression model. Based on the calculation results through R software, the Weibull regression model parameter estimator is obtained in Table 6 below:

Table 6. Weibull Regression Model Parameter Estimator

Parameters	Estimates
γ	1.636661
$\hat{\beta}_0$	4.96747
$\hat{\beta}_1$	-0.00323
$\hat{\beta}_2$	-1.05518
$\hat{\beta}_3$	1.48237
$\hat{\beta}_4$	0.01159
$\hat{\beta}_5$	-0.59542
$\hat{\beta}_6$	1.22265

Based on Table 6, Weibull's survival regression model is obtained, namely:

$$\hat{Y} = exp - y^{1,636661}\exp\left(1{,}636661(-4{,}96747 - (-0{,}00323)X_1 - (-1{,}05518)X_2 - 9{,}48237D_1 - 0{,}011159D_2 - (-0{,}59542)D_3 - 1{,}22265D_4)\right) \quad (23)$$

$$\hat{Y} = exp - y^{1,636661}\exp\left(-8{,}13006 + 0{,}0053X_1 + 1{,}727X_2 - 2{,}42614D_1 - 0{,}01897D_2 + 0{,}9745D_3 - 2{,}00106D_4\right) \quad (24)$$

5.6 Weibull Regression Best Model Selection

The criteria in choosing the best model can refer to the smallest AIC (Akaike Information Criterion) value and the smallest BIC (Bayesian Information Criterion), but in this study in the selection of the best model Weibull regression uses the AIC value, because the value produced by AIC is smaller than BIC [24]. The best Weibull regression model is selected based on the 4 independent variables used.

Table 7. Best Model of Weibull Regression

Model	AIC value
$Y \sim X_2 + X_3 + X_4$	622.1
$Y \sim X_1 + X_2 + X_3 + X_4$	623.9

Based on Table 7, the smallest AIC value is 622.1, which is the Weibull regression model which contains 3 independent variables, namely, Frail (X_2), Gender (X_3), and Disease (X_4).

5.7 Best Weibull Regression Model Hypothesis Testing

Parameter hypothesis testing Weibull Regression Model consists of simulating parameter hypothesis testing and partial parameter hypothesis testing.
Simultaneous parameter hypothesis testing aims to find out whether the estimated parameters provide a variable regression model (fit). Simultaneous testing also aims to determine whether the variables frail, sex, and disease together affect the Weibull regression model. Here's the hypothesis testing:

H_0 : $\beta_1 = \beta_2 = \beta_3 = 0$
(Weibull Regression Model is not feasible (not fit))
H_1 : There is at least one $\beta_k \neq 0; k = 1,2,3$
(Weibull Regression Model fit (fit))

The results of simultaneous testing using R software can be seen in the following table:

Table 8. Simultaneous Test Results

P-value	Results
1.7×10^{-14}	H_0 rejected

Based on Table 8, it can be seen that means H_0 is rejected so that it can be concluded that frail, sex, and disease simultaneously affect the patient's recurrence time data. Partial testing of parameter hypotheses is performed to determine whether certain independent variables individually have an effect on Weibull's regression model. The test statistics used are the Wald test with W~N(0,1), the following results are obtained:

Table 9. Partial Test Results

Variable	P-value	Result
Frail (X_2)	$<2 \times 10^{-16}$	H_0 rejected
Gender (X_3)	4.4×10^{-15}	H_0 rejected
Disease GN (X_4)	0.8167	H_0 failed to be rejected
Disease AN (X_4)	0.0009	H_0 rejected
Disease PKD (X_4)	8.5×10^{-15}	H_0 rejected

Based on Table 3.10, it is known that frail (X_2), sex (X_3), AN disease and PKD disease affect the recurrence time data. While GN disease did not have a partial effect on the data model of patient recurrence time.

Based on the results of the best Weibull regression modeling and parameter testing, it can be said that the Weibull regression model is proven to be able to obtain factors that affect the length of recurrence time of patients with kidney infections. This is also reinforced by the results of research from Cavalcante, T., et al [25], who explain that Weibull regression is more suitable for research that prioritizes the use of assumptions from a distribution that underlies the length of patient survival, in this case the distribution in question is the Weibull distribution.

6. Conclusion and Future Scope

Based on the results of the analysis and discussion, it can be concluded that the best weibull survival regression model formed is as follows:

$$\hat{Y} = \exp\left(-y^{1,631321}\exp\left(-7,95905 + 1,7233X_2 - 2,41077D_1 + 0,08418D_2 + 1,08891D_3 - 1,86721D_4\right)\right) \quad (27)$$

where:

y = Time Variable
X_2 = Frail Variables
D_1 = Gender Dummy *variable* (worth 1 if female, 0 if male)
D_2 = *Dummy* variable GN disease (worth 1 if GN, 0 if not)
D_3 = *Dummy* variable disease AN (value 1 if AN, 0 if not)
D_4 = *Dummy* variable PKD disease (worth 1 if PKD, 0 if not)

Factors that affect the length of time for relapse of kidney infection patients are *Frail*, Gender, and the patient's disease history.

Data Availability

The data used for this research as a secondary data. https://search.rproject.org/CRAN/refmans/SurvCorr/html/kidney.html

Conflict of Interest

All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations that could inappropriately influence, or be perceived to influence, their work. Otherwise, Authors declare that they do not have any conflict of interest.

Funding Source

This research was fully funded by the authors personally.

Authors' Contributions

A.T.R. Dani conducted initial identification, selection of case studies, and data analysis. In addition, this article serves as chairman, Q. Q. A'yun, F. B. Putra, dan L. Ni'matuzzahroh as research assistants who assited in the data analysis process, writing articles according to templates. At the end, we are re-examined the manuscripts to be submitted.

Acknowledgements

The authors gratefully acknowledge to the Mulawarman University for facilitating our team in conducting data analysis

References

[1] A. P. D. Alamsyah and Normalisa, "Sistem Pakar Diagnosa Penyakit Ginjal," International Journal of Artificial Inteligence, vol. 6, no. 1, pp. 53-74, 2019.

[2] L. Geografi and N. M. Y. D. Wahyani, "Duration of Antibiotic Treatment for Urinary Tract Infection," Jurnal Manajemen dan Pelayanan Farmasi, vol. 4, no. 1, 2014.

[3] P. D. Allison, Survival Analysis Using the SAS System: A Practical Guide, Cary , NC: SAS Institute, Inc., 1995.

[4] A. Solehah and M. Fatekurohman, "Analisis Ketahanan Hidup Pasien Kanker Paru Menggunakan Regresi Weibull," Indonesian Journal of Applied Statistics, vol. 1, no. 2, pp. 79-87, 2018.

[5] N. Safitri, Syafriandi and Helma, "faktor Dominan yang Mempengaruhi Penyakit Gagal Ginjal dengan Menggunakan Regresi Cox (Studi Kasus di RZUP Dr. M. Djamil Padang)," Journal of Mathematics UNP, vol. 1, no. 1, pp. 47-53, 2013.

[6] N. Safitri and Helma, "Faktor-Faktor Penyebab Penyakit Gagal Ginjal pada Pasien di RSUP Dr. M. Djamil Padang dengan Menggunakan Regresi Cox," EKSAKTA, pp. 60-66, 2016.

[7] Y. I. Z. Afni, M. Fatekurhoman and D. Anggraeni, "Perbandingan Model Cox Proportional Hazard dan Regresi Weibull untuk Menganalisis Ketahanan bank Syariah," Indonesian Journal of Applied Statistics, vol. 6, no. 1, pp. 53-74, 2019.

[8] Y. S. A. Monica and Purhadi, "Analisis Faktor yang Memengaruhi Laju Kesembuhan Pasien Tuberkulosis Paru di RSUD dr. Soetomo Tahun 2015 Menggunakan Regresi Weibull dan Regresi Cox Proportional Hazard," JURNAL SAINS & SENI ITS, vol. 5, no. 2, 2016.

[9] Mayawi, Nurhayati, N. S. Laamen, A. W. Bustan, M. Salmin and T. Talib, "Perbandingan Model Regresi Weibull dan Regresi Cox Proportional Hazard," Science Map Journal, vol. 4, no. 2, 2022.

[10] I. K. Hasan, W. A. Pakarya, N. Achamd and D. R. Isa, "Analisis Survival Menggunakan Regresi Weibull pada Laju Kesembuhan Pasien Tuberkulosis Paru di RSUD Aloei Saboe Kota Gorontalo," EULER: Jurnal Ilmiah Matematika, Sains, dan teknologi, vol. 9, pp. 40-51, 2021.

[11] Hasmiati, Suyitno and Nasution, "Model Regresi Weibull pada Data Waktu Rawat Inap Pasien Penderita Penyakit Jantung Koroner dengan Event Kematian di RSUD Abdul Wahab Sjahranie Samarinda," Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya, vol. 2, pp. 369-376, 2022.

[12] H. D. Panduwinata, Suyitno and M. N. Huda, "Model Regresi Weibull pada Daa Kontinu yang Diklasifikasikan (Studi Kasus: Indikator Pencemaran Air BOD di DAS Mahakam Tahun 2016)," EKSPONENSIAL, vol. 13, no. 2, pp. 123-130, 2022.

[13] Suyitno, "Penaksiran Parameter dan Pengujian Hipotesis Model Regresi Weibull Univariat," Jurnal EKSPONENSIAL, vol. 8, pp. 179-184, 2017.

[14] Suyitno, Purhadi, Sutikno and Irhamah, "Multivariate Weibull Regression Model," Far East Journal of Mathematical Sciences, vol. 101, pp. 1977-1992, 2017.

[15] S. F. Khairunnisa, Suyitno and S. Mahmuda, "Weibull Regression Model on Hospitalization Time Data of COVID-19 Patients at Abdul Wahab Sjahranie Hospital Samarinda," Jurnal Matematika, Statistika, dan Komputasi, vol. 19, no. 2, pp. 286-303, 2023.

[16] J. F. Lawless, Statistical Models and Methods for Lifetime Data, New York: John Willey and Sons, Inc, 1982.

[17] D. D. Hanagal, "Parametric Bivariate Regression Analysis Based on Censored Samples: A Weibull Model," Stochastics and Quality Control, vol. 19, no. 1, pp. 83-90, 2004.

[18] H. Rinnie, The Weibull Distribution A Handbook, New York: CRC Press Taylor and Francis Group, 2009.

[19] H. Rehman, N. Chandra, T. Emura and M. Pandey, "Estimation of The Modified Weibull Additive Hazard Regression Model under Competing Risks," Symmetry, vol. 15, p. 485, 2023.

[20] Y. Pawitan, In All Likelihood Statistical Modelling and Inference using Likelihood, Sweden: Claredon Press-Oxford, 2001.

[21] N. Draper and H. Smith, Analisis Regresi Terapan Edisi Kedua, Jakarta: PT. Gramedia Pustaka Utama, 1992.

[22] A. C. Renchar, Linier Model in Statistics, New York: John Wiley and Sons, Inc, 2000.

[23] E. Liu, R. Y. Liu and K. Lim, "Using the Weibull Accelerated Failure Time Regression Model to Predict Time to Health Events," Applied Sciences, vol. 13, no. 24, 2023.

[24] P. Gauthami, K. D. Mariyamma and P. Kavya, "A Modified

Inverse Weibull Distribution Using KM Transformation," Reliability: Theory & Applications, vol. 18, pp. 35-42, 2023.
[25] T. Cavalcante, R. Ospina, V. Leiva, X. Cabezas and C. M. Barreiro, "Weibull Regression and Machine Learning Survival Models: Methodology, Comparison, adn Application to Biomedical Data Related to cardiac Surgery," Biology, vol. 12, no. 3, p. 442, 2023

AUTHORS PROFILE

Andrea Tri Rian Dani is a Lecturer at the Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia. He received M. Stat. Degree at Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
His research interest is Statistical Modeling with the Nonparametric-Semiparametric Regression approach and Its Applications. The results of his research are published in several National Journals, International Proceedings, and International Journals Indexed by Scopus.

Qonita Qurrota A'yun earned her Bachelor of Science and Master of Science in Mathematics from Gadjah Mada University, Yogyakarta, Indonesia. She is currently working as a Lecturer in Mathematics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University since 2021. She is an active member of the Indonesian Mathematical Society (IndoMS). Her research interest is related to linear algebra.

Fachrian Bimantoro Putra is a graduate student in the Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences Mulawarman University, Samarinda, Indonesia. He is currently a member of the research team with Andrea in the Statistical Modeling.

Meirinda Fauziyah is a Lecturer at the Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia. She received M. Stat. Degree at Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Her research interest is Statistical Modeling with the Spatial and Applications. The results of his research are published in several National Journals, International Proceedings, and International Journals Indexed by Scopus.

