

Model Regresi Cox *Proportional Hazard* dengan Penaksiran Parameter *Efron Partial Likelihood*

St. Husnah Nur^{1*}, Hikmah², Rahmawati³

^{1,3} Program Studi Matematika, Universitas Sulawesi Barat, Majene, Indonesia

² Program Studi Statistika, Universitas Sulawesi Barat, Majene, Indonesia

Corresponding Email*: husnaunnha14@gmail.com

Abstrak

Penyakit Tuberkulosis (TBC) masih menjadi salah satu penyakit menular yang paling mematikan di dunia dan Indonesia termasuk delapan negara yang menyumbang 2/3 kasus TBC di seluruh dunia. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang signifikan mempengaruhi lama perawatan pasien penderita tuberkulosis, dan mendapatkan model regresi Cox *Proportional Hazard* dengan Penaksiran Parameter *Efron Partial Likelihood*. Metode penelitian ini menggunakan metode *Efron Partial Likelihood* pada data lama perawatan pasien penderita tuberkulosis di Puskesmas Banggae I, Majene tahun 2017-2022. Penelitian ini memiliki 5 variabel bebas yaitu: jenis kelamin, umur, riwayat penyakit, pemeriksaan dahak dan alamat domisili. Penaksiran parameter pada model terbaik dinyatakan oleh nilai AIC yang paling kecil yaitu 1409,283. Berdasarkan uji parsial pemeriksaan dahak berpengaruh secara signifikan pada data lama perawatan pasien penderita tuberkulosis. Hal ini menunjukkan bahwa setiap pasien yang terdeteksi adanya bakteri disaluran pernafasan yang dikategorikan positif akan mengalami peningkatan 1,474 kali terhadap lama perawatan pasien penderita tuberkulosis dari pada kategori negatif.

Kata Kunci: *cox ph, efron partial likelihood, analisis survival, ties, tuberkulosis*

Abstract

Tuberculosis (TB) is still one of the most deadly infectious diseases in the world and Indonesia is one of eight countries that contributes 2/3 of TB cases worldwide. This study aims to identify factors that significantly affect the length of stay of patient with tuberculosis and obtain a Cox Proportional Hazard regression model with Efron Partial Likelihood Parameter Estimation. This research method uses the Efron Partial Likelihood method on data on the length of stay of tuberculosis patients at the Banggae I Health Center, Majene in 2017-2022. This study has 5 independent variables: gender, age, medical history, sputum examination and domicile address. Parameter estimation in the best model is expressed by the smallest AIC value 1409,283. Based on the partial test, sputum examination has a significant effect on data the length of treatment for patients suffering from tuberculosis. This shows that every patient who is diagnosed with bacteria in the respiratory tract who is categorized as positive will experience a 1.474 times increase in the length of treatment for patients suffering from tuberculosis compared to those in the negative category.

Keywords: *cox ph, efron partial likelihood, survival analysis, ties, tuberculosis*

Received :20-06-2024, Revised :23-09-2024, Accepted :24-09-2024

1. Pendahuluan

Penyakit Tuberkulosis (TBC) adalah salah satu penyakit menular yang paling mematikan di dunia. Indonesia merupakan salah dari delapan negara yang menyumbang 2/3 kasus TBC di seluruh dunia, yaitu posisi ke dua setelah india dengan jumlah kasus dan kematian yang tinggi (Indramayu). Kasus TBC di Indonesia semakin meningkat setiap tahunnya. Peningkatan kasus tersebut cukup signifikan terjadi pada tahun 2022. Kelompok pekerja ditemukan mendominasi pasien TBC, kebanyakan dari mereka yang terpapar adalah pekerja atau buruh pabrik, petani dan nelayan, pegawai BUMN, hingga pegawai negeri sipil (PNS) [1].

Dari latar fenomena diatas, diperlukan suatu analisis statistika mengenai masalah tersebut. Salah satu metode statistika yang sering digunakan dalam menganalisis ketahanan hidup pasien adalah analisis *survival*. Analisis *survival* (*survival analysis*) atau yang biasa dikenal sebagai analisis ketahanan hidup

adalah kumpulan prosedur statistika untuk menganalisis data dari waktu yang dihasilkan sampai terjadinya suatu *event* yang diteliti [2].

Adapun yang membedakan analisis *survival* dengan prosedur analisis statistika lainnya, yaitu adanya data sensor. Data tersensor adalah data yang diperoleh tidak dapat diamati secara utuh atau tidak lengkap, dikarenakan subjek penelitian tersebut keluar dari pengamatan pada saat belum mengalami *event* atau pada akhir pengamatan individu tersebut belum mengalami *event* tertentu yang dikutip oleh [1].

Dalam analisis *survival* sering kali ditemukan adanya *event* bersama atau yang lebih sering disebut *ties*. *Ties* adalah keadaan dimana terdapat dua individu atau lebih yang mengalami *event* pada waktu yang bersamaan yang dikutip oleh [3]. Jika suatu data terdapat *ties*, maka akan menimbulkan permasalahan dalam membentuk *partial likelihood* yaitu saat menentukan anggota dari himpunan risikonya. Untuk mengatasi adanya *ties* dalam analisis *survival* terdapat tiga metode yaitu metode *Exact*, *Efron* dan *Breslow* yang dikutip oleh [8].

Penelitian sebelumnya yang terkait dengan model regresi Cox PH dilakukan oleh [5] dengan studi kasus lama mencari kerja, dengan hasil bahwa menggunakan metode *Exact Partial Likelihood* lebih baik. Sedangkan penelitian lain dilakukan oleh [2] yang membahas analisis *survival* data kejadian bersama dengan metode *Efron Partial Likelihood* yang diterapkan pada lama masa studi mahasiswa FMIPA, Universitas Mulawarman Angkatan 2011 dengan hasil penelitian bahwa variabel yang berpengaruh signifikan yaitu variabel jenis kelamin dan IPK mahasiswa.

Tujuan penelitian ini yaitu mengidentifikasi faktor-faktor yang secara signifikan mempengaruhi lama perawatan pasien penderita tuberkulosis di Puskesmas Banggae I, Majene. Selain itu, tujuan lainnya yaitu mendapatkan model regresi cox *proportional hazard* dengan Penaksiran Parameter *Efron Partial Likelihood* berdasarkan data lama perawatan pasien penderita tuberkulosis di Puskesmas Banggae I, Majene.

2. Landasan Teori

2.1 Analisis Survival

Analisis *survival* (*survival analysis*) atau yang biasa dikenal sebagai analisis ketahanan hidup adalah kumpulan prosedur statistika untuk menganalisis data dari waktu yang dihasilkan sampai terjadinya suatu *event* yang diteliti [2]. Mulai dari awal pengamatan hingga *event* yang diharapkan terjadi disebut dengan waktu *survival* atau *failure time*. Waktu *survival* dapat dinyatakan dalam tahun, bulan, mingguan, hari, jam, menit, detik dan lainnya. Sementara kejadian yang dimaksud disini adalah kejadian yang diprediksi berupa kematian, kambuh, sembuh, waktu kembali bekerja, kerusakan dari suatu alat atau benda tertentu, maupun kejadian lainnya.

2.2 Fungsi-fungsi Waktu Survival

Menurut Kleinbaum dan Klein yang dikutip oleh [6] waktu *survival* dinyatakan sebagai peubah respon yang merupakan peubah acak non-negatif dan dilambangkan dengan T . Fungsi-fungsi yang saling berhubungan dalam analisis *survival* adalah fungsi kepadatan peluang, fungsi *survival*, dan fungsi *hazard*.

2.2.1 Fungsi Kepadatan Peluang

Fungsi kepadatan peluang (FKP) adalah peluang suatu individu mengalami *event* dalam interval waktu dari t sampai $t + \Delta t$, dengan waktu T merupakan variabel random. Fungsi kepadatan peluang dinotasikan $f(t)$ dirumuskan dengan :

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t \leq T < t + \Delta t)}{\Delta t} \right] \quad (1)$$

2.2.2 Fungsi Survival

Fungsi *survival* didefinisikan sebagai peluang suatu individu *survive* (belum mengalami *event*) lebih lama dari pada waktu t (Lee dan Wang) yang dikutip oleh [7]. Fungsi *survival* dinotasikan $S(t)$ dirumuskan sebagai berikut :

$$\begin{aligned} S(t) &= P(T > t) = 1 - (P(T \leq t)) \\ &= 1 - F(t) \\ F(t) &= 1 - S(t) \end{aligned} \quad (2)$$

2.2.3 Fungsi Hazard

Fungsi *hazard* adalah peluang suatu individu mengalami *event* dalam waktu interval t sampai $t + \Delta t$, jika diketahui individu tersebut masih dapat bertahan hidup hingga waktu t , pendapat dari Lawless yang dikutip oleh [7]. Fungsi *hazard* dinotasikan dengan $h(t)$ yang dinyatakan sebagai berikut :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < (t + \Delta t) | T \geq t)}{\Delta t} \quad (3)$$

2.3 Data Tersensor

Perbedaan analisis *survival* dengan prosedur analisis statistika lainnya, yaitu adanya data sensor. Menurut Lee dan Wang yang dikutip oleh [1]. Data dikatakan tersensor apabila data tidak dapat diamati secara lengkap karena subjek penelitian hilang atau mengundurkan diri atau sampai akhir penelitian subjek tersebut belum mengalami kejadian tertentu, sedangkan data yang diamati secara lengkap sampai penelitian berakhir disebut data yang tidak tersensor. Penyebab terjadinya data tersensor antara lain :

1. *Loss to follow up*, terjadi bila objek yang diteliti hilang dalam pengamatan.
2. *Drop out*, terjadi bila perlakuan dihentikan karena alasan tertentu.
3. *Termination of study*, terjadi bila masa penelitian berakhir sementara objek yang diobservasi belum mencapai *event*.

2.4 Kejadian Bersama (Ties)

Terdapat tiga metode yang biasa digunakan untuk mengatasi kejadian bersama dalam analisis *survival* yaitu metode *Exact*, metode *Breslow*, dan metode *Efron*. Metode *Exact* merupakan metode yang paling tepat digunakan karena merupakan metode yang paling akurat, tetapi metode ini memiliki kelemahan dalam hal efisiensi komputasi dan semakin besar jumlah *ties* yang ditemukan maka semakin besar waktu komputasi yang dibutuhkan. Metode *Breslow* dan metode *Efron* keduanya memiliki perhitungan yang sederhana dan cepat, tetapi metode *Efron* lebih akurat dibandingkan metode *Breslow*, terutama ketika data mengandung banyak *ties* yang dinyatakan Allison dikutip oleh [2].

2.5 Model Cox

Model cox adalah model semiparametrik yang artinya data *survival time* tidak diketahui distribusi dari data sehingga tidak diketahui bentuk fungsional dari fungsi *baseline hazard*, akan tetapi parameter-parameter (β) dari model dapat diketahui distribusinya [9].

2.5.1 Model Regresi Cox Proportional Hazard (PH)

Model regresi cox *proportional hazard* adalah salah satu cara yang digunakan untuk mengetahui hubungan antara waktu *survival* dengan variabel-variabel yang diduga mempengaruhi waktu *survival*. Model tersebut memiliki asumsi bahwa fungsi *hazard* dari individu yang berbeda adalah *proportional* atau *ratio*, fungsi *hazard* dari dua individu yang berbeda adalah konstan, Lee dan Wang yang dikutip oleh [6]. Regresi ini lebih populer digunakan dalam penelitian tentang data kesehatan, data ekonomi yang variabel responnya berupa waktu *survival* (tahun, bulan, hari dan sebagainya). Secara umum bentuk dari model regresi cox PH adalah :

$$h(t, x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \tag{4}$$

Menurut Darwanto [10] bentuk dari model cox PH pada persamaan (4) memiliki sifat bahwa jika semua $x = 0$, maka rumus tereduksi menjadi fungsi *baseline hazard*. Berdasarkan persamaan (4), $h_0(t)$ dianggap sebagai fungsi awal atau dasar dari fungsi *hazard* dan dapat dituliskan sebagai berikut :

$$\begin{aligned} h(t, x) &= h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \\ h(t, x) &= h_0(t) \exp(\beta_1 \times 0 + \beta_2 \times 0 + \dots + \beta_p \times 0) \\ &= h_0(t) \exp(0) \\ &= h_0(t)(1) \\ &= h_0(t) \end{aligned} \tag{5}$$

dimana, $h(t, x)$: Risiko kematian individu pada waktu (t) dengan karakteristik x .

$h_0(t)$: Fungsi *hazard* dasar / fungsi *baseline hazard*.

$\beta_1, \beta_2, \dots, \beta_p$: Parameter dari model regresi.

x_1, x_2, \dots, x_p : Variabel bebasnya.

Model regresi cox sangat populer digunakan karena:

1. Dapat mengestimasi *hazard ratio* tanpa perlu diketahui $h_0(t)$ atau fungsi *baseline*.
2. Dapat mengestimasi $h_0(t), h_1(t)$ dan fungsi *survival* meskipun $h_0(t)$ tidak spesifik.
3. Cox model *robust* sehingga dari model Cox hampir sama dengan hasil model parametrik.

2.5.2 Pengujian Asumsi Proportional Hazard

Menurut Kleinbaum dan Klein yang dikutip oleh [2] ada beberapa pendekatan yang dapat dilakukan untuk pemeriksaan asumsi *proportional hazard* yaitu dengan metode grafik $\ln[-\ln S(t)]$ atau uji *goodness of fit* (GOF) dengan residual *Schoefeld*.

1. Metode Grafik $\ln[-\ln S(t)]$

Dalam menggunakan grafik $\ln[-\ln S(t)]$, data dikelompokkan sesuai dengan kategori pada masing-masing variabel bebas. Asumsi *proportional hazard* dapat dikatakan terpenuhi, apabila grafik $\ln[-\ln S(t)]$ menunjukkan kurva yang sejajar. Kurva dikatakan sejajar apabila kedua garis tersebut terletak pada satu bidang diantara yang tidak akan berpotongan meskipun diperpanjang tanpa batas. Kelemahan menggunakan metode grafik $\ln[-\ln S(t)]$ adalah bersifat subjektif karena sejajar atau tidaknya kurva sangat bergantung pada cara peneliti menilai.

2. Uji Goodness Of Fit (GOF)

Pengujian asumsi dengan *goodness of fit* merupakan variasi dari tes *Schoenfeld*, yaitu menggunakan residual *Schoenfeld*. Residual *Schoenfeld* didefinisikan hanya pada waktu *survival* yang tidak tersensor. Residual *Schoenfeld* untuk individu yang mengalami *event* pada saat t_1 pada variabel bebas ke $-j$ adalah sebagai berikut :

$$R_{ij} = \delta_i \left(x_{ij} \frac{\sum_{q \in R(t_i)} x_{qj} \exp(\hat{\beta} x_q)}{\sum_{q \in R(t_i)} \exp(\hat{\beta} x_q)} \right), \text{ dimana } j = 1, 2, \dots, p \text{ dan } i = 1, 2, \dots, n \tag{6}$$

Adapun langkah-langkah pengujian asumsi *proportional hazard* menggunakan residual *Schoenfeld* adalah sebagai berikut :

1. Membangun model *cox proportional hazard* dan mencari taksiran residual *Schoenfeld* untuk setiap variabel prediktor.
2. Mengurutkan waktu *survival* mulai dari individu yang mengalami *event* pertama kali.
3. Menguji korelasi antara residual *Schoenfeld* dengan peringkat waktu *survival*.

2.6 Estimasi Parameter dengan Metode Efron Partial Likelihood

Metode *Efron* dianggap sebagai metode yang lebih baik dari pada metode *Breslow*. Walaupun metode *Breslow* merupakan metode yang lebih sering digunakan. Semakin besar jumlah *ties* pada data, maka hasil estimasi parameter semakin tidak akurat karena fungsi *partial likelihood* menjadi semakin berbeda dari fungsi *partial likelihood* yang sebenarnya.

Menurut [11] pada metode *Efron* ini, himpunan resikonya diselesaikan dengan pengurangan terhadap rata-rata dari nilai fungsi variabel ke- *j*, karena tidak diketahui variabel mana yang mengalami kejadian terlebih dahulu. Fungsi *partial likelihood* dengan metode *Efron* adalah :

$$L(\beta) = \prod_{i=1}^r \frac{\exp\left(\sum_{l \in D(t_i)} \beta_j x_{lj}\right)}{\prod_{k=1}^{d_i} \left[\sum_{q \in R(t_i)} \exp(\beta_j x_{qj}) - \frac{k-1}{d_i} \sum_{l \in D(t_i)} \exp(\beta_j x_{lj}) \right]} \quad (7)$$

dengan

- $D(t_i)$: himpunan individu yang mengalami *event* pada saat t_i
- $R(t_i)$: himpunan individu yang berisiko mengalami *event* pada saat t_i
- k : individu yang mengalami *ties*
- d_i : banyaknya *ties* yang teramati pada saat t_i

Dari persamaan (7) diperoleh fungsi *log partial likelihood* sebagai berikut :

$$\begin{aligned} \ln L(\beta) &= \ln \left[\prod_{i=1}^r \frac{\exp\left(\sum_{l \in D(t_i)} \beta_j x_{lj}\right)}{\prod_{k=1}^{d_i} \left[\sum_{q \in R(t_i)} \exp(\beta_j x_{qj}) - \frac{k-1}{d_i} \sum_{l \in D(t_i)} \exp(\beta_j x_{lj}) \right]} \right] \\ &= \sum_{i=1}^r \left[\ln \left(\exp\left(\sum_{l \in D(t_i)} \beta_j x_{lj}\right) \right) - \sum_{k=1}^{d_i} \left[\ln \left(\sum_{q \in R(t_i)} \exp(\beta_j x_{qj}) - \frac{k-1}{d_i} \sum_{l \in D(t_i)} \exp(\beta_j x_{lj}) \right) \right] \right] \\ &= \sum_{i=1}^r \left[\left(\sum_{l \in D(t_i)} \beta_j x_{lj} \right) - \sum_{k=1}^{d_i} \left[\ln \left(\sum_{q \in R(t_i)} \exp(\beta_j x_{qj}) - \frac{k-1}{d_i} \sum_{l \in D(t_i)} \exp(\beta_j x_{lj}) \right) \right] \right] \end{aligned} \quad (8)$$

2.7 Pengujian Signifikansi Parameter

Pengujian signifikansi parameter bertujuan untuk memeriksa apakah variabel bebas memiliki pengaruh nyata dalam model regresi. Pengujian signifikansi parameter regresi *cox proportional hazard* dilakukan secara serentak menggunakan uji *partial likelihood ratio* dan secara *partial* atau individu menggunakan uji *Wald*.

2.7.1 Pengujian Signifikansi Parameter Secara Serentak

Uji serentak digunakan untuk menguji pengaruh variabel bebas dalam model regresi secara bersamaan. Statistik uji yang digunakan adalah distribusi *Chi-square* dengan derajat bebas p (banyaknya variabel bebas). Prosedur pengujian parameter yang dilakukan secara serentak adalah sebagai berikut:

1. Hipotesis dari pengujian ini adalah :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$ (variabel bebas tidak berpengaruh terhadap waktu *survival*).

H_1 : minimal ada satu $\beta_j \neq 0$, dimana $j=1,2,\dots,p$ (variabel bebas berpengaruh terhadap waktu *survival*).

2. Taraf signifikansi $\alpha = 0,05$.

3. Statistik uji ditentukan dengan *likelihood ratio test*, yang disimbolkan dengan G . Rumus umum statistik uji *likelihood ratio test* adalah sebagai berikut :

$$G = -2[\ln L(\hat{\beta}) - \ln L(0)] \quad (9)$$

dengan G berdistribusi $\chi^2_{\alpha:p}$.

4. Keputusan bahwa H_0 akan ditolak apabila nilai $G_{hitung} \geq \chi^2_{(\alpha:p)}$ atau $P_{value} \leq \alpha$

5. Pada pengujian ini jika H_0 ditolak berarti disimpulkan bahwa minimal ada satu variabel bebas yang berpengaruh signifikan terhadap fungsi *hazard*.

2.7.2 Pengujian Signifikansi Parameter Secara Partial

Uji parsial digunakan untuk menguji pengaruh parameter (β_j) secara individu. Statistik uji *Wald* mengikuti distribusi *Chi-square* dengan derajat bebas p . Prosedur pengujian parameter yang dilakukan secara *partial* adalah sebagai berikut:

1. Hipotesis dalam pengujian untuk variabel j dimana $j=1,2,\dots,p$ adalah :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

2. Taraf signifikan $\alpha = 0,05$.

3. Statistik uji ditentukan dengan metode pengujian *Wald*, yang disimbolkan dengan W^2 . Rumus umum statistik uji *Wald* adalah sebagai berikut :

$$W^2 = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \quad (10)$$

dengan W^2 berdistribusi $\chi^2_{(\alpha:p)}$ dimana

W^2 : Uji *Wald*

$\hat{\beta}_p$: Koefisien *covariate* ke- p .

$SE(\hat{\beta}_j)$: Standae *error* bagi $\hat{\beta}_p$.

4. Dengan aturan keputusan bahwa H_0 akan ditolak apabila nilai $W^2_{hitung} \geq \chi^2_{(\alpha:1)}$ atau $P_{value} \leq \alpha$.

5. Pada pengujian ini jika H_0 ditolak berarti disimpulkan bahwa varibel bebas ke- j berpengaruh signifikan terhadap fungsi *hazard*.

2.8 Interpretasi Hazard Ratio

Menurut [12] secara umum, *hazard ratio* (HR) didefinisikan sebagai perbandingan dua kelompok antara fungsi *hazard* individu satu dengan fungsi *hazard* individu yang lain. Misalkan individu pertama mempunyai nilai *hazard* $h(t, x_j)$ dimana $x_j = 1$ dan individu kedua mempunyai nilai *hazard* $h(t, x_j^*)$ dimana $x_j^* = 0$, maka berdasarkan persamaan (5) diperoleh bentuk *hazard ratio* yaitu :

$$\begin{aligned}
 HR &= \frac{h(t, x_j)}{h(t, x_j^*)} = \frac{h_0(t) \exp(\beta_j x_j)}{h_0(t) \exp(\beta_j x_j^*)} \\
 &= \exp(\beta_j x_j - \beta_j x_j^*) \\
 &= \exp(\beta_j (x_j - x_j^*)) \\
 &= \exp(\beta_j (1 - 0))
 \end{aligned}
 \tag{11}$$

Menurut [13] interpretasi dari nilai *hazard ratio* yang diperoleh adalah risiko atau laju terjadinya *event* pada individu yang diberikan perlakuan $x_j = 1$ adalah e^{β_j} kali lebih besar dari pada individu yang tidak diberi perlakuan $x_j^* = 0$.

2.9 Pemilihan Model Terbaik

Analisis regresi seringkali digunakan untuk mengkaji hubungan antara beberapa variabel. Agar diperoleh hasil analisis yang optimal, maka diperlukan model regresi terbaik. Pemilihan model regresi terbaik menggunakan metode *Akaike's Information Criterion* (AIC). Metode AIC adalah metode yang dapat digunakan untuk memilih model regresi terbaik dan memiliki nilai AIC terkecil yang ditemukan oleh Akaike yang dinyatakan oleh Faturahman yang dikutip oleh [12]. Metode ini didasarkan pada metode *maximum likelihood estimation* (MLE). Untuk menghitung nilai AIC digunakan rumus sebagai berikut :

$$AIC = e^{\frac{2m}{o} \sum_{i=1}^o \hat{u}_i^2}
 \tag{12}$$

dimana

m : jumlah parameter yang diestimasi dalam model regresi

o : jumlah observasi

\hat{u} : residual R_{ij}

3. Metode

Penelitian ini dilakukan dengan cara mengumpulkan data pasien Tuberkulosis dari rekam medis di Puskesmas Banggae I, Majene pada tahun 2017-2022. Jenis data penelitian ini yaitu bersifat data sekunder di mana data dikumpulkan setelah semua kejadian yang telah terjadi atau lewat. Setelah data dapat dikumpulkan maka data diolah pada *Software Microsoft Excel* untuk menentukan variabel-variabel tuberkulosis dapat dilihat pada **Tabel 1**. selanjutnya data diolah berdasarkan *Software Rstudio* 4.3.0. Penelitian ini merupakan studi literatur yang dirancang dengan menggunakan Model Regresi Cox *Proportional Hazard*.

Tabel 1. Variabel lama perawatan pasien tuberkulosis

Variabel tak bebas	Nama Variabel	Kategori
t	Lama Perawatan pasien penderita tuberculosi	-
d	Status sensor/waktu <i>survival</i>	Waktu <i>survival</i> tersensor = 0 Waktu <i>survival</i> tidak tersensor = 1
x_1	Jenis kelamin	Laki-laki=1, Perempuan=2
x_2	Umur	-
x_3	Riwayat penyakit	Belum pernah=1, Pernah=2
x_4	Pemeriksaan dahak	Dahak positif=1, Dahak negatif=2
x_5	Asal tempat tinggal	Tinggal didalam banggae I=1, Tinggal diluar banggae I=2.

3.1 Tahapan Analisis Data

Tahapan analisis data terdiri dari analisis deskriptif dan analisis model regresi cox. Analisis deskriptif menggunakan aplikasi *Microsoft Excel* 2010 dan analisis model regresi cox menggunakan aplikasi *Rstudio* 4.3.0.

1. Analisis Deskriptif

Analisis deskriptif dimaksudkan untuk memberikan gambaran karakteristik data dari objek penelitian.

2. Analisis Model Regresi Cox

a. Penentuan data tersensor materi *survival*

Jenis penyensoran data yang digunakan adalah sensor kanan, dengan ketentuan, jika pasien perawat jalan dan rawat inap atau dinyatakan sembuh, selama masih dalam waktu penelitian, maka dikategorikan sebagai data *survival* tidak tersensor yang dinyatakan dengan nilai 1 dan jika pasien dinyatakan sembuh melebihi waktu penelitian, maka data *survival* diberi nilai 0.

b. Pengujian asumsi *proportional hazard* dengan uji *goodness of fit*.

c. Penentuan model regresi cox

Membuat model cox PH pada kejadian bersama dilakukan dengan membentuk model awal regresi berdasarkan variabel bebas yang parameternya akan diuji signifikansi. Konstruksi model regresi cox dibentuk berdasarkan persamaan (4) dapat dituliskan sebagai berikut :

$$h(t, x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5) \quad (13)$$

d. Penaksiran parameter model regresi cox PH, dilakukan dengan pendekatan *efron partial likelihood*.

e. Uji sigifikansi parameter, yaitu uji secara serentak *partial likelihood ratio* dan secara *partial*.

f. Nilai *hazard ratio* dan uji kebaikan model.

g. Interpretasi model cox PH.

4. Hasil dan Pembahasan

4.1 Data Penelitian

Data penelitian yang digunakan adalah data sekunder yang berasal dari rekam medis pasien penderita tuberkulosis di Puskesmas Banggae I, Majene dengan ukuran sampel penelitian adalah 202 orang. Data penelitian terdiri dari data tidak tersensor dan data tersensor, dari 202 data pasien yang diperoleh berdasarkan lama perawatan, terdapat 167 data tidak tersensor sebab dapat diamati secara lengkap dan terdapat 35 data yang tersensor karena subjek penelitian hilang atau mengundurkan diri atau sampai akhir penelitian subjek tersebut belum mengalami kejadian tertentu.

4.2 Pengujian Asumsi *Proportional Hazard*

Pemeriksaan asumsi PH dilakukan sebelum penentuan model. Asumsi yang digunakan dalam penelitian ini yaitu pendekatan uji *goodness of fit*. Pengujian asumsi PH dengan pendekatan uji *goodness of fit* dilakukan dengan hipotesis:

$H_0 : p = 0$ (Asumsi *proportional hazard* terpenuhi).

$H_1 : p \neq 0$ (Asumsi *proportional hazard* tidak terpenuhi).

Statistik uji dalam pengujian *goodness of fit* diberikan oleh persamaan (6) dengan taraf signifikansi (0,05) kriteria H_0 diterima jika $P_{value} \geq 0,05$. Hasil perhitungan statistik uji P_{value} menggunakan *software Rstudio* 4.3.0 dan disajikan pada Tabel 2.

Berdasarkan hasil pengujian asumsi PH pada **Tabel 2**. diperoleh nilai P_{value} lebih besar dari α maka H_0 diterima. Sehingga dapat disimpulkan bahwa variabel jenis kelamin, umur, riwayat penyakit, pemeriksaan dahak dan alamat domisili telah memenuhi asumsi PH.

Tabel 2. Hasil pengujian asumsi PH pada variabel bebas

Variabel bebas	P_{value}	Keputusan
Jenis kelamin	0,310	H_0 Diterima
Umur	0,371	H_0 Diterima
Riwayat penyakit	0,496	H_0 Diterima
Pemeriksaan dahak	0,065	H_0 Diterima
Asal tempat tinggal	0,845	H_0 Diterima

4.3 Model Analisis Regresi Cox Proportional Hazard

Langkah selanjutnya setelah dilakukan pengujian asumsi cox PH maka parameter β_p pada model regresi Cox PH merupakan parameter yang akan digunakan untuk menduga *maximum Partial likelihood estimation* (MPLE) dengan metode *Efron*. Hasil penghitungan penaksiran parameter awal model cox PH menggunakan *software Rstudio* 4.3.0 dan semua variabel bebas dapat dilihat pada **Tabel 3.** yaitu :

Tabel 3. Hasil estimasi parameter model awal cox untuk variabel bebas

Variabel bebas	Parameter	β_p	$\exp(\beta_p)$
Jenis kelamin	β_1	0,081761	1,085148
Umur	β_2	-0,007767	0,002263
Riwayat penyakit	β_3	0,345787	1,326333
Pemeriksaan dahak	β_4	0,401393	1,493904
Asal tempat tinggal	β_5	-0,093660	0,910592

Berdasarkan hasil estimasi parameter awal model cox untuk variabel bebas pada **Tabel 3.** diasumsikan semua variabel berpengaruh terhadap model, maka semua variabel dimasukkan dalam persamaan umum model cox, sehingga diperoleh estimasi awal model cox *proportional hazard* dengan metode *Efron* sebagai berikut:

$$h(t, x) = h_0(t) \exp(0,081761x_1 - 0,007767x_2 + 0,345787x_3 + 0,401393x_4 + (-0,093660)x_5)$$

Setelah penaksiran parameter, maka pengujian selanjutnya yaitu pengujian parameter yang terdiri dari pengujian hipotesis parameter regresi secara serentak dan parameter regresi secara *partial*.

4.4 Pegujian Signifikansi Parameter

4.4.1 Pengujian Signifikansi Parameter Secara Serentak

Statistik uji dalam pengujian serentak diberikan oleh persamaan (9) dengan taraf signifikansi 0,05 kriteria H_0 ditolak jika $G_{hitung} \geq \chi^2_{(0,05;5)}$ atau $P_{value} \leq 0,05$. Perhitungan statistik uji menggunakan *software Rstudio* 4.3.0 dan hasil pengujian hipotesis parameter cox PH secara serentak disajikan pada Tabel 4. Berdasarkan uji serentak untuk semua variabel pada **Tabel 4.** diperoleh nilai $G_{hitung} = 9,2266 < \chi^2_{(0,05;5)} = 11,0705$ atau $P_{value} = 0,1 > \alpha = 0,05$, maka H_0 gagal ditolak dan dapat disimpulkan bahwa model regresi dengan 5 variabel bebas di uji secara bersamaan tidak berpengaruh terhadap model cox PH atau model cox PH belum dapat digunakan.

Tabel 4. Uji serentak semua variable bebas

Variabel bebas	Nilai G_{hitung}	Nilai β_p	P_{value}
x_1, x_2, x_3, x_4, x_5	9,2266	11,0705	0,1

Langkah selanjutnya yaitu pemilihan model terbaik didasarkan pada hasil pengujian parameter secara serentak yang ditunjukkan oleh nilai statistika uji G_{hitung} atau P_{value} juga didasarkan pada nilai AIC. Hasil dari pengujian metode *backward elimination* yang memuat dua variabel bebas yaitu variabel umur dan variabel pemeriksaan dahak yang dinyatakan oleh nilai AIC yang paling kecil yaitu sebesar 1409,283. Berdasarkan nilai G , didapat nilai $G_{hitung} = 7,69$ lebih besar dari pada $\chi^2_{tabel} = 5,99$ atau $P_{value} = 0,02$ lebih kecil dari pada nilai $\alpha = 0,05$ sehingga dapat disimpulkan bahwa model regresi cox PH dengan variabel umur (x_2) dan pemeriksaan dahak (x_4) sudah layak digunakan.

Hasil penaksiran model cox PH dengan variabel bebas yang berpengaruh yaitu umur (x_2) dan pemeriksaan dahak (x_4) dengan menggunakan *software Rstudio* 4.3.0 dan disajikan pada **Tabel 5.** yaitu:

Tabel 5. Hasil estimasi parameter model cox untuk variabel bebas berpengaruh

Variabel bebas	Parameter	β_p	$\exp(\beta_p)$
Umur	β_2	-0,007767	0,002263
Pemeriksaan dahak	β_4	0,401393	1,493904

Berdasarkan Tabel 5. model regresi cox PH pada data lama perawatan pasien penderita tuberkulosis adalah:

$$h(t, x) = h_0(t) \exp((-0,006)x_2 + 0,388x_4)$$

Tahapan selanjutnya setelah pengujian signifikansi parameter secara serentak, maka selanjutnya pengujian signifikansi parameter secara *Partial*.

4.4.2 Pengujian Signifikansi Parameter Secara *Partial*

Statistik uji dalam pengujian *partial* diberikan oleh persamaan (10) dengan taraf signifikansi $\alpha = 0,05$. Kriteria H_0 ditolak jika $P_{value} \leq 0,05$. Perhitungan statistik uji menggunakan *software Rstudio* 4.3.0 dan hasil pengujian hipotesis parameter cox PH secara *Partial* disajikan pada **Tabel 6.** sebagai berikut:

Tabel 6. Hasil keputusan uji *partial* untuk variabel bebas berpengaruh

Variabel bebas	Parameter	P_{value}	Keputusan
Umur	β_2	0,2	H_0 gagal ditolak
Pemeriksaan dahak	β_4	0,02	H_0 ditolak

Berdasarkan uji *partial* untuk variabel bebas pada Tabel 6. diperoleh nilai $P_{value} \leq 0,05$ yaitu sebesar 0,02 maka H_0 ditolak. Dapat disimpulkan bahwa secara individual variabel yang berpengaruh terhadap lama waktu perawatan pasien penderita tuberkulosis adalah variabel pemeriksaan dahak.

4.5 Interpretasi Hazard Ratio (HR)

Setelah dilakukan analisis di atas diperoleh model regresi cox *proportional hazard* yang dihasilkan dengan metode *efron* adalah sebagai berikut:

$$h(t, x) = h_0(t) \exp(0,388x_4)$$

Dari variabel yang berpengaruh didapat nilai *hazard rasio* untuk variabel pemeriksaan dahak adalah $e^{0,388} = 1,474$, hal ini menunjukkan bahwa setiap pasien yang terdeteksi adanya bakteri disaluran pernafasan yang dikategorikan positif akan mengalami peningkatan 1,474 kali terhadap lama perawatan pasien penderita tuberkulosis dari pada kategori negatif.

5. Kesimpulan

Berdasarkan hasil analisis dan pembahasan, diperoleh faktor yang berpengaruh signifikan terhadap lama perawatan pasien penderita tuberkulosis di Puskesmas Banggae I, Majene yaitu faktor pemeriksaan dahak (x_4), dengan model

$$h(t, x) = h_0(t) \exp(0,388x_4)$$

Dari variabel yang berpengaruh diperoleh nilai *hazard rasio* untuk variabel pemeriksaan dahak adalah $e^{0,388} = 1,474$, hal ini menunjukkan bahwa setiap pasien yang terdeteksi adanya bakteri disaluran pernafasan yang dikategorikan positif akan mengalami peningkatan 1,474 kali terhadap lama perawatan pasien penderita tuberkulosis dari pada kategori negatif.

Referensi

- [1] C. H. Imanina, "ANALISIS SURVIVAL TERHADAP PASIEN PENYAKIT GINJAL KRONIS DENGAN MENGGUNAKAN COX REGRESSION," Universitas Islam Indonesia, 2018. [Daring]. Tersedia pada: <https://dspace.uii.ac.id/handle/123456789/6525>
- [2] S. Prabawati, Y. N. Nasution, dan S. Wahyuningsih, "Analisis Survival Data Kejadian Bersama dengan Pendekatan Efron Partial Likelihood," *EKSPONENSIAL*, vol. 9, no. 1, hal. 75–84, 2018.
- [3] R. Isnaeni, Y. N. Nasution, dan S. Wahyuningsih, "Analisis Survival pada Data Kejadian Bersama Menggunakan Metode Exact Partial Likelihood," *EKSPONENSIAL*, vol. 9, no. 2, hal. 145–152, 2019.
- [4] CNNIndonesia, "Indonesia Peringkat Dua Kasus TBC Terbanyak di Dunia," *www.cnnindonesia.com*, 2023. [Daring]. Tersedia pada: <https://www.cnnindonesia.com/gaya-hidup/20230324112557-255-928806/indonesia-peringkat-dua-kasus-tbc-terbanyak-di-dunia>
- [5] S. Fatimah, "Analisis Regresi Cox Proportional Hazard Dengan Pendekatan Metode Exact Partial Likelihood" Universitas Sulawesi Barat," Universitas Sulawesi Barat, 2022. [Daring]. Tersedia pada: <https://repository.fmipa.unsulbar.ac.id/posts/14>
- [6] D. M. Ulfah, "Penerapan Analisis Survival Menggunakan Metode Regresi Cox Pada Pasien Hipertensi (Studi Kasus Pasien Rawat Inap Di Rsud Dr. Soedomo Trenggalek)," Universitas Brawijaya, 2018. [Daring]. Tersedia pada: <https://repository.ub.ac.id/id/eprint/168583/>
- [7] I. Ihsan, S. Suyitno, dan S. Wahyuningsih, "Model Regresi Cox Weibull Dengan Metode Penaksiran Parameter Efron Partial Likelihood," *EKSPONENSIAL*, vol. 11, no. 1, hal. 1–8, 2021.
- [8] N. Soraya, Y. N. Nasution, dan S. Wahyuningsih, "Model Cox Proportional Hazard Pada Kejadian Bersama (Ties) dengan Metode Breslow," *Eksponensial*, vol. 9, no. 1, hal. 95–104, 2018.
- [9] R. T. Yohanes, "Pengecekan Asumsi Proportional Hazard Pada Model Cox PH," Universitas Indonesia, 2011. [Daring]. Tersedia pada: <https://lib.ui.ac.id/detail?id=20289275&lokasi=lokal>
- [10] Y. K. Darwanto, "Penerapan Model Cox Proportional Hazard untuk Menganalisis Jangka Waktu Kemampuan Pembayaran Premi Asuransi Kendaraan," Universitas Jember, 2020. [Daring]. Tersedia pada: <https://repository.unej.ac.id/handle/123456789/102830>
- [11] D. K. Vianda, "ANALISIS SURVIVAL PADA MODEL REGRESI COX PROPORTIONAL HAZARD DENGAN MENGGUNAKAN METODE BRESLOW DAN METODE EFRON,"

- Universitas Lampung, 2021. [Daring]. Tersedia pada: <https://digilib.unila.ac.id/60967/>
- [12] N. Nurmala, “MODEL HAZARD RATIO UNTUK ANALISIS KETAHANAN HIDUP PASIEN KANKER PARU-PARU,” Universitas Jember, 2017. [Daring]. Tersedia pada: <https://repository.unej.ac.id/handle/123456789/79534>
- [13] M. Ilmiah, “Analisis Survival Dengan Model Regresi Cox (Studi Kasus : Pasien Tuberkulosis Di RSUD Dr. Soetomo Surabaya),” Institut Teknologi Sepuluh Nopember Surabaya, 2015. [Daring]. Tersedia pada: <https://repository.its.ac.id/72068/>



© **The Author(s) 2024.** This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. Editorial of Journal of Mathematics: Theory and Applications, Department of Mathematics, Universitas Sulawesi Barat, Jalan Prof. Dr. Baharuddin Lopa, S.H., Talumung, Majene 91412, Sulawesi Barat.