

ANALISIS MODEL REGRESI COX PROPORTIONAL HAZARD PADA STUDI KASUS PASIEN KANKER PARU-PARU

JUNITA AMALIA
INSTITUT TEKNOLOGI DEL
junita.amalia@del.ac.id

ABSTRACT

The high mortality rate caused by lung cancer is at an alarming rate. Lung cancer patients must be treated quickly and appropriately. One method often used in survival analysis is the cox proportional hazard regression model. In the cox proportional hazard regression model, the independent variables used must meet the proportional hazard assumption. The results showed that the variables that had a significant effect on the Cox Proportional Hazard regression analysis using a 90% confidence level in the survival time of lung cancer patients were cell type 2, cell type 3 and performance status. Meanwhile, other variables namely treatment, cell type 1, disease duration, age, and prior therapy did not significantly influence.

Keywords : *Lung Cancer, Survival Analysis, Cox Proportional Hazard*

PENDAHULUAN

Kanker paru adalah penyakit yang ditandai dengan tidak terkendalinya pertumbuhan sel dalam jaringan paru, terutama sel-sel yang melapisi bagian pemapasan. Tingginya angka kematian yang disebabkan oleh kanker paru-paru ini berada pada tingkat yang mengkhawatirkan. Pasien kanker paru-paru harus segera ditangani dengan cepat dan tepat. Oleh sebab itu, sangat penting mengetahui faktor-faktor yang dapat mempengaruhi waktu ketahanan hidup pasien kanker paru-paru. Faktor faktor ini dapat dianalisis dengan menggunakan analisis *survival*.

Analisis ketahanan hidup (*survival*) adalah analisis mengenai data yang diperoleh dari catatan waktu yang dicapai suatu obyek sampai terjadinya peristiwa tertentu yang disebut sebagai *failure event*. Salah satu metode yang sering digunakan dalam analisis survival adalah model *cox proportional hazard* (Kleinbaum & Klein, 2005). Metode ini pertama kali dikenalkan oleh Cox dan respon yang digunakan adalah data yang diperoleh dari perhitungan waktu ketahanan hidup (*survival*) dari suatu peristiwa. Pada model *cox proportional hazard*, variabel independen yang digunakan harus memenuhi asumsi proportional hazard artinya seluruh variabel independen harus konstan sepanjang waktu.

TINJAUAN PUSTAKA

Analisis Survival

Analisis survival adalah metode statistika yang digunakan dalam mempelajari ketahanan hidup yang berhubungan dengan waktu, mulai waktu awal (*time origin*) yang sudah ditentukan dalam penelitian sampai waktu akhir penelitian (end point). Sehingga analisis survival membutuhkan waktu survival dari suatu individu. Kejadian yang terjadi (*failure event*) dapat berupa kejadian meninggal, sakit, respon terhadap suatu percobaan yang dilakukan dalam penelitian yaitu sesuai dengan kejadian yang diteliti (Kleinbaum dan Klen, 2005).

Menurut Collet (1994) ada tiga jenis sensor dalam analisis survival, yaitu:

- a.) Sensor Kanan (*Right Censored*) jika observasi dari awal penelitian belum mengalami *failure event* sampai akhir penelitian.
- b.) Sensor Kiri (*Left Censored*) jika *failure event* telah terjadi sebelum penelitian dimulai.

c.) Sensor Interval (*Interval Censored*) jika *failure event* dari pasien terjadi pada interval penelitian akan tetapi tidak teramati

Fungsi Survival dan Fungsi Hazard

Jika T melambangkan waktu survival dan mempunyai distribusi peluang $f(t)$, dimana T adalah variabel random yang tidak negatif. Fungsi distribusi kumulatif $F(t)$ dinyatakan sebagai berikut :

$$F(t) = P(T \leq t) = \int_0^t f(t)dt$$

Fungsi survival $S(t)$, didefinisikan sebagai probabilitas suatu obyek bertahan setelah waktu ke- t , dinyatakan dengan persamaan sebagai berikut (Le, 1997):

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

Fungsi hazard $h(t)$ merupakan laju failure atau kegagalan sesaat dengan asumsi obyek telah bertahan sampai waktu ke- t , oleh Le (1997) dinyatakan sebagai berikut :

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

Kemudian setelah mengintegrasikan dan mengeksponensialkan diperoleh hubungan antara fungsi kumulatif hazard dengan fungsi survival adalah:

$$H(t) = -\ln S(t)$$

dimana:

$f(t)$ = Fungsi kepadatan peluang/Probability density function (PDF)

$h(t)$ = Fungsi hazard proporsional

$S(t)$ = Fungsi survival

$F(t)$ = Fungsi distribusi kumulatif (CDF)

$H(t)$ = Fungsi kumulatif hazard

Regresi Cox

Regresi Cox pertama kali dikenalkan oleh Cox dan respon yang digunakan adalah data yang diperoleh dari perhitungan waktu suatu peristiwa tertentu (waktu survival). Misalnya data tentang waktu pasien menderita penyakit tertentu, dimana perhitungannya dimulai dari awal sakit hingga terjadi pada kejadian khusus, yaitu seperti kematian, sembuh atau kejadian khusus lainnya (Cox, 1972).

METODE PENELITIAN

Sumber Data

Data yang digunakan merupakan data *survival time* dalam hari pada 137 pasien dari *Veteran's Administration Lung Cancer Trial cited by Kalbfleisch dan Prentice* pada bukunya (*The Statistical Analysis of Survival Time Data*), John Wiley, pp 223-224, 1980).

Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel dependen dan variabel independen. Berikut merupakan uraian dari variabel-variabel yang digunakan.

i. Variabel Dependen Variabel dependen yang digunakan adalah *survival time (days)* dan status (*censored=0, died=1*).

ii. Variabel Independen

Variabel Independen yang digunakan dirincikan pada Tabel 1, sebagai berikut.

Tabel 1. Variabel Independen

Variabel	Deskripsi	
<i>Treatment</i>	<i>Standart = 1</i>	<i>Test = 2</i>
<i>Cell type 1</i>	<i>Large = 1</i>	<i>Other = 0</i>
<i>Cell type 2</i>	<i>Adeno = 1</i>	<i>Other = 0</i>
<i>Cell type 3</i>	<i>Small= 1</i>	<i>Other = 0</i>
<i>Cell type 4</i>	<i>Squamous =1</i>	<i>Other = 0</i>
<i>Survival time</i>	<i>Days</i>	
<i>Performance Status</i>	<i>(0 = worst ... 100= best</i>	
<i>Disease duration</i>	<i>Months</i>	
<i>Age</i>		
<i>Prior therapy</i>	<i>None = 0</i>	<i>Some =1</i>

Tahapan Analisis Data

Tahapan analisis data yang akan dilakukan sebagai berikut:

1. Mendeskripsikan karakter waktu *survival* pasien kanker paru-paru beserta faktor-faktor yang diduga berpengaruh terhadap ketahanan hidup pasien kanker paru-paru meliputi *treatment, cell type 1, cell type 2, cell type 3, cell type 4, performance status, diasese duration, age dan prior therapy*.
2. Membuat kurva *survival Kaplan-Meier* dari faktor faktor yang diduga berpengaruh terhadap ketahanan hidup pasien kanker paru-paru.
3. Melakukan *uji log rank* untuk mengetahui ada tidaknya perbedaan kurva *survival* pasien kanker paru-paru.
4. Memeriksa asumsi *portional hazard*.
5. Membuat model regresi *cox propotional hazard*.

HASIL DAN PEMBAHASAN

Fungsi *Survival* dan Fungsi *Hazard* Pasien Kanker Paru-paru

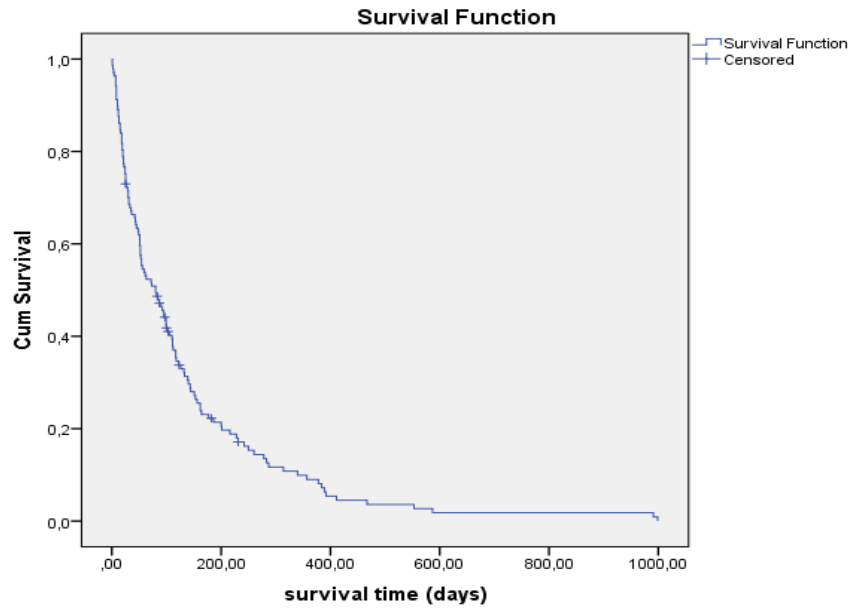
Karakteristik waktu *survival* pasien kanker paru-paru dapat ditunjukkan dengan melihat nilai mean dan median *survival time* serta menggunakan kurva kaplan meier. Pada Tabel 2 diberikan nilai mean dan median dari *survival time* pasien kanker paru-paru.

Tabel 2. Mean dan Median *Survival Time* Pasien Kanker Paru Paru

Mean			Median		
Estimate	95% Confidence Interval		Estimate	95% Confidence Interval	
	Lower Bound	Upper Bound		Lower Bound	Upper Bound
132,777	102,655	162,898	80,000	49,187	110,813

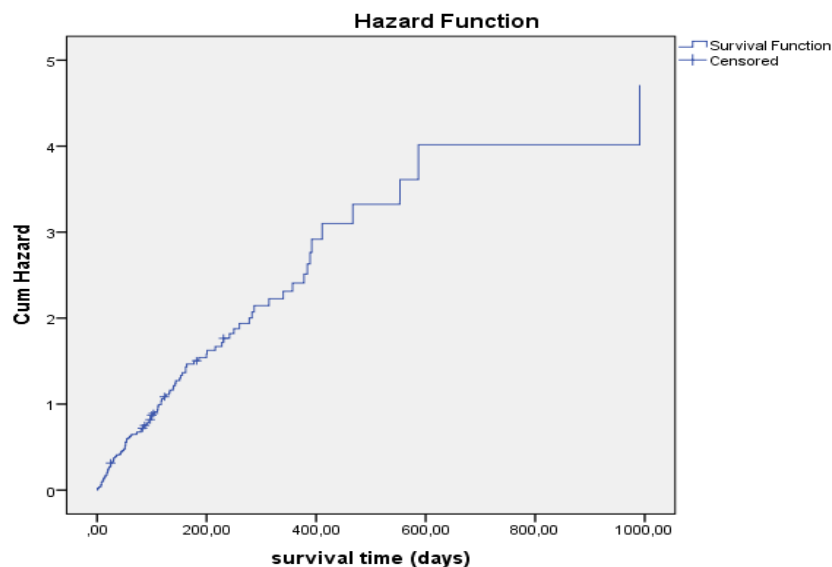
Dari Tabel 2 dapat diketahui, rata-rata dari 137 penderita kanker paru-paru mampu bertahan hidup hingga 133 hari, dengan selang kepercayaan anatra 103 sampai 163 hari. Sedangkan median dari *survival time* adalah 80 artinya 50% dari penderita kanker paru-paru akan mampu bertahan hidup hingga hari ke 80 dan 50% dari penderita kanker paru-paru lainnya akan meninggal setelah hari ke 80, dengan selang kepercayaan antara 49 sampai 111.

Selain menggunakan nilai mean dan median, karakteristik *survival time* dapat ditunjukkan melalui kurva kaplan meier. Gambar 1 menunjukkan kurva Kaplan Meier pada 137 pasien penderita kanker paru-paru.



Gambar 1. Kurva Kaplan Meier Pasien Kanker Paru Paru

Berdasarkan Gambar 1 dapat diketahui bahwa pada hari ke-0 sampai hari ke-400 kurva kaplan meier turun cepat dari nilai mendekati 1 hingga mendekati 0. Sementara itu pada hari ke 500 hingga 1000 nilainya cenderung turun lambat dan konstan mulai hari ke- 600. Hal ini menunjukkan bahwa pada hari dimulainya pengamatan peluang pasien bertahan hidup sangat tinggi yaitu mendekati 1 namun peluangnya semakin hari menurun dengan cepat hingga hari ke 600 peluang pasien bertahan hidup sangat kecil bahkan hampir mendekati 0. Berdasarkan data *survival time* untuk 137 pasien, dapat dibuat kurva *hazard function* seperti berikut.



Gambar 2. Kurva Hazard Function Pasien Kanker Paru Paru

Fungsi *hazard* dikenal juga sebagai *hazard rate* yang dinotasikan dengan $h(t)$. Fungsi ini didefinisikan sebagai kelajuan suatu individu untuk mengalami event pada interval waktu t sampai apabila diketahui individu tersebut belum mengalami event sampai 5 dengan waktu t . *Hazard function* memberikan informasi yang berlawanan dengan *survival function*. Berdasarkan Gambar 2 dapat dilihat bahwa nilai *hazard function* selalu naik. Kurva *hazard function* menggambarkan semakin lama laju terjadinya event terhadap seseorang semakin besar, sampai $t=600$ nilai $h(t)$ cenderung konstan. Karakteristik waktu *survival* yang ditunjukkan pada Gambar 2 merupakan kurva *survival* kaplan meier secara umum. Kemudian karakteristik *survival time* pada masing-masing variabel faktor yang diduga mempengaruhi ketahanan hidup pasien penderita kanker paru-paru yaitu tipe sel, usia, *treatment*, *performance status*, *disease duration*, dan *Prior therapy* dilakukan analisis seperti di atas.

Log Rank

Log rank digunakan untuk mengevaluasi apakah kaplan meier pada dua atau lebih grup yang ada pada tiap-tiap variabel dalam studi kasus ini ekuivalen secara statistik atau tidak. Hipotesis yang digunakan dalam pengujian ini adalah, H_0 tidak terdapat perbedaan yang signifikan pada kaplan meier antar grup pada variabel tertentu, sedangkan H_1 terdapat perbedaan yang signifikan diantara grup pada variabel tertentu. Statistik log rank merupakan pendekatan dari chi-square dengan derajat bebas sama dengan $G-1$ dimana G adalah jumlah grup yang ada pada variabel. Tabel 12 menunjukkan hasil statistik log-rank pada masing-masing variabel.

Tabel 3. Statistik Uji Log Rank pada masing-masing Variabel

Variabel	Chi-Square	df	P-Value
Treatment	0,008	1	0,928
Cell type 1	3,021	1	0,082
Cell type 2	8,192	1	0,004
Cell type 3	10,203	1	0,001
Cell type 4	10,531	1	0,001
Performance status	29,181	2	0,000
Disease duration	0,049	1	0,825
Age	3,221	1	0,073
Prior therapy	0,501	1	0,479

Dengan menggunakan $\alpha=5\%$, terdapat beberapa variabel yang hasil statistik log-rank nya signifikan. variabel tersebut diantaranya, *Cell type 2*, *Cell type 3*, *Cell type 4* dan *Performance status*. Untuk *cell type 2* maka Kaplan meier antara tipe sel *adeno* dan selain *adeno* (0) berbeda signifikan, begitu pula untuk *cell type 3* dan *cell type 4*. Terdapat perbedaan yang di signifikan pada kaplan meier antara tipe sel *small* dan tipe sel selain *small* dan terdapat perbedaan yang signifikan pula antara tipe sel *squamos* dan tipe sel yang lain. Selanjutnya yaitu untuk variabel *performance status*, nilai log rank yang signifikan pada variabel ini berarti terdapat perbedaan kaplan meier antara *performance status* pada grup nilai 0-59, 60-74, dan 75-100.

Sementara itu statistik log rank tidak signifikan pada variabel *treatment*, *cell type 1*, *disease duration*, *age*, dan *Prior therapy*. Untuk variabel *treatment* maka tidak terdapat perbedaan kaplan meier pada individu yang diberi *treatment standard* dan yang diberi *treatment test*, sedangkan untuk *cell type 1* tidak terdapat perbedaan kaplan meier pada pasien dengan tipe sel *large* dan tipe sel selain *large*, begitu pula untuk variabel *disease duration* yaitu tidak ada perbedaan antara kaplan meier untuk grup *disease duration* yang bernilai diatas 8,7 bulan dan dibawah 8,7 bulan. Untuk variabel usia tidak ada perbedaan antara kaplan meier untuk usia pasien yang usianya dibawah 58 tahun dan pasien yang usianya diatas 68 tahun, untuk variabel *Prior therapy* tidak terdapat perbedaan antara kaplan meier untuk pasien yang pernah mendapatkan terapi dan yang belum pernah mendapatkan terapi.

Pengujian Asumsi Proportional Hazard

Uji asumsi proportional hazard dilakukan sebelum melakukan pembentukan model. Pembentukan model bertujuan untuk mengetahui faktor-faktor mana saja yang memenuhi asumsi proportional hazard dan yang tidak memenuhi asumsi untuk selanjutnya dilakukan pemodelan berdasarkan hasil pengujian yang diperoleh. Metode yang digunakan ada 2 yaitu secara visual dengan menggunakan grafik $\ln(-\ln \hat{S}(t))$ dan inferensi menggunakan uji Goodness of fit. Kedua metode tersebut digunakan untuk memperoleh hasil yang akurat.

Secara ringkas, hasil uji asumsi proporsional hazard dengan menggunakan uji Goodness of fit ditampilkan dalam Tabel 4 sebagai berikut.

Tabel. 4 Hasil Uji Goodness of Fit

Variabel	Korelasi	P-Value
<i>Treatment</i>	0,043	0.628
<i>Cell type 1</i>	0,188	0.033
<i>Cell type 2</i>	0.155	0.081
<i>Cell type 3</i>	-0.156	0.078
<i>Performance Status</i>	0.313	0.000
<i>Disease duration</i>	-0.009	0.919
<i>Age</i>	0.115	0.198
<i>Prior therapy</i>	-0.129	0.145

Berdasarkan Tabel 4 dapat disimpulkan bahwa pada taraf signifikan 5% variabel yang tidak memenuhi asumsi proportional hazard adalah cell type 1 dan performance status sedangkan jika menggunakan taraf signifikan 1% variabel yang tidak memenuhi asumsi proportional hazard adalah cell type 1, cell type 2, cell type 3 dan performance status. Alternatifnya, kita kombinasikan kategori sebagai cell type 1, cell type 2, cell type 3 dan performance status. Namun, kombinasi kategori seharusnya tidak hanya berdasarkan keputusan statistik tetapi juga berdasarkan kajian biologi. Berdasarkan kajian biologi yang perlu distratifikasi hanya variabel cell type 3, yang mana variabel 3 dijadikan dikotomus (1=type cell small, 0= type cell lain). Jadi variabel yang akan menjadi variabel stratifikasi adalah cell type 3(small cell) dan performance status.

Pemodelan Regresi Cox Proportional Hazard

Berikut merupakan hasil estimasi parameter model regresi Cox proportional hazard untuk data waktu survival pasien-pasien kanker paru-paru.

Tabel. 5 Estimasi Parameter Model Regresi Cox proportional hazard

Variabel	Estimasi	Hazard Rasio	P-Value
<i>Treatment</i>	0,290	1,336	0,162
<i>Cell type 1</i>	0,400	1,491	0,157
<i>Cell type 2</i>	1,188	3,281	0,000
<i>Cell type 3</i>	0,856	2,355	0,002
<i>Performance Status</i>	-0,033	0,986	0,000
<i>Disease duration</i>	0,000	1,000	0,992
<i>Age</i>	-0,009	0,991	0,359
<i>Prior therapy</i>	0,007	1,007	0,755

Dari Tabel 5 maka diperoleh model regresi cox proportional hazard:

$$\hat{h}(t) = \hat{h}_0(t)\exp(0,29x_1 + 0,4x_2 + 1,188x_3 + 0,856x_4 - 0,033x_5 - 0,009x_7 + 0,007x_8)$$

Berdasarkan nilai p-value, dengan menggunakan kepercayaan 90% terdapat beberapa variabel yang berpengaruh signifikan terhadap survival time pasien kanker paru-paru yaitu *Cell type 2*, *Cell type 3* dan *Performance Status*. Sementara itu variabel yang lain yaitu *treatment*, *cell type 1*, *disease duration*, *age*, dan *prior therapy* tidak berpengaruh secara signifikan.

Interpretasi dari tabel 5, laju terjadinya event yaitu meninggal pada pasien kanker paru-paru yang mendapat *treatment standard* 1,336 kali dari pasien yang diberi *treatment test*. Setiap pasien kanker paru-paru dengan *cell type 1 Large* maka kemungkinan laju terjadinya event akan meningkat 1,491 kali dari pasien dengan *cell type 1* lain, pasien dengan *cell type 2 Adeno* laju terjadinya event 3,281 kali dari pasien *cell type 2* lain dan pasien *cell type 3 small* laju terjadinya event 2,355 kali dari pasien *cell type 3* lain. Untuk pasien kanker paru-paru apabila *performance status Worst* maka laju terjadinya event yaitu meninggal sebesar 0,986 kali. Setiap penambahan satu tahun usia pasien kanker paru-paru, maka laju terjadinya event yaitu meninggal pada pasien kanker paru-paru akan meningkat 0.991 kali. Laju terjadinya event yaitu meninggal pada pasien kanker paru-paru yang mendapat *prior therapy* 1,007 kali dari pasien yang tidak mendapat prior therapy.

KESIMPULAN

Variabel yang berpengaruh signifikan pada analisis *Regresi Cox Proportional Hazard* dengan menggunakan taraf kepercayaan 90% pada survival time pasien kanker paru-paru yaitu *Cell type 2*, *Cell type 3* dan *Performance Status*. Sementara itu, variabel yang lain yaitu *treatment*, *cell type 1*, *disease duration*, *age*, dan *prior therapy* tidak berpengaruh secara signifikan.

DAFTAR PUSTAKA

Collet, D., (1994), *Modelling Survival Data in Medical Reseach*. London, Chapman and Hall.

Cox, D. R., (1972), *Regression Model and Life Table*. J Roy Stat Soc B, 34, 187- 202.

Kleibaum, D.G & Klein, M., (2005), *Survival Analysis: A Self-Learning Text, Second Edition*. New York, Springer.

Le, C. T., (1997), *Applied Survival Analysis*. New York, John Wiley and Sons, Inc.