

# Stratified Cox Regression Approach to Identifying Prognostic Factors for Survival in Breast Cancer Patients

**Dhio Ervandi, Aisyah Novriani , Andini Diva Luthfiyah, Fauzan Alhamdani Siregar, dan Tessy Octavia Mukhti\***

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: [tessyoctaviam@fmipa.unp.ac.id](mailto:tessyoctaviam@fmipa.unp.ac.id)

Submitted : 09 Desember 2024

Revised : 24 Oktober 2025

Accepted : 24 Oktober 2025

## ABSTRACT

*Breast cancer is the number one type of cancer that commonly affects women. Around 2.3 million women were diagnosed with breast cancer in 2022, and about 670,000 died from the disease globally. By 2040, it is estimated that breast cancer will increase by 40%, reaching 3 million annually with the number of deaths increasing by 50% to 1 million in 2020. This highlights breast cancer as a serious threat to world health. This study utilized secondary data from METABRIC or the Molecular Taxonomy of Breast Cancer International Consortium obtained from the website [www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric/data](http://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric/data). The independent variables analyzed were, Age at Diagnosis ( $X_1$ ), Surgery Type ( $X_2$ ), Chemotherapy ( $X_3$ ), Hormone Therapy ( $X_4$ ), Tumor Size ( $X_5$ ), Radio Therapy ( $X_6$ ), Pam50. The dependent variables were Survival Time (Overall Survival Month) and Patient Status. In this study, we used the Stratified Cox model to predict the predictor variables of survival time. The total number of patients used was 1868, with 1080 censored patients and 788 uncensored patients. The Stratified Cox model without interaction revealed that the patients who underwent breast-conserving surgery had a 1.35 times higher risk of death compared to those who underwent mastectomy. Patients who received chemotherapy had a 2.01 times higher risk of death than those who did not, while patients who did hormone therapy had a 1.83 times higher risk of death than those who did not undergo this therapy.*

**Keywords:** *Cancer, Hazard ratio, Breast Cancer, Stratified Cox, Survival Analysis*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Kanker payudara merupakan jenis kanker nomor satu yang paling banyak menyerang wanita. Pada tahun 2022 terdapat 2,3 juta wanita yang terdiagnosis kanker payudara dengan 670.000 kematian yang tercatat secara global (WHO, 2024). Pada tahun 2040, diperkirakan kanker payudara akan meningkat 40% dengan 3 juta kasus setiap tahunnya dan jumlah kematian yang meningkat 50% menjadi 1 juta kematian berdasarkan data 2020 (Arnold et al., 2022). Hal ini menjadikan kanker payudara sebagai ancaman yang serius bagi kesehatan dunia.

Penelitian-penelitian terdahulu telah berupaya meneliti faktor-faktor yang memengaruhi ketahanan hidup pasien kanker payudara. Farisyi (2018) menemukan adanya hubungan antara ukuran tumor dengan kelangsungan hidup bebas penyakit pada pasien kanker payudara. Penelitian Agustina (2019) menunjukkan bahwa usia saat terdiagnosis, stadium kanker, dan ukuran tumor memiliki pengaruh signifikan terhadap ketahanan hidup penderita. Hal ini juga menjadi perhatian dalam penelitian sebelumnya yang dilakukan oleh Maajani dkk (2019), yang menunjukkan bahwa tingkat kelangsungan hidup pasien kanker payudara dipengaruhi oleh berbagai faktor, dua di antaranya yang paling penting adalah usia dan stadium saat diagnosis. Meskipun demikian, penelitian-penelitian tersebut masih terbatas pada variabel tertentu dan belum memperhitungkan efek dari karakteristik pasien yang tidak memenuhi asumsi proportional hazard.

Berdasarkan penelitian sebelumnya, peneliti ingin melakukan penelitian terkait dengan menggunakan lebih banyak variabel. Pada penelitian kali ini akan dilakukan pemodelan waktu *survival* pasien penderita kanker payudara dengan metode *stratified cox survival*. *Stratified cox survival* merupakan metode yang digunakan jika terdapat variabel yang tidak memenuhi asumsi *proportional hazard*. Metode ini tidak hanya memodelkan waktu *survival* namun juga memperhitungkan perbedaan karakteristik pada pasien. Penelitian ini dilakukan untuk mengetahui faktor-faktor yang berperan dalam menentukan lamanya ketahanan hidup pasien kanker payudara, sehingga dapat digunakan sebagai acuan dalam merancang strategi pengobatan yang lebih baik di masa mendatang.

## II. METODE PENELITIAN

Sumber data dalam penelitian ini adalah data sekunder, yakni METABRIC atau *The Molecular Taxonomy of Breast Cancer International Consortium* yang diperoleh dari website [www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric/data](http://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric/data). Data ini merupakan data pasien kanker payudara yang dikumpulkan dari tahun 1977 sampai 2005 di Inggris dan Kanada dengan jumlah 1904 pasien. Variabel dependen yang digunakan adalah waktu survival (*overall survival month*) dan status penderita. Sedangkan variable independen yang digunakan pada penelitian ini adalah usia saat terdiagnosis, tipe operasi payudara, kemoterapi, terapi hormonal, jumlah mutasi, ukuran tumor, terapi radiasi, dan PAM50 yang merupakan profil genetik yang digunakan untuk mengklasifikasikan kanker payudara. Berikut adalah tabel dari variabel yang digunakan:

**Tabel 1.** Variabel Penelitian

Variabel	Keterangan
<i>Overall Survival Month</i>	Lama waktu bertahan hidup
<i>Overall Survival</i>	0 : Tersensor 1 : Tidak tersensor
Umur Saat Terdiagnosis ( $X_1$ )	Umur ketika terkena kanker payudara
Tipe Operasi ( $X_2$ )	<i>Breast Conserving Mastectomy</i>
Kemoterapi ( $X_3$ )	0 : tidak dilakukan 1 : dilakukan
Terapi Hormonal ( $X_4$ )	0 : tidak dilakukan 1 : dilakukan
Ukuran Tumor ( $X_5$ )	Ukuran tumor dalam satuan mm
Terapi Radiasi ( $X_6$ )	0 : tidak dilakukan 1 : dilakukan
PAM50 ( $X_7$ )	Basal Claudin-low Her2 LumA LumB NC Normal

Penelitian ini dilakukan menggunakan metode *stratified cox survival*. Adapun prosedur analisis yang dilakukan adalah sebagai berikut :

1. Melakukan *cleaning* data dan analisis deskriptif

*Cleaning* data dilakukan untuk mendapatkan data yang layak untuk dianalisis sehingga hasilnya tidak bias. Sedangkan analisis deskriptif digunakan untuk memberikan informasi umum dari data. Salah satu yang dapat dilakukan adalah melihat perbandingan antara nilai atau elemen dari variabel yang digunakan.

2. Membuat model regresi cox

Model regresi cox merupakan model regresi berbasis *proportional hazard*, di mana fungsi *baseline hazard* dimodelkan secara nonparametrik, sedangkan pengaruh variabel independen dimodelkan secara parametrik. Oleh karena itu, model ini juga dikenal dengan istilah *Cox proportional hazards model()*. Model cox memiliki bentuk umum sebagai berikut :

$$h(\mathbf{t}|\mathbf{X}) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p) \quad (1)$$

Keterangan :

$h(\mathbf{t}|\mathbf{X})$  : Hazard pada waktu  $t$  untuk individu dengan kovariat  $\mathbf{X}$ .

$h_0(t)$  : Baseline Hazard (ketika  $\mathbf{X} = 0$ )

$\beta_i$  : koefisien regresi untuk kovariat  $X_i$

3. Melakukan uji asumsi *proportional hazard*

Untuk melihat terpenuhi atau tidaknya asumsi *proportional hazard*, dapat dilakukan dengan dua cara. Yang pertama adalah menggunakan kurva kaplan-meier, dimana tidak terjadi pelanggaran *proportional hazard* terjadi ketika garis kurva tidak saling berpotongan atau sejajar. Yang kedua adalah menggunakan uji *goodness of fit*, salah satunya dengan metode residual schoenfeld.

Persamaan schoenfeld didefinisikan sebagai berikut :

$$r_{ij} = X_{ij} - \hat{X}_j(t_i) \quad (2)$$

Keterangan :

$X_{ij}$  : nilai kovariat j untuk individu ke-i

$\hat{X}_j(t_i)$ : nilai rata-rata kovariat j pada waktu  $t_i$  , dihitung sebagai berikut :

$$\hat{X}_j(t_i) = \frac{\sum_{k \in R(t_i)} X_{kj} \exp(\beta^T X_k)}{\sum_{k \in R(t_i)} \exp(\beta^T X_k)} \quad (3)$$

$R(t_i)$  : set individu yang masih beresiko pada waktu  $t_i$

Statistik uji yang digunakan adalah metode chi-square, asumsi *proportional hazard* dilanggar ketika  $p\text{-value} < \alpha$ , sebaliknya asumsi *proportional hazard* diterima.

4. Membuat model *Stratified cox survival*

Model *stratified cox survival* merupakan model lanjutan ketika terdapat variabel independen yang tidak memenuhi asumsi *proportional hazard*. Variabel independen yang memenuhi asumsi *proportional hazard* (PH) dimasukkan ke dalam model, sedangkan variabel independen yang dikelompokkan berdasarkan strata tidak dimasukkan ke dalam model(). Model ini dapat berupa model tanpa interaksi dan model dengan interaksi.

Model yang akan digunakan akan memiliki persamaan dan karakteristik sebagai berikut :

a. Model tanpa interaksi

Model tanpa interaksi merupakan model umum dari *stratified cox* dengan persamaan sebagai berikut :

$$h_g(t|X) = h_{0g}(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad (4)$$

Keterangan :

$g$  : 1,2,...,k\*

$h_{0g}(t)$ : *baseline* fungsi *hazard*

Model tanpa interaksi menunjukkan bahwa  $Z^*$  tidak berinteraksi dengan variabel independen lainnya.

b. Model dengan interaksi

Model dengan interaksi memiliki persamaan sebagai berikut :

$$h_g(t|X) = h_{0g}(t) \exp(\beta_{1g} X_1 + \beta_{2g} X_2 + \dots + \beta_{pg} X_p) \quad (5)$$

Keterangan :

$g$  : 1,2,...,k\*

$h_{0g}(t)$ : *baseline* fungsi *hazard*

Model yang akan digunakan ditentukan menggunakan uji interaksi pada model *stratified cox survival*. Uji interaksi dilakukan dengan uji *likelihood ratio* sebagai berikut :

$$LR = -2 \ln L_R - (-2 \ln L_F) \sim \chi^2_{p(k^*-1)} \quad (6)$$

Keterangan :

$R$  : model tanpa interaksi

$F$  : model dengan interaksi

$p$  : jumlah kategori pada variabel strata

$k^*$  : jumlah kategori variabel kovariat

Hipotesis :

$H_0$  = Model tanpa interaksi cocok untuk data

$H_1$  = Model dengan interaksi cocok untuk data

Dengan kriteria keputusan tolak  $H_0$  jika  $p\text{-value} < \alpha$ , lainnya gagal tolak  $H_0$ .

5. Uji signifikansi model

Terdapat dua pendekatan yang dapat digunakan dalam pengujian signifikansi model, yaitu secara parsial dan simultan yang digambarkan pada persamaan berikut:

a. Uji Parameter secara simultan atau *partial likelihood*

Statistik uji :

$$G = -2[\ln L(\hat{\Omega}) - L(\hat{\Omega})] \sim \chi^2_{\alpha, p} \quad (7)$$

Hipotesis :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$

$H_1 : \text{Minimal ada satu } \beta_j \neq 0, \text{ dimana } j = 1, 2, \dots, p$

Tolak  $H_0$  jika  $G_{hit}^2 > \chi^2_{\alpha, p}$  atau  $p\text{-value} < \alpha$

b. Uji parameter secara parsial

Hipotesis :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Statistik uji :

$$W^2 = \frac{\widehat{\beta}_j^2}{(SE(\widehat{\beta}_j))^2} \sim \chi_{\alpha,1}^2 \quad (8)$$

Tolak  $H_0$  jika  $W_{hit}^2 > \chi_{\alpha,1}^2$  atau  $p\text{-value} < \alpha$

### III. HASIL DAN PEMBAHASAN

#### A. Analisis Deskriptif dan Pembuatan Model Awal

Jumlah pasien kanker payudara pada dataset METABRIC adalah 1904 pasien, setelah dilakukan *cleaning* data, jumlah pasien menjadi 1868 pasien dengan jumlah pasien tersensor sebanyak 1080 pasien dan yang tidak tersensor sebanyak 788 pasien. Mayoritas pasien dalam data METABRIC menjalani jenis operasi *mastectomy* (59.73%) dibandingkan *breast conserving* (40.27%), dengan proporsi kejadian meninggal lebih tinggi pada pasien *mastectomy* (65.56%) dibandingkan *breast conserving* (46.27%). Sebagian besar pasien tidak menerima kemoterapi (87.95%) atau terapi hormonal (68.01%), namun pasien yang menerima terapi ini memiliki proporsi kejadian meninggal yang lebih rendah dibandingkan yang tidak menerima. Selain itu, pasien yang menerima terapi radiasi (55.86%) juga menunjukkan proporsi kejadian meninggal yang lebih rendah (53.13%) dibandingkan yang tidak menerima (64.92%). Berdasarkan subtipe PAM50, subtipe LumA dan LumB mendominasi populasi pasien, dengan LumA memiliki proporsi kejadian meninggal yang lebih rendah dibandingkan subtipe lainnya.

Model yang terbentuk berdasarkan data tersebut adalah sebagai berikut:

$$\hat{h}_g(t) = \hat{h}_{0g}(t) \exp(-0,006X_1 - 0,282X_2 + 0,638X_3 + 0,638X_4 - 0,005X_5 - 0,011X_6 + 0,009X_7(2) - 0,308X_7(3) - 0,123X_7(4) - 0,174X_7(5) - 0,673X_7(6) = 0,038X_7(7))$$

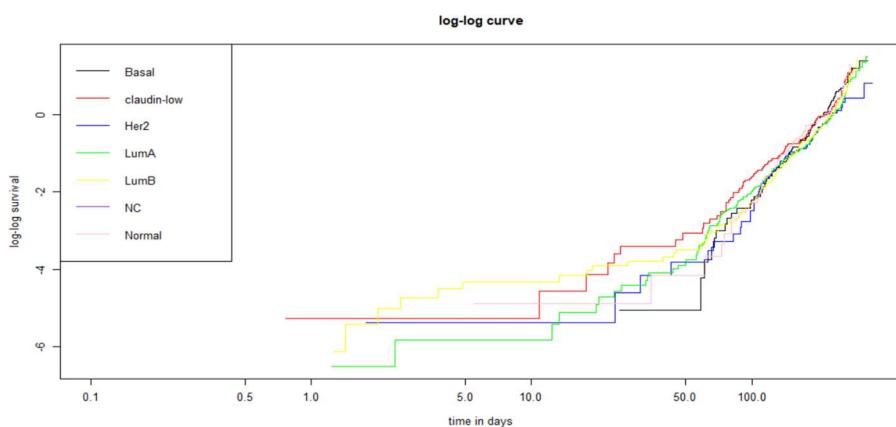
#### B. Pengujian Asumsi *Proportional Hazard*

Dilakukan uji asumsi *proportional hazard* dengan metode *goodness of fit*. Pada taraf signifikansi 0,1 diperoleh hasil uji asumsi *proportional hazard* sebagai berikut:

**Tabel 2.**Tabel Pengujian Asumsi Proportional Hazard

Variable	chisq	P(PH)	df	Keputusan
X <sub>1</sub>	0,813414	0,367	1	$H_0$ diterima
X <sub>2</sub>	0,141032	0,707	1	$H_0$ diterima
X <sub>3</sub>	0,208483	0,648	1	$H_0$ diterima
X <sub>4</sub>	2,041017	0,153	1	$H_0$ diterima
X <sub>5</sub>	0,791311	0,374	1	$H_0$ diterima
X <sub>6</sub>	1,644771	0,200	1	$H_0$ diterima
X <sub>7</sub> (2)	0,032763	0,856	1	$H_0$ diterima
X <sub>7</sub> (3)	0,353352	0,552	1	$H_0$ diterima
X <sub>7</sub> (4)	0,446161	0,504	1	$H_0$ diterima
X <sub>7</sub> (5)	3,499684	0,061	1	$H_0$ ditolak
X <sub>7</sub> (6)	1,062364	0,303	1	$H_0$ diterima

Pada taraf signifikansi 0,1 diperoleh kesimpulan bahwa asumsi *proportional hazard* pada semua variabel, satu diantaranya gagal memenuhi asumsi *proportional hazard* yakni pada NC yang merupakan bagian dari variabel PAM50, sedangkan variabel lainnya memenuhi asumsi *proportional hazard*. Hal ini menunjukkan secara statistik PAM50 tidak memenuhi asumsi *proportional hazard*. Untuk melihat lebih lanjut, diperoleh hasil kurva kaplan-meier untuk variabel PAM50 pada gambar 1. Dapat dilihat bahwa garis-garis pada kurva tidak sejajar dan saling berpotongan, sehingga diperoleh kesimpulan telah terjadi pelanggaran asumsi proportional hazard pada variabel PAM50.



**Gambar 1.** Kurva Kaplan-Meier variabel PAM50

### C. Pembentukan Model Cox Stratifikasi

Model *Cox proportional hazard* tidak dapat digunakan karena variabel PAM50 tidak memenuhi asumsi *proportional hazard*. Alternatif metode yang dapat digunakan salah satunya adalah model *stratified cox* dengan variabel tidak memenuhi asumsi *proportional hazard* akan distartifikasi. Terdapat dua model dari *stratified cox*, yaitu dengan interaksi dan tanpa interaksi.

Dalam pemodelan dengan interaksi, berikut merupakan hasil estimasi parameternya:

**Tabel 3.** Estimasi Parameter Model Dengan Interaksi

Variabel	Estimasi parameter	Variabel	Estimasi parameter
Umur saat terjangkit	0,010084	Tipe operasi x Lum B	0,251914
Tipe operasi	-0,54417	Tipe operasi x NC	-
Kemoterapi	0,485016	Tipe operasi x Normal	0,027269
Terapi hormonal	0,781555	Terapi hormonal x claudin low	-0,55839
Ukuran tumor	0,002162	Terapi hormonal x Her 2	-0,37994
Terapi radiasi	0,177026	Terapi hormonal x Lum A	-0,10841
Umur x claudin low	0,01075	Terapi hormonal x Lum B	-0,01171
Umur x Her 2	-0,03535	Terapi hormonal x NC	-
Umur x Lum A	-0,02585	Terapi hormonal x Normal	-0,14313
Umur x Lum B	-0,01268	Ukuran tumor x claudin low	-0,01513
Umur x NC	-	Ukuran tumor x Her 2	0,007223
Umur x Normal	-0,00643	Ukuran tumor x Lum A	-0,0129
Tipe operasi x claudin low	0,476127	Ukuran tumor x Lum B	-0,01003
Tipe operasi x Her 2	-0,43413	Ukuran tumor x NC	-
Tipe operasi x Lum A	0,226863	Ukuran tumor x Normal	-0,02017
Tipe operasi x Lum B	0,442584	Terapi radiasi x claudin low	-0,11314
Tipe operasi x NC	-	Terapi radiasi x Her 2	-0,7253
Tipe operasi x Normal	0,514943	Terapi radiasi x Lum A	-0,15506
Kemoterapi x claudin low	0,084681	Terapi radiasi x Lum B	-0,17207
Kemoterapi x Her 2	-0,1318	Terapi radiasi x NC	-
Kemoterapi x Lum A	0,601436	Terapi radiasi x Normal	-0,04547
<b>Likelihood ratio</b>		<b>158,4</b>	

Dalam pemodelan tanpa interaksi, berikut merupakan hasil estimasi parameternya:

**Tabel 4.** Estimasi Parameter Tanpa Interaksi

Variabel	Estimasi parameter
Umur saat terjangkit	-0,00561
Tipe operasi	-0,29809
Kemoterapi	0,641919

Variabel	Estimasi parameter
Terapi hormonal	0,631212
Ukuran tumor	-0,00565
Terapi radiasi	-0,01222
<i>Likelihood ratio</i>	<b>116,2</b>

#### D. Pemilihan Model Terbaik

Sebelum memilih model yang akan terbaik, perlu dilakukan uji interaksi pada model. Pengujian asumsi ini dilakukan untuk melihat apakah terdapat interaksi antara variabel PAM50 yang distratifikasi dengan variabel yang masuk ke dalam model meliputi umur saat terjangkit ( $X_1$ ), tipe operasi ( $X_2$ ), kemoterapi ( $X_3$ ), terapi hormonal ( $X_4$ ), ukuran tumor ( $X_5$ ), dan terapi radiasi ( $X_6$ ). Diperoleh hasil sebagai berikut:

**Tabel 5.**Tabel Uji Interaksi

LR	df	p-value	Keputusan
42,222	36	0,22	$H_0$ diterima

Karena hasil uji menunjukkan  $H_0$  diterima, sehingga dapat ditarik kesimpulan bahwa model yang terpilih adalah model tanpa interaksi. Model yang terbentuk adalah sebagai berikut :

$$\hat{h}_g(t) = \hat{h}_{0g}(t) \exp(-0,006X_1 - 0,298X_2 + 0,642X_3 + 0,631X_4 - 0,006X_5 - 0,012X_6)$$

#### E. Pengujian Signifikansi Model

Hasil uji parameter secara simultan adalah sebagai berikut:

$$G^2 = 116,2 \text{ atau } p\text{-value} = 2 \times 10^{-16}$$

Pada  $\alpha = 0,01$  dengan  $df = 5$

Keputusan yang diambil adalah tolak  $H_0$ , jadi dapat disimpulkan bahwa paling sedikit terdapat satu variabel independen yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara.

Hasil uji parsial adalah sebagai berikut:

**Tabel 6.** Tabel uji signifikansi parsial

Variabel	Estimasi parameter	p-value	keputusan
Umur saat terjangkit	-0,00561	0,11809	$H_0$ diterima
Tipe operasi	-0,29809	0,00147	$H_0$ ditolak
Kemoterapi	0,641919	$5,62 \times 10^{-9}$	$H_0$ ditolak
Terapi hormonal	0,631212	$4,34 \times 10^{-14}$	$H_0$ ditolak
Ukuran tumor	-0,00565	0,07983	$H_0$ diterima
Terapi radiasi	-0,01222	0,90319	$H_0$ diterima

Berdasarkan uji parsial, didapatkan variabel yang berpengaruh secara signifikan terhadap model adalah tipe operasi, kemoterapi, terapi hormonal, dan ukuran tumor pada  $\alpha = 0,05$ .

#### F. Interpretasi Model Cox Stratifikasi

Model cox stratifikasi yang terbentuk merupakan model tanpa interaksi, sehingga hanya memiliki satu nilai *hazard ratio*. Berikut merupakan nilai *hazard ratio* dari model yang terbentuk:

**Tabel 7.** Tabel Hazard Ratio

Variabel	Hazard Ratio
Tipe operasi	0,7404
Kemoterapi	2,0078
Terapi hormonal	1,8301

Berdasarkan tabel dapat dilihat *hazard ratio* untuk tipe operasi adalah 0,7404. Hal ini berarti untuk tipe operasi *breast conserving* memiliki peluang kematian  $1/0,7404=1,35$  kali lebih besar dibanding tipe operasi *mastectomy*. Untuk variabel kemoterapi nilai *hazard ratio* (HR) 2,0078 tidak berarti kemoterapi meningkatkan risiko kematian, tetapi menunjukkan bahwa pasien yang menerima kemoterapi umumnya memiliki kondisi kanker yang lebih berat. Dengan demikian, tingginya HR lebih mencerminkan derajat keparahan penyakit (confounding by indication), bukan efek

negatif dari terapi itu sendiri. Demikian pula, HR sebesar 1,8301 pada terapi hormonal menunjukkan bahwa pasien yang mendapatkan terapi hormona memiliki risiko dasar (baseline risk) lebih tinggi.

Penelitian sebelumnya yang menggunakan model *Stratified Cox* tanpa interaksi menunjukkan bahwa faktor-faktor yang mempengaruhi waktu kelangsungan hidup pasien dari kanker payudara meliputi ukuran tumor, dengan risiko kematian sebesar 16,6%. Pasien yang menjalani kemoterapi memiliki risiko kematian 0,682 kali lebih tinggi dibandingkan pasien yang tidak menjalani kemoterapi. Selain itu, pasien dengan kanker payudara stadium 1 memiliki risiko kematian 0,516 kali lebih tinggi dibandingkan pasien dengan kanker payudara stadium 4, sementara pasien stadium 2 memiliki risiko kematian 0,544 kali lebih tinggi dibandingkan pasien stadium 4. Pasien yang menjalani operasi *mastectomy* memiliki risiko kematian 0,637 kali lebih tinggi dibandingkan pasien yang menjalani operasi konservasi (Samosir & Di Asih, 2024). Hal ini jelas berbanding terbalik dengan hasil penelitian ini, namun data yang peneliti gunakan lebih besar sehingga dapat menghindari *overfitting* yang dapat terjadi.

Pada penelitian lain dengan metode serupa, tetapi mempertimbangkan 4 tipe stadium, ditemukan bahwa untuk pasien dengan kanker payudara stadium 1 dan 2, risiko kematian lebih tinggi pada mereka yang menjalani terapi radiasi dan kemoterapi dibandingkan pasien yang tidak menjalani kedua terapi tersebut. Sementara itu, untuk pasien stadium 3 dan 4, risiko kematian terkait dengan jenis operasi penanganan yang dilakukan (Abadi et al., 2014).

Berdasarkan kedua penelitian serupa, ditemukan bahwa kemoterapi, terapi radiasi, dan jenis operasi mempengaruhi waktu *survival* penderita kanker payudara. Sementara itu, hasil analisis kami menunjukkan bahwa tipe operasi, kemoterapi, dan hormon terapi mempengaruhi waktu *survival* pasien. Meskipun terdapat kesamaan dalam beberapa faktor yang mempengaruhi, terdapat perbedaan dalam prioritas pengaruh faktor tertentu, seperti tipe operasi dan ukuran tumor, yang perlu diperhatikan dalam pengobatan dan perencanaan pengobatan kanker payudara.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian mengenai pengujian metode regresi cox pada data penderita kanker payudara dari database METABRIC maka dapat ditarik kesimpulan pada penerapan data dengan menggunakan metode regresi *stratified cox* diperoleh model akhir *stratified cox* sebagai berikut

$$\hat{h}_g(t) = \hat{h}_{0g}(t) \exp(-0,298X_2 + 0,642X_3 + 0,631X_4)$$

Dari model tersebut, dapat disimpulkan bahwa faktor-faktor yang mempengaruhi waktu *survival* pasien kanker payudara adalah tipe operasi yang dilakukan ( $X_2$ ), dilakukan atau tidaknya kemoterapi ( $X_3$ ) dan terapi hormonal ( $X_4$ ). Disarankan untuk penelitian kedepannya menambahkan variabel lain seperti riwayat keluarga, gaya hidup, maupun status hormonal pasien untuk memberikan gambaran yang lebih komprehensif.

#### DAFTAR PUSTAKA

- Abadi, A., Yavari, P., Dehghani-Arani, M., Alavi-Majd, H., Ghasemi, E., Amanpour, F., & Bajdik, C. (2014). Cox models survival analysis based on breast cancer treatments. *Iranian journal of cancer prevention*, 7(3), 124.
- Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66, 15–23.
- Al Farisyi, M., & Khambri, D. (2018). Analisis Survival Pasien Kanker Payudara Usia Muda di RSUP DR. M. Djamil Padang Tahun 2008-2017. *Jurnal Kesehatan Andalas*, 7, 25-29.
- Cox, D.R., 1972. Regression Model and Life Table, Journal Royal Statistical Society, London
- Kleinbaum, D. G., & Klein, M. (1996). *Survival analysis a self-learning text*. Springer.
- Maajani, K., Jalali, A., Alipour, S., Khodadost, M., Tohidinik, H. R., & Yazdani, K. (2019). The global and regional survival rate of women with breast cancer: a systematic review and meta-analysis. *Clinical Breast Cancer*, 19(3), 165-177.

Maulida, E. A. (2019). Analisis Ketahanan Hidup Penderita Kanker Payudara menggunakan Regresi Cox Proportional Hazard dan Metode Kaplan Meier. *Tugas Akhir, Surabaya*.

Samosir, J. A., & Di Asih, I. M. (2024). PERBANDINGAN MODEL REGRESI STRATIFIED COX DAN EXTENDED COX PADA ANALISIS SURVIVAL PENDERITA KANKER PAYUDARA. *Jurnal Gaussian*, 13(1), 59-69.

World Health Organization. (2024). *Breast Cancer*. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (diakses pada 3 Januari 2025 pukul 21.07 WIB)