

## PERBANDINGAN MODEL REGRESI *STRATIFIED COX* DAN *EXTENDED COX* PADA ANALISIS *SURVIVAL* PENDERITA KANKER PAYUDARA

Jessika Aurora Samosir<sup>1</sup>, Sudarno<sup>2</sup>, Di Asih I Maruddani<sup>3</sup>

<sup>1,2,3</sup> Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

\*e-mail: [jejessicaauroras@gmail.com](mailto:jejessicaauroras@gmail.com)

DOI: 10.14710/j.gauss.13.1.59-69

### Article Info:

Received: 2023-06-05

Accepted: 2024-09-18

Available online: 2024-09-25

### Keywords:

*Breast cancer; stratified Cox; extended Cox; METABRIC.*

**Abstract:** Breast cancer is a tumor that develops in the breast where cells in the mammary gland divide and develop uncontrollably. Breast cancer is the most common cancer that causes death in women among other cancers. This study aims to determine the factors that affect the survival of breast cancer patients from the METABRIC database. This study was analyzed using survival analysis method. The method that is often used is the Cox proportional hazard model where the proportional hazard assumption must be met. There is variable that do not meet the assumption so that the methods used are stratified Cox and extended Cox. The stratified Cox model overcomes variables that do not meet the assumption by stratifying variables that do not meet the assumption. The extended Cox model overcomes variables that do not meet the assumption by interacting the variables with a time function. The time functions used in this study are linear time functions and logarithmic time functions. Based on the smallest AIC value, the best model is the stratified Cox regression model without interaction. Factors that affect the survival of breast cancer patients from the METABRIC database are tumor size, chemotherapy, stage 1, stage 2, and type of surgery.

## 1. PENDAHULUAN

Kanker payudara adalah kanker yang terbentuk di jaringan payudara yang terjadi karena sel pada payudara tumbuh tidak terkendali. Kanker payudara merupakan penyakit kanker yang paling banyak menyebabkan kematian pada wanita diantara penyakit kanker lainnya. Banyaknya kasus kanker payudara sendiri terhitung 25% dari semua penyakit kanker lainnya dan kematian yang disebabkan kanker payudara terhitung 15% dari semua kematian yang disebabkan oleh kanker pada wanita. Terhitung sekitar 50% dari kasus kanker payudara dan 38% kematian yang disebabkan oleh kanker tersebut ditemukan pada negara berkembang (Torre *et al.*, 2012). Pada tahun 2020 banyaknya kasus kanker payudara sendiri telah melampaui kasus kanker paru-paru dengan perkiraan 2,3 juta kasus baru. Meskipun kebanyakan kanker payudara disebabkan oleh faktor keturunan dan faktor genetik, terhitung sekitar 5% sampai 10% kasus kanker payudara bukan disebabkan oleh faktor keturunan maupun faktor genetik. (Sung *et al.*, 2020).

Berdasarkan informasi tentang penyakit kanker payudara tersebut dilakukan analisis ketahanan hidup terhadap penderita kanker payudara. Pada analisis ketahanan hidup, variabel yang dilihat adalah waktu hingga *peristiwa* terjadi. Pada analisis ketahanan hidup, ada banyak model yang dapat digunakan. Salah satu model yang paling umum digunakan dalam penelitian analisis *survival* adalah regresi Cox *proportional hazard*. Model ini digunakan untuk mengetahui hubungan antara variabel independen dan variabel dependen, yaitu waktu *survival*, yang dapat dihitung melalui fungsi *hazardnya*. Pada model Cox *proportional hazard*, asumsi *proportional hazard* harus terpenuhi yang merupakan suatu kondisi ketika nilai *hazard ratio* konstan terhadap waktu (Kleinbaum & Klein, 2012). Jika

asumsi ini tidak terpenuhi, maka model *stratified* Cox atau model *extended* Cox dapat digunakan untuk mengatasi variabel independen yang tidak memenuhi asumsi.

Faktor-faktor yang mempengaruhi lama penderita stroke hemoragik *unspecified* dirawat pernah diteliti oleh Pahlevi (2016) menggunakan metode *stratified* Cox. Faktor-faktor yang berpengaruh adalah umur dan kadar kolesterol. Selain itu, pada tahun 2017 Nurfain melakukan penelitian terhadap penderita kusta dengan metode *extended* Cox. Adapun faktor-faktor yang mempengaruhi laju perbaikan klinis adalah keteraturan berobat dan tipe kusta.

Agustina (2019) menggunakan model Cox *propotional hazard* untuk meneliti faktor-faktor yang mempengaruhi ketahanan hidup penderita kanker payudara dan menyimpulkan bahwa faktor-faktor tersebut adalah usia awal terjangkit kanker payudara, stadium klinik, ukuran tumor. Sedangkan Sinaga *et al.* (2017) melakukan penelitian menggunakan metode Cox *proportional hazard* menyimpulkan bahwa ketahanan hidup penderita kanker payudara dipengaruhi oleh faktor umur saat didiagnosis.

*The Molecular Taxonomy of Breast Cancer International Consortium* (METABRIC) adalah kelompok studi yang dipimpin oleh Professor Carlos Cadas dari Institut Penelitian *Cambridge* dan Professor Sam Aparico dari Pusat Kanker *British Columbia* di Kanada. Para peneliti di METABRIC menyatakan kanker payudara bisa menjadi paling tidak sepuluh penyakit yang berbeda karena setiap penyakitnya memiliki sidik jari molekuler sendiri dan memiliki kelemahannya sendiri (Scrowcroft, 2012).

## 2. TINJAUAN PUSTAKA

Asumsi *proportional hazard* harus terpenuhi pada model Cox *proportional hazard* yang berarti nilai *hazard ratio*-nya bersifat konstan terhadap waktu. Pengujian dilakukan dengan metode *goodness of fit* dengan menggunakan residual Schoenfeld

Hipotesis:

$H_0 : \rho = 0$  (Asumsi *proportional hazard* terpenuhi)

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

Statistik Uji:

$$t_{hit} = \frac{r_{RT,PR_k} \sqrt{r-2}}{\sqrt{1-r_{RT,PR_k}^2}} \quad (1)$$

Tolak  $H_0$  jika  $|t_{hitung}| > t_{\frac{\alpha}{2}, n-2}$  atau p-value  $< \alpha$

Persamaan model Cox *proportional hazard* dapat ditulis sebagai berikut

$$h(t, \mathbf{X}) = h_0(t) \exp \sum_{k=1}^p \beta_k X_k ; \mathbf{X} = (X_1, X_2, \dots, X_p) \quad (2)$$

Variabel independen yang tidak memenuhi asumsi *proportional hazard* akan distratifikasikan dan menjadi variabel stratifikasi  $Z^*$  yang tidak masuk dalam model. Untuk menguji interaksi digunakan uji *likelihood ratio*

Hipotesis:

$H_0$  : variabel stratifikasi  $Z^*$  tidak berinteraksi dengan variabel independen lainnya

$H_1$  : variabel stratifikasi  $Z^*$  berinteraksi dengan variabel independen lainnya

Statistik uji

$$LR = -2 \ln L_R - (-2 \ln L_F) \sim \chi_{p(k^*-1)}^2 \quad (3)$$

Tolak  $H_0$  jika  $LR > \chi_{p(k^*-1)}^2$  atau p-value  $< \alpha$ .

1. Model *Stratified Cox* tanpa Interaksi

Model *stratified Cox* tanpa interaksi menunjukkan bahwa variabel stratifikasi  $Z^*$  tidak berinteraksi dengan variabel independent lainnya yang memiliki persamaan sebagai berikut

$$h_g(t, x) = h_{og}(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad (4)$$

Dengan

$$g = 1, 2, \dots, k^*$$

2. Model *Stratified Cox* dengan Interaksi

Model *stratified Cox* dengan interaksi menunjukkan bahwa variabel stratifikasi  $Z^*$  berinteraksi dengan variabel bebas lainnya dengan persamaan sebagai berikut

$$h_g(t, X) = h_{og}(t) \exp(\beta_{1g} x_1 + \beta_{2g} x_2 + \dots + \beta_{pg} x_p) \quad (5)$$

Dengan  $g = 1, 2, \dots, k^*$ .

Alternatif persamaan *stratified Cox* dengan interaksi lainnya dapat dituliskan sebagai berikut

$$h_g(t, X) = h_{og}(t) \exp[\beta_1 x_1 + \dots + \beta_p x_p + \beta_{11}(Z_1^* x_1) + \dots + \beta_{p1}(Z_1^* x_p) + \beta_{12}(Z_2^* x_1) + \dots + \beta_{p2}(Z_2^* x_p) + \dots + \beta_{1, k^*-1}(Z_{k^*-1}^* x_1) + \dots + \beta_{p, k^*-1}(Z_{k^*-1}^* x_p)] \quad (6)$$

dengan  $g = 1, 2, \dots, k^*$

a. Estimasi Parameter

Estimasi parameter pada model *stratified Cox* ini menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE). Pada model *stratified Cox* fungsi partial likelihoodnya adalah hasil perkalian fungsi *partial likelihood* dari setiap kategori (Kleinbaum & Klein, 2012).

$$L(\beta) = \prod_{g=1}^{k^*} L_g(\beta) \quad (7)$$

$$L(\beta) = \prod_{g=1}^{k^*} \left[ \prod_{j=1}^{r_g} \frac{\exp(\beta' x_{gj})}{\sum_{l \in R(t_{gj})} \exp(\beta' x_{gl})} \right] \quad (8)$$

$$\begin{aligned} \ln L(\beta) &= \sum_{g=1}^{k^*} \left[ \sum_{j=1}^{r_g} (\beta' x_{gj}) - \sum_{j=1}^{r_g} \ln \left( \sum_{l \in R(t_{gj})} \exp(\beta' x_{gl}) \right) \right] \\ &= \sum_{g=1}^{k^*} \left[ \sum_{j=1}^{r_g} \sum_{i=1}^p (\beta_i x_{gji}) - \sum_{j=1}^{r_g} \ln \left( \sum_{l \in R(t_{gj})} \exp \left( \sum_{i=1}^p (\beta_i x_{gli}) \right) \right) \right] \\ &= \sum_{g=1}^{k^*} \left[ \sum_{j=1}^{r_g} \left( \sum_{i=1}^p (\beta_i x_{gji}) - \ln \left( \sum_{l \in R(t_{gj})} \exp \left( \sum_{i=1}^p (\beta_i x_{gli}) \right) \right) \right) \right] \end{aligned} \quad (9)$$

Berdasarkan persamaan (11), maka didapatkan turunan pertama dan kedua seperti persamaan (12) dan (13).

$$\frac{\partial \ln(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_p} = \sum_{g=1}^{k^*} \left[ \sum_{j=1}^{r_g} \left( x_{gjp} - \frac{\sum_{l \in R(t_{gj})} x_{glp} \exp(\sum_{i=1}^p (\beta_i x_{gli}))}{\sum_{l \in R(t_{gj})} \exp(\sum_{i=1}^p (\beta_i x_{gli}))} \right) \right] \quad (10)$$

$$\frac{\partial^2 \ln L(\beta)}{\partial \beta_z \partial \beta_{z^*}} = \sum_{g=1}^{k^*} \left[ \sum_{j=1}^{r_g} \left( - \frac{\sum_{l \in R(t_{gj})} x_{glz} x_{glz^*} \exp(\sum_{i=1}^p (\beta_i x_{gli}))}{\sum_{l \in R(t_{gj})} \exp(\sum_{i=1}^p (\beta_i x_{gli}))} \right) \right] + \sum_{g=1}^{k^*} \left[ \sum_{j=1}^{r_g} \left( \frac{\sum_{l \in R(t_{gj})} x_{glz} \exp(\sum_{i=1}^p (\beta_i x_{gli})) \sum_{l \in R(t_{gj})} x_{glz^*} \exp(\sum_{i=1}^p (\beta_i x_{gli}))}{\sum_{l \in R(t_{gj})} \exp(\sum_{i=1}^p (\beta_i x_{gli}))^2} \right) \right] \quad (11)$$

dengan  $z, z^* = 1, 2, \dots, p$

Karena turunan pertama implisit, maka digunakan interasi numerik dengan menggunakan metode Newton Raphson.

b. *Hazard Ratio*

Pada model *stratified Cox* nilai *hazard ratio* merupakan *hazard* untuk individu kategori satu dibagi dengan *hazard* untuk individu yang berbeda, seperti persamaan berikut (Kleinbaum & Klein, 2012).

$$\widehat{HR} = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})} = \frac{\hat{h}_0(t) e^{\sum_{i=1}^p \hat{\beta}_i x_i^*}}{\hat{h}_0(t) e^{\sum_{i=1}^p \hat{\beta}_i x_i}} = e^{\sum_{i=1}^p \hat{\beta}_i (x_i^* - x_i)} \quad (12)$$

c. Pengujian Signifikansi Parameter

Pengujian signifikansi parameter dilakukan secara serentak dan parsial (Hosmer & Lemeshow, 1997).

1. Uji Serentak

Hipotesis:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_j = \dots = \beta_p = 0$  (secara bersama-sama variabel bebas tidak berpengaruh terhadap variabel terikat)

$H_1 : \text{minimal ada satu } \beta_j \neq 0; j = 1, 2, \dots, p$  (minimal ada satu variabel bebas yang berpengaruh terhadap variabel terikat)

Statistik uji

Uji rasio likelihood:

$$G = -2 [\ln L(\hat{\omega}) - L(\hat{\Omega})] \sim \chi_{\alpha, p}^2 \quad (13)$$

Kriteria Penolakan

Tolak  $H_0$  jika  $G_{hit}^2 > \chi_{\alpha, p}^2$

2. Uji Parsial

Hipotesis

$H_0 : \beta_j = 0$  (variabel bebas  $j$  tidak berpengaruh terhadap waktu ketahanan hidup)

$H_1 : \beta_j \neq 0$ , dengan  $j=1, 2, \dots, p$  (variabel bebas  $j$  berpengaruh terhadap waktu ketahanan hidup)

Statistik Uji

$$W^2 = \frac{\hat{\beta}_j^2}{(SE(\hat{\beta}_j))^2} \sim \chi_{\alpha, 1}^2 \quad (14)$$

Kriteria Penolakan

Tolak  $H_0$  jika  $W_{hit}^2 > \chi_{\alpha, 1}^2$

Model *extended Cox* merupakan model *Cox proportional hazard* yang diperluas karena memiliki variabel yang bergantung terhadap waktu. Bentuk umum model *extended Cox* adalah sebagai berikut.

$$h(t, x(t)) = h_o(t) \exp \left[ \sum_{k=1}^p \beta_k x_k + \sum_{m=1}^q \delta_m x_m g_m(t) \right] \quad (15)$$

Hazard ratio untuk model *extended Cox* memiliki rumus seperti pada persamaan 18

$$\begin{aligned} \widehat{HR}(t) &= \frac{\hat{h}(t, \mathbf{X}^*(t))}{\hat{h}(t, \mathbf{X}(t))} \\ &= \exp \left[ \sum_{k=1}^p \hat{\beta}_k [X_k^* - X_k] + \sum_{m=1}^q \delta_m [X_m^*(t) - X_m(t)] \right] \end{aligned} \quad (16)$$

Pada model *extended Cox* digunakan persamaan *maximum partial likelihood estimation* untuk menaksir parameter (Collet, 2003) sebagai berikut

$$\ln L(\beta) = \sum_{i=1}^n d_i \left\{ \beta' x_i(t_i) - \ln \sum_{l \in R(t_i)} \exp(\beta' x_l(t_i)) \right\} \quad (17)$$

Selanjutnya persamaan di atas dimaksimumkan dengan metode Newton Raphson untuk mendapatkan taksiran parameter model *extended Cox*. Sehingga dapat mengestimasi model *extended Cox* pada nilai *hazard function* yaitu sebagai berikut

$$\hat{h}_0(t) = -\ln \hat{S}_0(t) = \sum_{j=1}^n \frac{d_j}{\sum_{f \in R(t_j)} \exp(\hat{\beta}' x_f(t))} \quad (18)$$

Nilai *Akaike's Information Criterion* (AIC) dapat digunakan sebagai pembanding sejumlah model yang diperoleh untuk memperoleh model terbaik. Model yang memiliki nilai AIC terkecil adalah model terbaik (Collet, 2003). Nilai AIC dapat diperoleh dari

$$AIC = -2 \log \hat{L} + 2q \quad (19)$$

Salah satu jenis kanker yang paling banyak diderita wanita di Indonesia adalah kanker payudara yang merupakan tumor ganas pada jaringan payudara yang berasal dari epitel lobulus dan duktus. Berdasarkan badan registasi kanker perhimpunan dokter spesialis patologi Indonesia, kanker payudara memiliki frekuensi relatif sebesar 18,6% pada tahun 2010. Pada tahun 2020 kanker payudara masih menempati urutan pertama sebagai kasus baru terbanyak di Indonesia dengan frekuensi sebesar 16,6% (Sung *et al.*, 2020). Jenis kelamin wanita, riwayat menopause yang terlambat yaitu lebih dari 50 tahun, riwayat genetik pada keluarga, usia lebih dari 50 tahun, riwayat kanker payudara pada keluarga, kelebihan berat badan atau obesitas, konsumsi alkohol, pemakaian alat kontrasepsi hormonal, riwayat menstruasi dini yaitu kurang dari 12 tahun, riwayat reproduksi, dan faktor lingkungan merupakan faktor risiko yang berpengaruh dalam peningkatan insiden kanker payudara (Komite Nasional Penanggulangan Kanker, 2015).

### 3. METODE PENELITIAN

Jenis data yang digunakan pada penelitian ini adalah data sekunder diperoleh secara tidak langsung atau melalui media perantara. Data yang digunakan adalah data penderita kanker payudara dari *database* METABRIC yang diperoleh dari website [www.cbioportal.org](http://www.cbioportal.org). Variabel penelitian yang digunakan terdiri atas variabel dependen dan variabel independen. Variabel dependen yang digunakan adalah waktu *survival* (T) dan status penderita (d) sedangkan variabel independent yang digunakan adalah usia (X<sub>1</sub>), ukuran tumor (X<sub>2</sub>), tipe kanker (X<sub>3</sub>), kemoterapi (X<sub>4</sub>), stadium (X<sub>5</sub>), jenis operasi (X<sub>6</sub>), menopause (X<sub>7</sub>), dan tumor utama (X<sub>8</sub>).

Penelitian Tugas Akhir ini data dianalisis menggunakan metode *stratified Cox* dan *extended Cox*. Langkah-langkah analisis yang dilakukan adalah sebagai berikut:

1. Melakukan analisis deskriptif
2. Melakukan uji asumsi *proportional hazard* dengan uji *goodness of fit*.
3. Membuat model *stratified Cox* pada data ketahanan hidup penderita kanker payudara.
4. Membuat model *extended Cox* pada data ketahanan hidup penderita kanker payudara.
5. Menentukan model terbaik untuk memodelkan ketahanan hidup penderita kanker payudara dengan kriteria AIC terkecil

#### 4. HASIL DAN PEMBAHASAN

Jumlah penderita kanker payudara dari *database* METABRIC pada penelitian ini adalah 551 dengan 382 penderita merupakan data tidak tersensor dan sisanya 169 penderita merupakan data tersensor. Penderita kanker payudara yang terdaftar dalam *database* METABRIC memiliki usia rata-rata sekitar 59 tahun. Penderita kanker payudara yang terdaftar dalam *database* METABRIC memiliki ukuran tumor rata-rata 25,9010889 mm. Tipe kanker yang paling banyak adalah karsinoma duktal invasif yaitu sebanyak 440 penderita dengan 300 penderita tersensor dan 140 penderita tidak tersensor. Persentase kemoterapi yang paling besar adalah tidak melakukan kemoterapi yaitu sebanyak 383 penderita dengan 286 penderita tersensor dan 97 penderita tidak tersensor. Persentase stadium yang paling besar adalah stadium 2 yaitu sebanyak 304 penderita dengan 215 penderita tersensor dan 89 penderita tidak tersensor. Persentase jenis operasi paling besar adalah jenis operasi mastektomi yaitu sebanyak 213 penderita dengan 168 penderita tersensor dan 45 penderita tidak tersensor. Persentase menopause paling besar adalah penderita kanker payudara yang telah mengalami menopause yaitu sebanyak 415 penderita dengan 287 penderita tersensor dan 128 penderita tidak tersensor. Persentase tumor utama paling besar adalah tumor utama yang terdapat di daerah payudara kiri yaitu sebanyak 280 penderita dengan 187 penderita tersensor dan 93 penderita tidak tersensor.

Dilakukan uji asumsi *proportional hazard* sebelum membuat model dengan metode *goodness of fit*. Pada taraf signifikansi 0,05 diperoleh kesimpulan bahwa asumsi *proportional hazard* tidak terpenuhi karena variabel tipe kanker 1 dan tipe kanker 2 ( $X_3$ ) memiliki  $p\text{-value} < \alpha = 0,05$ . Variabel tipe kanker 1 dan tipe kanker 2 merupakan kategori dari variabel tipe kanker sehingga variabel tipe kanker menolak  $H_0$ . Sedangkan variabel usia, ukuran tumor, kemoterapi, jenis operasi, menopause, dan tumor utama memiliki  $p\text{-value} > \alpha = 0,05$  sehingga menerima  $H_0$ .

Variabel tipe kanker tidak memenuhi asumsi sehingga tidak bisa menggunakan model *Cox proportional hazard*. Alternatif metode yang dapat digunakan salah satunya adalah model *stratified Cox* dengan variabel yang tidak memenuhi asumsi *proportional hazard* akan distratifikasi. Variabel independen yang tidak memenuhi asumsi tidak masuk ke dalam model dan menjadi variabel stratifikasi  $Z^*$  sebagai berikut:

$Z^*_1$  = Tipe Kanker 1 (Karsinoma Duktal Invasif)

$Z^*_2$  = Tipe Kanker 2 (Karsinoma Gabungan Duktal dan Lobular)

$Z^*_3$  = Tipe Kanker 3 (Karsinoma Lobular Invasif)

$Z^*_4$  = Tipe Kanker 4 (Karsinoma Musinus)

Variabel yang memenuhi asumsi *proportional hazard* dimasukkan ke dalam model yaitu usia ( $X_1$ ), ukuran tumor ( $X_2$ ), kemoterapi ( $X_4$ ), stadium ( $X_5$ ), jenis operasi ( $X_6$ ), menopause ( $X_7$ ), dan tumor utama ( $X_8$ ).

Langkah pertama yang dilakukan sebelum membuat model *stratified Cox* adalah melakukan uji interaksi untuk mengetahui apakah terdapat interaksi antara variabel yang telah distratifikasi dengan variabel bebas lainnya. Setelah dilakukan uji interaksi diperoleh kesimpulan bahwa tidak terdapat interaksi antara variabel yang telah distratifikasi dengan variabel independen yang masuk dalam model karena  $LR = 26,633$  lebih kecil dari



$\chi_{0,05;21}^2=32,671$  atau  $p\text{-value} = 0,18333$  lebih besar dari  $\alpha = 0,05$ . Setelah disimpulkan tidak terdapat interaksi, langkah selanjutnya adalah membuat model. Diperoleh model awal sebagai berikut:

$$\hat{h}_g(t) = \hat{h}_{0g}(t) \exp (0,1677X_1 + 0,1436X_2 - 0,6121X_4 - 1,3276X_5(1) - 1,3751X_5(2) - 0,9175X_5(3) - 0,3979X_6 - 0,0007X_7 + 0,1264X_8)$$

Setelah model *stratified* Cox tanpa interaksi diperoleh, maka dilakukan uji signifikansi parameter secara serentak dan parsial. Dari model yang terbentuk dilakukan uji serentak untuk mengetahui kesesuaian model sehingga diperoleh hasil bahwa paling sedikit ada satu variabel independen yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara karena  $G = 58,1075$  lebih besar dari  $\chi_{\alpha,p}^2=16,91898$  atau  $p\text{-value} = <0,0001$  lebih kecil dari  $\alpha = 0,05$ . Untuk mengetahui variabel independen yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara maka dilakukan uji secara parsial.

Tabel 1. Uji Parsial Model Awal *Stratified* Cox tanpa Interaksi

Variabel	Parameter Estimasi	Standard Error	p-value	Keputusan
Usia (X <sub>1</sub> )	0,1677	0,1361	0,2177	H <sub>0</sub> diterima
Ukuran (X <sub>2</sub> )	0,1436	0,0591	0,0152	H <sub>0</sub> ditolak
Kemoterapi (X <sub>4</sub> )	-0,6121	0,2178	0,0050	H <sub>0</sub> ditolak
Stadium (1) (X <sub>5</sub> )	-1,3276	0,4701	0,0047	H <sub>0</sub> ditolak
Stadium (2) (X <sub>5</sub> )	-1,3751	0,4429	0,0019	H <sub>0</sub> ditolak
Stadium (3) (X <sub>5</sub> )	-0,9175	0,4845	0,0583	H <sub>0</sub> diterima
Jenis Operasi (X <sub>6</sub> )	-0,3979	0,1844	0,0310	H <sub>0</sub> ditolak
Menopause (X <sub>7</sub> )	-0,0007	0,2638	0,9980	H <sub>0</sub> diterima
Tumor Utama (X <sub>8</sub> )	0,1264	0,1566	0,4196	H <sub>0</sub> diterima

Pada uji secara parsial diperoleh kesimpulan bahwa variabel independen yang signifikan terhadap model adalah variabel ukuran (X<sub>2</sub>), kemoterapi (X<sub>4</sub>), stadium 1(X<sub>5</sub> (1)), stadium 2(X<sub>5</sub> (2)), dan jenis operasi (X<sub>6</sub>). Variabel yang tidak signifikan antara lain usia (X<sub>1</sub>), stadium 3 (X<sub>5</sub> (3)), menopause (X<sub>7</sub>), dan tumor utama (X<sub>8</sub>) dikeluarkan dari model. Model kedua *stratified* Cox tanpa interaksi yang terbentuk adalah sebagai berikut:

$$\hat{h}_g(t) = \hat{h}_{0g}(t) \exp (0,1539X_2 - 0,3828X_4 - 0,6618X_5(1) - 0,6090X_5(2) - 0,4505X_6)$$

Setelah model kedua *stratified* Cox tanpa interaksi didapatkan, maka dilakukan uji signifikansi parameter dari model tersebut untuk mengetahui kesesuaian model sehingga diperoleh hasil bahwa paling sedikit ada satu variabel independen dari persamaan yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara karena  $G= 50,6624$  lebih besar dari  $\chi_{(5\%,5)}^2= 11,070$  atau  $p\text{-value} = <0,0001$  lebih kecil dari  $\alpha = 0,05$ . Untuk mengetahui variabel independen yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara maka dilakukan uji secara parsial.

Tabel 2. Uji Parsial Model Kedua *Stratified* Cox tanpa Interaksi

Variabel	Parameter Estimasi	Standard Error	p-value	Keputusan
Ukuran (X <sub>2</sub> )	0,1539	0,0568	0,0067	H <sub>0</sub> ditolak
Kemoterapi (X <sub>4</sub> )	-0,3828	0,1836	0,0371	H <sub>0</sub> ditolak
Stadium 1 (X <sub>5</sub> (1))	-0,6618	0,2847	0,0201	H <sub>0</sub> ditolak
Stadium 2 (X <sub>5</sub> (2))	-0,6090	0,2366	0,0100	H <sub>0</sub> ditolak
Jenis Operasi (X <sub>6</sub> )	-0,4505	0,1807	0,0127	H <sub>0</sub> ditolak

Pada uji secara parsial diperoleh bahwa variabel ukuran ( $X_2$ ), kemoterapi ( $X_4$ ), stadium 1 ( $X_5(1)$ ), stadium 2 ( $X_5(2)$ ) dan jenis operasi ( $X_6$ ) signifikan terhadap model.

Metode lain yang dapat digunakan ketika asumsi *proportional hazard* tidak terpenuhi adalah metode *extended Cox* yaitu dengan menginteraksikan variabel yang tidak memenuhi asumsi *proportional hazard* dengan fungsi waktu. Model *extended Cox* dengan fungsi waktu linier  $g_m(t) = t$  adalah sebagai berikut

$$\begin{aligned} \hat{h}(t, x(t)) = & \hat{h}_0(t) \exp(0,1844X_1 + 0,1463X_2 + 1,2382X_3(1) + 0,1195X_3(2) \\ & + 0,4114X_3(3) - 0,6027X_4 - 1,3358X_5(1) - 1,3751X_5(2) \\ & - 0,9229X_5(3) - 0,397X_6 + 0,0284X_7 + 0,1219X_8 - 0,0001X_3(1)t \\ & + 0,0003X_3(2)t + 0,0002X_3(3)t) \end{aligned}$$

Setelah model *extended Cox* dengan fungsi waktu linier didapatkan, maka selanjutnya akan dilakukan uji signifikansi parameter terhadap model tersebut. Setelah dilakukan uji secara serentak diperoleh bahwa paling sedikit terdapat satu variabel independen dari persamaan yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara karena  $G = 73,3117$  lebih besar dari  $\chi^2_{\alpha,p} = 24,996$  atau  $p\text{-value} = <0,0001$  lebih kecil dari  $\alpha = 0,05$ . Untuk mengetahui variabel bebas yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara maka dilakukan uji secara parsial.

Tabel 3. Uji Parsial Model Awal *Extended Cox* Fungsi Waktu Linier

Variabel	Parameter Estimasi	Standard Error	p-value	Keputusan
Umur ( $X_1$ )	0,1844	0,1365	0,1766	$H_0$ diterima
Ukuran ( $X_2$ )	0,1463	0,0590	0,0132	$H_0$ diterima
Tipe Kanker (1) ( $X_3(1)$ )	1,2382	1,6662	0,4574	$H_0$ diterima
Tipe Kanker (2) ( $X_3(2)$ )	0,1195	1,7150	0,9445	$H_0$ diterima
Tipe Kanker (3) ( $X_3(3)$ )	0,4114	1,7431	0,8134	$H_0$ diterima
Kemoterapi ( $X_4$ )	-0,6027	0,2178	0,0057	$H_0$ ditolak
Stadium (1) ( $X_5(1)$ )	-1,3358	0,4668	0,0042	$H_0$ diterima
Stadium (2) ( $X_5(2)$ )	-1,3751	0,4402	0,0018	$H_0$ ditolak
Stadium (3) ( $X_5(3)$ )	-0,9229	0,4829	0,0560	$H_0$ diterima
Jenis Operasi ( $X_6$ )	-0,3972	0,1843	0,0312	$H_0$ ditolak
Menopause ( $X_7$ )	0,0284	0,2647	0,9145	$H_0$ diterima
Tumor Utama ( $X_8$ )	0,1219	0,1568	0,4372	$H_0$ diterima
Tipe Kanker 1 x t	-0,0001	0,0005	0,8684	$H_0$ diterima
Tipe Kanker 2 x t	0,0003	0,0005	0,5157	$H_0$ diterima
Tipe Kanker 3 x t	0,0002	0,0005	0,7567	$H_0$ diterima

Setelah dilakukan uji secara parsial dapat disimpulkan bahwa variabel ukuran ( $X_2$ ), kemoterapi ( $X_4$ ), stadium 1 ( $X_5(1)$ ), stadium 2 ( $X_5(2)$ ) dan jenis operasi ( $X_6$ ) secara parsial signifikan terhadap model. Variabel yang tidak signifikan terhadap model akan dikeluarkan dari model. Model kedua *extended Cox* dengan fungsi waktu linier yang terbentuk adalah sebagai berikut

$$\begin{aligned} \hat{h}(t, x(t)) = & \hat{h}_0 \exp(0,1520X_2 - 0,4307X_4 - 0,6413X_5(1) - 0,5758X_5(2) \\ & - 0,4587X_6) \end{aligned}$$

Setelah model kedua *extended Cox* dengan fungsi waktu linier didapatkan, maka akan dilakukan uji signifikansi parameter terhadap model tersebut. Pada uji secara serentak dapat disimpulkan bahwa paling sedikit ada satu variabel independen dari persamaan yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara karena  $G = 73,3117$  lebih besar dari  $\chi^2_{(5\%|5)} = 11,070$  atau  $p\text{-value} = <0,0001$  lebih kecil dari



$\alpha = 0,05$ . Untuk mengetahui variabel bebas yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara maka dilakukan uji secara parsial.

Tabel 4. Uji Parsial Model Kedua *Extended Cox Fungsi Waktu Linier*

Variabel	Parameter Estimasi	Standard Error	p-value	Keputusan
Ukuran (X <sub>2</sub> )	0,1520	0,0575	0,0082	H <sub>0</sub> ditolak
Kemoterapi (X <sub>4</sub> )	-0,4307	0,1798	0,0166	H <sub>0</sub> ditolak
Stadium (X <sub>5</sub> (1))	-0,6413	0,2851	0,0245	H <sub>0</sub> ditolak
Stadium (X <sub>5</sub> (2))	-0,5758	0,2363	0,0148	H <sub>0</sub> ditolak
Jenis Operasi (X <sub>6</sub> )	-0,4587	0,1799	0,0108	H <sub>0</sub> ditolak

Pada uji secara parsial diperoleh bahwa variabel ukuran (X<sub>2</sub>), kemoterapi (X<sub>4</sub>), stadium 1 (X<sub>5</sub> (1)), stadium 2 (X<sub>5</sub> (2)) dan jenis operasi (X<sub>6</sub>) signifikan terhadap model.

Alternatif lain dari metode *extended Cox* adalah dengan menginteraksikan fungsi waktu logaritma  $g_m(t) = \ln t$  dengan variabel yang tidak memenuhi asumsi *proportional hazard*. Model *extended Cox* dengan fungsi waktu logaritma adalah sebagai berikut:

$$\begin{aligned} \hat{h}(t, x(t)) = & \hat{h}_0(t) \exp (0,1878X_1 + 0,1438X_2 + 5,7777X_3(1) - 1,9699X_3(2) \\ & + 2,9043X_3(3) - 0,6111X_4 - 1,2990X_5(1) - 1,3460X_5(2) \\ & - 0,8912X_5(3) - 0,4054X_6 + 0,0248X_7 + 0,1292X_8 - 0,6165X_3(1) \ln t \\ & + 0,3985X_3(2) \ln t - 0,2643X_3(3) \ln t) \end{aligned}$$

Setelah model *extended Cox* dengan fungsi waktu logaritma didapatkan, maka akan dilakukan uji signifikansi parameter. Setelah dilakukan uji secara serentak disimpulkan bahwa paling sedikit ada satu variabel independen dari persamaan yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara karena  $G = 72,3096$  lebih besar dari  $\chi^2_{\alpha,p} = 24,996$  atau  $p\text{-value} = <0,0001$  lebih kecil dari  $\alpha = 0,05$ . Untuk mengetahui variabel bebas yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara maka akan dilakukan uji secara parsial

Tabel 5. Uji Parsial Model Awal *Extended Cox Fungsi Waktu Logaritma*

Variabel	Parameter Estimasi	Standard Error	p-value	Keputusan
Umur (X <sub>1</sub> )	0,1878	0,1365	0,1689	H <sub>0</sub> diterima
Ukuran (X <sub>2</sub> )	0,1438	0,0592	0,0151	H <sub>0</sub> ditolak
Tipe Kanker (1) (X <sub>3</sub> (1))	5,7777	10,3278	0,5759	H <sub>0</sub> diterima
Tipe Kanker (2) (X <sub>3</sub> (2))	-1,9699	10,6060	0,8527	H <sub>0</sub> diterima
Tipe Kanker (3) (X <sub>3</sub> (3))	2,9043	10,7109	0,7863	H <sub>0</sub> diterima
Kemoterapi (X <sub>4</sub> )	-0,6111	0,2176	0,0050	H <sub>0</sub> ditolak
Stadium (1) (X <sub>5</sub> (1))	-1,2990	0,4655	0,0053	H <sub>0</sub> ditolak
Stadium (2) (X <sub>5</sub> (2))	-1,3460	0,4397	0,0022	H <sub>0</sub> ditolak
Stadium (3) (X <sub>5</sub> (3))	-0,8912	0,4821	0,0645	H <sub>0</sub> diterima
Jenis Operasi (X <sub>6</sub> )	-0,4054	0,1841	0,0277	H <sub>0</sub> ditolak
Menopause (X <sub>7</sub> )	0,0248	0,2649	0,9255	H <sub>0</sub> diterima
Tumor Utama (X <sub>8</sub> )	0,1292	0,1567	0,4095	H <sub>0</sub> diterima
Ln(T) x Tipe Kanker 1	-0,6165	1,3073	0,6372	H <sub>0</sub> diterima
Ln(T) x Tipe Kanker 2	0,3985	1,3417	0,7664	H <sub>0</sub> diterima
Ln(T) x Tipe Kanker 3	-0,2643	1,3607	0,8460	H <sub>0</sub> diterima

Pada uji secara parsial diperoleh bahwa variabel ukuran (X<sub>2</sub>), kemoterapi (X<sub>4</sub>), stadium 1 (X<sub>5</sub> (1)), stadium 2 (X<sub>5</sub> (2)) dan jenis operasi (X<sub>6</sub>) secara parsial signifikan terhadap model. Variabel yang tidak signifikan terhadap model dikeluarkan dari model. Model kedua

*extended Cox* fungsi logaritma yang terbentuk setelah variabel yang tidak signifikan dikeluarkan adalah sebagai berikut

$$\hat{h}(t, x(t)) = \hat{h}_0 \exp(0,1520X_2 - 0,4307X_4 - 0,6413X_5(1) - 0,5758X_5(2) - 0,4587X_6)$$

Setelah model kedua *extended Cox* dengan fungsi waktu logaritma didapatkan, maka akan dilakukan uji signifikansi parameter terhadap model tersebut baik secara serentak maupun parsial. Setelah dilakukan uji secara serentak dapat disimpulkan paling sedikit ada satu variabel independen dari persamaan yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara karena  $G = 73,3117$  lebih besar dari  $\chi^2_{(5\%;5)} = 11,070$  atau  $p\text{-value} = <0,0001$  lebih kecil dari  $\alpha = 0,05$ . Untuk mengetahui variabel bebas yang berpengaruh secara signifikan terhadap waktu *survival* penderita kanker payudara maka dilakukan uji secara parsial.

Tabel 6. Uji Parsial Model Kedua *Extended Cox* Fungsi Waktu Logaritma

Variabel	Parameter Estimasi	Standard Error	p-value	Keputusan
Ukuran (X <sub>2</sub> )	0,1520	0,0575	0,0082	H <sub>0</sub> ditolak
Kemoterapi (X <sub>4</sub> )	-0,4307	0,1798	0,0166	H <sub>0</sub> ditolak
Stadium 1 (X <sub>5</sub> (1))	-0,6413	0,2851	0,0245	H <sub>0</sub> ditolak
Stadium 2 (X <sub>5</sub> (2))	-0,5758	0,2363	0,0148	H <sub>0</sub> ditolak
Jenis Operasi (X <sub>6</sub> )	-0,4587	0,1799	0,0108	H <sub>0</sub> ditolak

Pada uji secara parsial diperoleh bahwa variabel ukuran (X<sub>2</sub>), kemoterapi (X<sub>4</sub>), stadium 1 (X<sub>5</sub> (1)), stadium 2 (X<sub>5</sub> (2)) dan jenis operasi (X<sub>6</sub>) signifikan terhadap model.

Berdasarkan hasil analisis di atas terdapat tiga model yang terbentuk dari kedua metode yaitu *stratified Cox* tanpa interaksi, *extended Cox* fungsi linier dan *extended Cox* fungsi logaritma. Diperoleh bahwa model terbaik yang digunakan untuk memodelkan probabilitas ketahanan hidup penderita kanker payudara adalah dengan model *stratified Cox* tanpa interaksi dengan AIC terkecil yaitu 1748,774.

Interpretasi *hazard ratio* pada penderita kanker payudara dari *database* METABRIC dari model *stratified Cox* tanpa interaksi yaitu semakin besar ukuran tumor maka semakin besar risiko yang dimiliki oleh penderita kanker payudara untuk gagal bertahan hidup yaitu sebesar 16,6%. Penderita kanker payudara yang melakukan kemoterapi memiliki peluang gagal 0,682 kalinya penderita kanker payudara yang tidak melakukan kemoterapi. Penderita kanker payudara stadium 1 memiliki risiko untuk mengalami *event* lebih besar sebesar 0,516 kali dibanding penderita kanker payudara stadium 4, penderita kanker payudara stadium 2 memiliki risiko untuk mengalami *event* lebih besar sebesar 0,544 kali dibanding penderita kanker payudara stadium 4. Penderita kanker payudara yang melakukan operasi mastektomi memiliki peluang gagal sebesar 0,637 kalinya penderita kanker payudara yang melakukan operasi konservasi.

## 5. KESIMPULAN

Berdasarkan hasil penelitian mengenai pengujian metode regresi Cox pada data penderita kanker payudara dari *database* METABRIC maka dapat ditarik kesimpulan sebagai berikut:

1. Pada penerapan data dengan menggunakan metode regresi *stratified Cox* diperoleh model akhir *stratified Cox* sebagai berikut.

$$\hat{h}_g(t) = \hat{h}_{0g}(t) \exp(0,15390X_2 - 0,38275X_4 - 0,66175X_5(1) - 0,60895X_5(2) - 0,45050X_6)$$

2. Pada penerapan data dengan menggunakan metode regresi *extended Cox* diperoleh model akhir regresi *extended Cox* sebagai berikut

$$\hat{h}(t, x(t)) = \hat{h}_0 \exp(0,15203X_2 - 0,43074X_4 - 0,64133X_5(1) - 0,64133X_5(2) - 0,45870X_6)$$

3. Berdasarkan nilai AIC terkecil didapatkan model terbaik yaitu model regresi *stratified* Cox tanpa interaksi. Faktor-faktor yang mempengaruhi ketahanan hidup penderita kanker payudara dari *database* METABRIC adalah faktor ukuran tumor ( $X_2$ ), kemoterapi ( $X_4$ ), stadium 1 ( $X_5(1)$ ), stadium 2 ( $X_5(2)$ ), dan jenis operasi ( $X_6$ ).

#### DAFTAR PUSTAKA

- Aini, I.N., (2011). “*Extended Cox Model untuk Time-Independent Covariate Yang Tidak Memenuhi Asumsi Proportional Hazard Pada Model Cox Proportional Hazard*”. Depok : Universitas Indonesia.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L.A., & Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 68(6). 394-424.
- Coleman, M.P., *et al.* (2008) Cancer Survival in Five Continents: A Worldwide Population-Based Study (CONCORD). *Lancet Oncology*, 9, 730-756.
- Collet, D. (2003). *Modelling Survival Data in Medical Research Second Edition*. Chapman & Hall.
- Cox, D.R., & Oakes, D. (1984). *Analysis of Survival Data*. London : Chapman and Hall.
- Hosmer, W. D., & Lemeshow, S. (1997). *Applied Survival Analysis Regression Modelling of Time to Event Data*. New York : Wiley.
- Klein, J.P., & Moeschberger, M.L. (2003). *Survival Analysis Techniques for Censored and Truncated Data Second Edition* . Berlin : Springer.
- Kleinbaum, D.G., & Klein, M. (2012). *Survival Analysis A Self-Learning Text Third Edition*. New York : Springer.
- Lee, E.T., & Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*. New York : Wiley.
- Maulida, E.A., (2019). “*Analisis Ketahanan Hidup Penderita Kanker Payudara Menggunakan Regresi Cox Proportional Hazard dan Metode Kaplan Meier*”. Skripsi. Surabaya : UIN Sunan Ampel.
- Nurfain, Purnami, S.W. (2017). Analisis Regresi Cox Extended pada Penderita Kusta di Kecamatan Brondong Kabupaten Lamongan. *Jurnal Sains dan Seni ITS*, 6(1), 89-93.
- Pahlevi, M.R., Mustafid, & Wuryandari, T. (2016). Model Regresi Cox *Stratified* Pada Data Ketahanan. *Jurnal Gaussian*, 5(3), 455-464.
- Scrowcroft, H., (2012). *Increasing the resolution on breast cancer – the METABRIC study*. <https://news.cancerresearchuk.org/2012/04/18/increasing-the-resolution-on-breast-cancer-the-metabric-study/>. Diakses : 14 Februari 2023
- Sinaga, E. S., Ahmad, R.A., & Hutajulu, S.H. (2017). Analisis Ketahanan Hidup 5 Tahun pada Penderita Kanker Payudara di RS Sardjito Provinsi Yogyakarta Indonesia. *Berita Kedokteran Rakyat*, 33(2), 67-72.
- Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2020). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 71(3). 209-249.
- Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2012). Global Cancer Statistics 2012. *CA: A Cancer Journal for Clinicians*. 65(2). 87-108.