

**ANALISIS *SURVIVAL COX PROPORTIONAL HAZARD*
DENGAN ESTIMASI *MAXIMUM LIKELIHOOD* DAN
FIRTH'S PENALIZED LIKELIHOOD
(Studi Kasus pada Pasien Kanker Paru-Paru
Jenis *Non-Small Cell Lung Cancer*)**

SKRIPSI

**OLEH
AISYA FATIMATUZ ZAHRA
NIM. 220601110004**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026**

**ANALISIS *SURVIVAL COX PROPORTIONAL HAZARD*
DENGAN ESTIMASI *MAXIMUM LIKELIHOOD* DAN
FIRTH'S PENALIZED LIKELIHOOD
(Studi Kasus pada Pasien Kanker Paru-Paru
Jenis *Non-Small Cell Lung Cancer*)**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
Aisya Fatimatuz Zahra
NIM. 220601110004**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026**

**ANALISIS *SURVIVAL COX PROPORTIONAL HAZARD*
DENGAN ESTIMASI *MAXIMUM LIKELIHOOD* DAN
FIRTH'S PENALIZED LIKELIHOOD
(Studi Kasus pada Pasien Kanker Paru-Paru
Jenis *Non-Small Cell Lung Cancer*)**

SKRIPSI

Oleh
Aisya Fatimatuz Zahra
NIM. 220601110004

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 5 Juni 2026

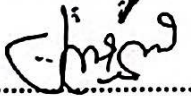
Ketua Penguji : Prof. Dr. Hj. Sri Harini, M.Si.

.....

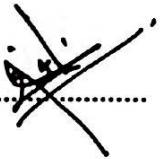

Anggota Penguji 1 : Abdul Aziz, M.Si.

.....


Anggota Penguji 2 : Ria Dhea Layla Nur Karisma, M.Si.

.....


Anggota Penguji 3 : Mohammad Nafie Jauhari, M.Si.

.....


Mengetahui,
Ketua Program Studi Matematika

Dr. Fachrudin Rozi, M.Si.
NIP. 19800527 200801 1 012


**ANALISIS SURVIVAL COX PROPORTIONAL HAZARD
DENGAN ESTIMASI MAXIMUM LIKELIHOOD DAN
FIRTH'S PENALIZED LIKELIHOOD**
(Studi Kasus pada Pasien Kanker Paru-Paru
Jenis Non-Small Cell Lung Cancer)

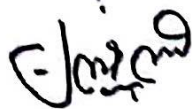
SKRIPSI

Oleh
Alsya Fatimatuz Zahra
NIM. 220601110004

Telah Disetujui untuk Diuji

Malang, 17 April 2026

Dosen Pembimbing I



Ria Dhea Layla Nur Karisma, M.Si.
NIPPPK. 19900709 202321 2 037

Dosen Pembimbing II



Mohammad Naife Jauhari, M.Si.
NIPPPK. 19870218 202321 1 018



Mengetahui,
Ketua Program Studi Matematika

Dr. Fashur Rozi, M.Si.
NIP. 19800528 200801 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Aisyah Fatimatuz Zahra
NIM : 220601110004
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Analisis *Survival Cox Proportional Hazard* dengan Estimasi
Maximum Likelihood dan *Firth's Penalized Likelihood*

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 5 Juni 2026

Yang membuat pernyataan,



Aisyah Fatimatuz Zahra

220601110004

MOTO

“I always thought that life was about standing your ground, no matter how strong the current was. But going with the flow isn’t so bad after all. As long as it takes you forward.”

— Nana Osaki, *Nana* (Ai Yazawa)

“The past cannot be changed, and the future cannot be foretold. As a result, I make sure to always cherish the here and now.”

— (Cyno)

PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala rahmat, kemudahan, dan pertolongan-Nya, karya ini penulis persembahkan kepada:

Pertama, Allah SWT yang senantiasa memberikan kekuatan, petunjuk, dan jalan terbaik dalam setiap proses kehidupan penulis.

Kedua, Ayah tercinta, motivator terbesar penulis, yang selalu memberikan dukungan, doa, dan pengorbanan tanpa henti demi masa depan penulis.

Ketiga, Ibu tercinta, sahabat terbaik penulis, yang selalu menghadirkan kasih sayang, doa, ketulusan, serta menjadi tempat pulang dan sumber kekuatan bagi penulis hingga dapat menyelesaikan karya ini.

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Alhamdulillah rabbil 'alamin, segala puji bagi Allah Tuhan Semesta Alam. Atas segala rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi dengan judul “Analisis *Survival Cox Proportional Hazard* dengan Estimasi *Maximum Likelihood* dan *Firth's Penalized Likelihood*”. Skripsi ini disusun sebagai salah satu syarat untuk menyusun tugas akhir di Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Selama penulisan draf skripsi ini, tak lepas dari peran, doa, dukungan, serta bantuan dari banyak pihak yang sangat berarti. Oleh karena itu, dengan penuh rasa tulus dan hormat, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrur Rozi, M.Si., selaku Ketua Program Studi Matematika yang telah membimbing, memberikan arahan, serta masukan kepada penulis sehingga penulis dapat menyelesaikan proposal skripsi ini.
4. Ria Dhea Layla Nur Karisma, M.Si., selaku Dosen Pembimbing I yang telah memberikan arahan, bimbingan selama penulisan proposal skripsi sehingga penulis dapat menyelesaikan skripsi dengan baik dan tepat waktu.
5. Mohammad Nafie Jauhari, M.Si., selaku Dosen Pembimbing II yang telah melimpahkan pengetahuan, bimbingan, nasihat, motivasi, dan saran yang membangun kepada peneliti.
6. Prof. Dr. Hj. Sri Harini, M.Si. selaku Ketua Penguji yang telah meluangkan waktu, tenaga, dan perhatian dalam memberikan masukan, arahan, serta koreksi yang sangat berharga demi penyempurnaan skripsi ini. Saran dan pandangan beliau menjadi dorongan penting bagi penulis untuk meningkatkan kualitas penelitian ini.

7. Abdul Aziz, M.Si. selaku Penguji I yang dengan penuh kesabaran memberikan kritik, saran, serta bimbingan selama proses ujian dan perbaikan skripsi. Setiap masukan yang diberikan sangat membantu penulis dalam memperbaiki kekurangan dan menyempurnakan penulisan skripsi ini.
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim yang telah memberikan ilmu, bimbingan, serta pengalaman berharga selama masa perkuliahan.
9. Kedua orang tua penulis, Ayah dan Ibu yang tercinta atas doa, kasih sayang, dan dukungan tanpa henti sebagai sumber kekuatan dalam setiap langkah penulis. Terima kasih atas segala kesabaran, kepercayaan, pengorbanan, dan cinta yang diberikan, serta kepada adik tersayang, atas kehadiran, dukungan, doa, dan kebersamaan yang selalu menyertai penulis.
10. Teman-teman baik penulis, terima kasih sudah selalu ada, memberi semangat, motivasi, doa, dukungan, dan kebersamaannya selama ini.
11. Seluruh teman seangkatan Matematika Angkatan 2022, yang telah menjadi teman seperjalanan dan juga memberikan semangat kepada penulis.
12. Semua pihak yang terlibat, baik secara langsung maupun tidak langsung, yang tidak bisa disebutkan satu persatu.
13. Kepada diri sendiri, yang telah melangkah dan berjuang tanpa henti, yang tetap mencoba. Terima kasih telah tidak menyerah.

Akhir kata, penulis menyadari bahwa penulisan ini masih jauh dari kata sempurna. Oleh karena itu, penulis dengan senang hati menerima kritik dan saran yang membangun.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 5 Juni 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xvi
ABSTRAK	xvii
ABSTRACT	xviii
مستخلص البحث.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah.....	7
1.6 Definisi Istilah	7
BAB II KAJIAN TEORI	10
2.1 Teori Pendukung	10
2.1.1 Analisis <i>Survival</i>	10
2.1.2 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i>	10
2.1.3 <i>Survival Time</i> dan Data Tersensor.....	12
2.1.4 Kurva <i>Survival</i> Kaplan–Meier dan Pengujian Log-Rank.....	14
2.1.5 Asumsi <i>Proportional Hazard</i>	17
2.1.6 Model <i>Cox Proportional Hazard</i>	19
2.1.7 Estimasi Parameter pada Model <i>Cox Proportional Hazard</i>	21
2.1.8 Seleksi Model Terbaik.....	28
2.1.9 Evaluasi Model.....	28
2.1.10 <i>Hazard Ratio</i>	29
2.1.11 Kanker Paru-paru	30
2.2 Kajian Integrasi Topik dengan Al-Qur'an/Hadits	33
2.3 Kajian Topik dengan Teori Pendukung	34
BAB III METODE PENELITIAN	36
3.1 Jenis Penelitian.....	36
3.2 Jenis Data dan Sumber Data	36
3.3 Variabel Penelitian	37
3.4 Langkah Penelitian.....	38
3.5 Diagram Alir	40

BAB IV HASIL DAN PEMBAHASAN	41
4.1 Deskripsi Data	41
4.1.1 Karakter Pasien Berdasarkan Kelompok Usia	41
4.1.2 Karakteristik Pasien Berdasarkan Jenis Kelamin	42
4.1.3 Karakteristik Pasien Berdasarkan Status Merokok	43
4.1.4 Karakteristik Pasien Berdasarkan Stadium Kanker	43
4.1.5 Karakteristik Pasien Berdasarkan Histologi	44
4.1.6 Karakteristik Pasien Berdasarkan Komorbiditas	45
4.2 Kurva <i>Survival</i> dan Uji Log-Rank	46
4.2.1 Kurva <i>Survival</i> Pasien Kanker Paru-paru NSCLC	46
4.2.2 Kurva <i>Survival</i> Usia dan Uji Log-Rank	49
4.2.3 Kurva <i>Survival</i> Jenis Kelamin dan Uji Log-Rank	50
4.2.4 Kurva <i>Survival</i> Status Merokok dan Uji Log-Rank	51
4.2.5 Kurva <i>Survival</i> Stadium Kanker dan Uji Log-Rank	52
4.2.6 Kurva <i>Survival</i> Histologi dan Uji Log-Rank	53
4.2.7 Kurva <i>Survival</i> Komorbiditas dan Uji Log-Rank	54
4.3 Uji Asumsi Proportional Hazard	55
4.4 Pemodelan Regresi CoxPH Pasien Kanker NSCLC	56
4.4.1 Estimasi Parameter dengan <i>Maximum Likelihood Estimation</i> .	56
4.4.2 Estimasi Parameter dengan <i>Firth's Penalized Likelihood</i>	61
4.5 Perbandingan Model <i>Cox Proportional Hazard</i>	64
4.6 <i>Hazard Ratio</i>	67
4.7 Integrasi Al-Quran	69
BAB V PENUTUP	71
5.1 Kesimpulan	71
5.2 Saran	72
DAFTAR PUSTAKA	73
LAMPIRAN	76
RIWAYAT HIDUP	90

DAFTAR TABEL

Tabel 3.1	Variabel-variabel Penelitian.....	37
Tabel 4.1	Uji Log-Rank Faktor Usia	50
Tabel 4.2	Uji Log-Rank Faktor Jenis Kelamin	51
Tabel 4.3	Uji Log-Rank Faktor Status Merokok	52
Tabel 4.4	Uji Log-Rank Faktor Stadium Kanker.....	53
Tabel 4.5	Uji Log-Rank Faktor Histologi.....	54
Tabel 4.6	Uji Log-Rank Faktor Komorbiditas	55
Tabel 4.7	Hasil Uji <i>Goodness of Fit</i>	56
Tabel 4.8	Hasil Uji Asumsi <i>Proportional Hazard</i>	57
Tabel 4.9	Hasil <i>Likelihood Ratio Test</i>	58
Tabel 4.10	Hasil Uji Parsial	59
Tabel 4.11	Hasil Regresi <i>Cox Proportional Hazard</i> Kedua	60
Tabel 4.12	Hasil <i>Likelihood Ratio Test</i> Model Kedua.....	60
Tabel 4.13	Hasil Uji Parsial Model Kedua	61
Tabel 4.14	Hasil Regresi CoxPH dengan Estimasi <i>Firth's Penalized Likelihood</i>	62
Tabel 4.15	Hasil <i>Penalized Likelihood Ratio Test</i>	62
Tabel 4.16	<i>Penalized Likelihood Ratio Test</i> Model Kedua	63
Tabel 4.17	Uji Parsial Model Kedua.....	64
Tabel 4.18	Perbandingan Model CoxPH dengan MLE dan FPL	65
Tabel 4.19	Perbandingan <i>Calibration Slope</i> dan <i>C-Index</i> Model CoxPH	66

DAFTAR GAMBAR

Gambar 2.1	Kurva Survival Kaplan-Meier (Kleinbaum & Klein, 2012)	15
Gambar 2.2	Tipe Kanker Paru-paru (Verta, 2019)	31
Gambar 3.1	Diagram Alir Penelitian	40
Gambar 4.1	Karakteristik Pasien Berdasarkan Kelompok Usia	41
Gambar 4.2	Karakteristik Pasien Berdasarkan Jenis Kelamin	42
Gambar 4.3	Karakteristik Pasien Berdasarkan Status Merokok	43
Gambar 4.4	Karakteristik Pasien Berdasarkan Stadium Kanker	44
Gambar 4.5	Karakteristik Pasien Berdasarkan Histologi.....	45
Gambar 4.6	Karakteristik Pasien Berdasarkan Komorbiditas.....	46
Gambar 4.7	Kurva <i>Survival</i> Kaplan–Meier Pasien Kanker NSCLC	47
Gambar 4.8	Kurva <i>Survival</i> Kaplan–Meier Faktor Usia.....	49
Gambar 4.9	Kurva <i>Survival</i> Kaplan–Meier Faktor Jenis Kelamin	50
Gambar 4.10	Kurva <i>Survival</i> Kaplan–Meier Faktor Status Merokok	51
Gambar 4.11	Kurva <i>Survival</i> Kaplan–Meier Faktor Stadium Kanker	52
Gambar 4.12	Kurva <i>Survival</i> Kaplan–Meier Faktor Histologi.....	53
Gambar 4.13	Kurva <i>Survival</i> Kaplan–Meier Faktor Komorbiditas	54
Gambar 4.14	Kurva <i>Survival</i> Observasi dan Prediksi Model CoxPH	66

DAFTAR SIMBOL

Simbol-simbol yang digunakan dalam proposal ini memiliki makna yang dijabarkan sebagai berikut:

T	: Waktu <i>survival</i>
$f(t)$	: Fungsi densitas probabilitas
$F(t) = P(T \leq t)$	: Fungsi distribusi kumulatif
$S(t)$	: Fungsi <i>survival</i>
$h(t)$	: Fungsi <i>hazard</i>
G	: Banyaknya kelompok
O_i	: Nilai observasi individu kelompok ke- i
E_i	: Nilai ekspektasi individu kelompok ke- i
m_{if}	: Jumlah objek yang mengalami <i>event</i> dalam kelompok ke- i dan waktu ke- t
e_{if}	: Nilai ekspektasi individu kelompok ke- i pada waktu t_j
n_{if}	: Jumlah objek yang bertahan hidup dalam kelompok ke- i dan waktu ke- t
n	: Banyaknya observasi
f	: <i>Failure event</i>
i	: $1, 2, \dots, G$
$h(t, X_i)$	: Fungsi <i>hazard</i> individu ke- i
$h_0(t)$	: Fungsi <i>baseline hazard</i>
β_j	: Koefisien regresi ke- j , $j = 1, 2, \dots, p$
x_{ij}	: Variabel bebas observasi ke- i , $j = 1, 2, \dots, p$ dan $i = 1, 2, \dots, n$
$\hat{h}_0(t)$	: Fungsi <i>baseline hazard</i>
$L(\beta)$	: Fungsi <i>likelihood</i> untuk regresi <i>Cox Proportional Hazard</i>
β	: Koefisien regresi
x_{ij}	: Vektor variabel dari individu yang gagal pada waktu ke- j
$R(t_j)$	: Himpunan individu yang bertahan pada waktu ke- j

x_{lj}	: Vektor variabel individu yang masih hidup dan merupakan elemen $R(t_j)$
$U(\beta_r)$	: Fungsi skor dari <i>partial likelihood</i> biasa
$I(\beta)^{-1}$	: <i>Invers</i> dari matriks informasi Fisher
$\frac{\partial \text{Log} L}{\partial \beta_r}$	: Turunan <i>log-partial likelihood</i> terhadap parameter ke- r
$L(\hat{\omega})$	: Nilai <i>likelihood</i> untuk model tanpa menyertakan variabel prediktor
$L(\hat{\Omega})$	: Nilai <i>likelihood</i> untuk model yang menyertakan semua variabel
$SE(\hat{\beta}_j)$	: Standar deviasi dari $\hat{\beta}_j$
$\text{var}(\hat{\beta}_j)$	: Varians dari $\hat{\beta}_j$
$H(t)$	: <i>Hazard</i> pada waktu tertentu
$H_0(t)$	: <i>Baseline hazard</i> pada waktu tertentu
e	: Bilangan natural, $e = 2.714$
HR	: <i>Hazard ratio</i>
$H(t)^*$	: <i>Hazard</i> pada waktu tertentu untuk satu kelompok individu

DAFTAR LAMPIRAN

Lampiran 1. Data Pasien Kanker Paru-Paru Tahun 2016 – 2020	76
Lampiran 2. <i>Script</i> RStudio.....	77

ABSTRAK

Zahra, Aisya Fatimatuz. 2026. **Analisis Survival Cox Proportional Hazard dengan Estimasi Maximum Likelihood dan Firth's Penalized Likelihood**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Mohammad Nafie Jauhari, M.Si.

Kata Kunci: Ketahanan Hidup, Regresi *Cox Proportional Hazard*, NSCLC, *Maximum Likelihood Estimation*, *Firth's Penalized Likelihood*

Analisis ketahanan hidup merupakan metode statistika yang digunakan untuk menganalisis waktu *survival* pasien. Salah satu metode yang digunakan adalah regresi *Cox Proportional Hazard*. Penelitian ini bertujuan untuk membandingkan tingkat kecocokan model *Cox Proportional Hazard* menggunakan estimasi *Maximum Likelihood Estimation* (MLE) dan *Firth's Penalized Likelihood* pada data *survival* pasien *Non-Small Cell Lung Cancer* (NSCLC), serta menganalisis model dengan performa terbaik dalam menggambarkan pola *survival* pasien NSCLC. Hasil penelitian menunjukkan bahwa variabel yang berpengaruh signifikan terhadap ketahanan hidup pasien NSCLC adalah stadium kanker dan jenis histologi. Pasien dengan Stadium IV memiliki risiko kematian 2,3 kali lebih tinggi dibandingkan Stadium III, sedangkan pasien dengan histologi *Squamous* memiliki risiko kematian 1,7 kali lebih tinggi dibandingkan *Adenocarcinoma*. Hasil perbandingan model menunjukkan bahwa model dengan estimasi *Firth's Penalized Likelihood* memiliki nilai AIC yang lebih rendah, sedangkan model dengan estimasi *Maximum Likelihood* memiliki performa kalibrasi yang lebih baik sehingga dipilih sebagai model yang lebih baik pada penelitian ini.

ABSTRACT

Zahra, Aisya Fatimatuz. 2026. **Cox Proportional Hazard Survival Analysis Using Maximum Likelihood Estimation and Firth's Penalized Likelihood.** Undergraduate Thesis. Mathematics Program Study, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Mohammad Nafie Jauhari, M.Si.

Keywords: Survival Analysis, Cox Proportional Hazard Regression, NSCLC, Maximum Likelihood Estimation, Firth's Penalized Likelihood

Survival analysis is a statistical method used to analyze patients' survival time. One method commonly used is the Cox Proportional Hazard regression model. This study aimed to compare the goodness-of-fit of Cox Proportional Hazard models estimated using Maximum Likelihood Estimation (MLE) and Firth's Penalized Likelihood on survival data of Non-Small Cell Lung Cancer (NSCLC) patients, as well as to analyze the model with the best performance in describing the survival pattern of NSCLC patients. The results showed that cancer stage and histological type significantly affected the survival of NSCLC patients. Patients with Stage IV had a mortality risk 2.3 times higher than those with Stage III, while patients with Squamous histology had a mortality risk 1.7 times higher than those with Adenocarcinoma. Model comparison results indicated that the model estimated using Firth's Penalized Likelihood achieved a lower Akaike Information Criterion (AIC) value, whereas the model estimated using Maximum Likelihood Estimation demonstrated better calibration performance. Therefore, the MLE-based model was considered more suitable for the data analyzed in this study.

مستخلص البحث

الزهرى، عائشة فاطمة. ٢٠٢٦. تحليل البقاء باستخدام نموذج كوكس للخطر النسبي بتقدير الإمكان الأعظم وتقدير فيرث الاحتمالي المعاقب. البحث العلمي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانغ. المشرفان: (١) ريا ديا ليلي نور كاريسما، الماجستير في العلوم. (٢) محمد نافع جوهرى، الماجستير في العلوم.

الكلمات الأساسية: تحليل البقاء، انحدار كوكس للخطر النسبي، سرطان الرئة ذو الخلايا غير الصغيرة (NSCLC)، تقدير الإمكان الأعظم، تقدير فيرث الاحتمالي المعاقب.

استخدم تحليل البقاء بوصفه طريقة إحصائية لتحليل زمن بقاء المرضى، واستخدم نموذج انحدار كوكس للخطر النسبي بوصفه أحد الأساليب المستخدمة في هذا المجال. وهدفت الدراسة إلى مقارنة مدى ملاءمة نموذج كوكس للخطر النسبي باستخدام تقدير الإمكان الأعظم (MLE) وتقدير فيرث الاحتمالي المعاقب في بيانات بقاء مرضى سرطان الرئة ذو الخلايا غير الصغيرة ($NSCLC$)، وكذلك تحليل النموذج ذي الأداء الأفضل في وصف نمط البقاء لدى المرضى. وأظهرت النتائج أن مرحلة السرطان والنوع النسيجي أثرا تأثيراً معنوياً في بقاء المرضى. وامتلك المرضى في المرحلة الرابعة خطر وفاة أعلى بمقدار 2.3 مرة مقارنة بالمرحلة الثالثة، في حين امتلك المرضى ذوو النسيج الحشفي خطر وفاة أعلى بمقدار 1.7 مرة مقارنة بالأدينوكارسينوما. كما أظهرت نتائج المقارنة أن نموذج فيرث حقق قيمة (AIC) أقل، بينما أظهر نموذج الإمكان الأعظم أداءً أفضل من حيث المعايير، ولذلك اختير بوصفه النموذج الأفضل في هذه الدراسة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis *survival* adalah metode statistika yang mempelajari durasi hingga terjadinya suatu peristiwa tertentu (*event*). Waktu tersebut dapat diukur sejak awal pengamatan dan dapat dinyatakan dalam satuan hari, minggu, atau bulan, serta berfungsi untuk menilai lamanya individu bertahan hingga munculnya peristiwa yang diteliti (Kleinbaum & Klein, 2012). Metode ini bertujuan untuk menilai kondisi kesehatan individu berdasarkan kejadian yang dialami selama periode pengamatan. Dalam analisis *survival*, dikenal istilah *waktu survival*, yaitu rentang waktu antara awal pengamatan hingga terjadinya suatu peristiwa (Kleinbaum & Klein, 2012). Dengan kata lain, waktu *survival* menggambarkan lamanya individu bertahan hidup sejak awal pengamatan sampai munculnya *event*. Pada penelitian ini, *event* yang dimaksud adalah kematian pasien selama periode observasi. Sementara itu, apabila pasien masih hidup hingga akhir periode tersebut, maka datanya dikategorikan sebagai data tersensor, yaitu data yang hanya menunjukkan lama bertahan hidup tanpa mengetahui waktu pasti terjadinya peristiwa tersebut.

Salah satu model yang sering diaplikasikan dalam studi *survival* adalah model *Cox Proportional Hazard*, yang digolongkan sebagai model semi-parametrik (Harrell, 2015). Dalam analisis *survival*, terdapat pengujian asumsi yang perlu dilakukan, yaitu untuk memastikan terpenuhinya asumsi *proportional hazard* (PH). Asumsi ini menyatakan bahwa nilai *hazard ratio* harus konstan sepanjang waktu (Kleinbaum & Klein, 2012). Pengujian terhadap asumsi *proportional hazard* dapat dilakukan melalui berbagai pendekatan, seperti pendekatan grafik, uji *goodness of*

fit, serta penggunaan variabel kovariat yang bergantung pada waktu (*time-dependent covariate*). Beberapa metode dalam analisis *survival* mensyaratkan terpenuhinya asumsi ini. Salah satu metode yang membutuhkan asumsi *proportional hazard* adalah regresi Cox. Dalam regresi ini, hubungan antara variabel bebas (X) dan variabel terikat (Y) direpresentasikan melalui fungsi *hazard* berbentuk eksponensial terhadap waktu tertentu. Regresi Cox digunakan untuk mengkaji pengaruh beberapa variabel terhadap data *survival* secara simultan (Cox, 1972). Apabila asumsi *proportional hazard* telah terpenuhi, maka analisis dapat dilakukan menggunakan model regresi *Cox Proportional Hazard* (Monica & Purhadi, 2016).

Bentuk fungsi *baseline hazard*, $h_0(t)$ pada regresi Cox, tidak diketahui secara eksplisit. Namun, model ini tetap memungkinkan perhitungan *hazard ratio*, yakni perbandingan fungsi *hazard* antar individu, tanpa harus mengetahui bentuk pasti *baseline hazard* (Kleinbaum & Klein, 2012). *Baseline hazard* merupakan komponen penting dalam fungsi *hazard*, sehingga estimasinya diperlukan untuk membangun model secara lengkap dan valid. Model *Cox Proportional Hazard* yang diestimasi menggunakan metode *Maximum Likelihood Estimation* (MLE) sering mengalami kendala ketika diterapkan pada data yang memiliki *censoring* tinggi, ukuran sampel kecil, atau jumlah kejadian (*event*) yang sangat sedikit. Dalam kondisi seperti ini, estimasi MLE dapat menjadi bias, tidak stabil, bahkan tidak terhingga. Permasalahan ini juga berkaitan dengan fenomena *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap data pelatihan sehingga kinerjanya menurun saat diuji pada data baru. Masalah bias dalam estimasi MLE, Firth (1993) mengembangkan metode *penalized likelihood* yang bertujuan untuk

memperbaiki bias kecil pada estimasi parameter. Metode ini memodifikasi fungsi *likelihood* dengan menambahkan penalti berbasis informasi Fisher, sehingga menghasilkan estimasi parameter yang lebih stabil dan bias yang lebih kecil, bahkan dalam kondisi data kecil atau kejadian langka. Heinze (1999) kemudian mengadaptasi pendekatan *Firth's penalized likelihood* ke dalam regresi Cox untuk analisis *survival*. Mereka menunjukkan bahwa metode ini efektif dalam mengatasi masalah *infinite estimates* yang sering terjadi pada data dengan rasio kejadian rendah, serta meningkatkan validitas hasil analisis *survival*.

Penelitian oleh Adhikary & Shafiqur Rahman (2021) menunjukkan bahwa metode Firth memberikan peningkatan yang signifikan dalam performa prediksi dibandingkan model Cox standar berbasis MLE, khususnya pada data dengan *censoring* tinggi atau kejadian langka. Melalui studi simulasi dan penerapan pada data nyata (seperti data kanker payudara dan data jarak kelahiran), mereka menemukan bahwa model Firth berhasil mengurangi *overfitting* dan bias parameter, terutama ketika rasio *Events Per Variable* (EPV) sangat rendah. Kondisi serupa juga berpotensi ditemukan dalam data pasien kanker paru-paru, terutama pada studi dengan ukuran sampel terbatas, banyak variabel klinis, dan masa pengamatan yang tidak terlalu panjang.

Secara global, kanker paru-paru termasuk salah satu jenis kanker dengan insidensi tertinggi, yakni sebanyak 2,5 juta kasus baru, atau sekitar 12,4% dari total kasus kanker, diikuti oleh kanker payudara, kolorektal, prostat, dan lambung. Selain itu, kanker paru-paru juga menempati posisi utama sebagai penyebab kematian terkait kanker. dengan 1,8 juta kematian, atau sekitar 18,7% dari seluruh kematian akibat kanker di dunia. Tingginya angka kejadian kanker paru-paru ini erat

kaitannya dengan penggunaan tembakau yang masih tinggi, khususnya di Asia (WHO, 2024).

Konsep waktu dan kejadian yang menjadi inti dari analisis *survival* memiliki makna yang sangat dekat dengan ajaran Islam. Al-Qur'an menegaskan bahwa setiap peristiwa yang terjadi di dunia ini telah berada dalam ketentuan Allah SWT, sebagaimana dalam QS. Al-Hadid ayat 22 (Kemenag RI, 2022):

“Tidak ada bencana (apa pun) yang menimpa di bumi dan tidak (juga yang menimpa) dirimu, kecuali telah tertulis dalam Kitab (Lauhulmahfuz) sebelum Kami mewujudkannya. Sesungguhnya hal itu mudah bagi Allah.”

Ayat tersebut mengingatkan bahwa segala sesuatu, termasuk waktu terjadinya suatu peristiwa telah ditetapkan oleh Allah SWT. Namun, manusia tetap diberi akal untuk meneliti, memahami, dan mengambil pelajaran dari setiap kejadian.

Selain itu, penelitian ini juga berupaya membandingkan dua metode estimasi, yaitu *Maximum Likelihood* dan *Firth's Penalized Likelihood*, untuk memperoleh model yang paling sesuai dan akurat. Upaya memilih metode yang lebih baik ini sejalan dengan pesan dalam QS. Az-Zumar ayat 18 (Kemenag RI, 2022):

“(Yaitu) mereka yang mendengarkan perkataan lalu mengikuti apa yang paling baik di antaranya. Mereka itulah orang-orang yang telah diberi petunjuk oleh Allah dan mereka itulah ulul albab (orang-orang yang mempunyai akal sehat).”

Ayat ini menjadi dasar motivasi spiritual bagi peneliti untuk senantiasa berusaha mencari serta menetapkan pendekatan yang paling tepat dalam penyelidikan ilmiah. Dengan demikian, penelitian ini tidak semata berorientasi pada ketepatan hasil secara statistik, tetapi juga dilandasi oleh nilai-nilai keislaman

tentang kebijaksanaan, ketelitian, dan tanggung jawab ilmiah dalam memahami takdir dan waktu melalui ilmu pengetahuan.

Oleh karena itu, penelitian ini bertujuan menerapkan metode *Firth's Penalized Likelihood* pada model *Cox Proportional Hazard* untuk menganalisis data *survival* pasien NSCLC, sekaligus membandingkan tingkat kecocokan model dengan model *Cox* standar. Hasil penelitian diharapkan dapat memberikan informasi yang akurat bagi dunia medis dan sejalan dengan nilai-nilai Islam dalam menjaga kehidupan (*hifz al-nafs*).

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana perbandingan performa model *Cox Proportional Hazard* yang diestimasi menggunakan metode *Maximum Likelihood Estimation* (MLE) dan *Firth's Penalized Likelihood* pada data *survival* pasien NSCLC?
2. Bagaimana analisa hasil model *Cox Proportional Hazard* berdasarkan metode estimasi parameter yang lebih baik dalam menganalisis data *survival* pasien *Non-Small Cell Lung Cancer* (NSCLC)?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah dijelaskan, tujuan dari penelitian ini adalah sebagai berikut.

1. Membandingkan performa model *Cox Proportional Hazard* yang diestimasi menggunakan metode *Maximum Likelihood Estimation* (MLE) dan *Firth's*

Penalized Likelihood pada data *survival* pasien *Non-Small Cell Lung Cancer* (NSCLC).

2. Menganalisis hasil model *Cox Proportional Hazard* berdasarkan metode estimasi parameter yang memberikan performa yang lebih baik dalam menganalisis data *survival* pasien *Non-Small Cell Lung Cancer* (NSCLC).

1.4 Manfaat Penelitian

Manfaat dari penelitian ini dibedakan berdasarkan beberapa kepentingan pihak, yaitu sebagai berikut.

1. Manfaat bagi penulis

Penelitian ini diharapkan dapat meningkatkan pemahaman peneliti dalam bidang analisis *survival*, khususnya dalam penerapan metode estimasi *Firth's Penalized Likelihood* dan estimasi *Maximum Likelihood* pada model *Cox Proportional Hazard*.

2. Manfaat bagi pembaca

Hasil penelitian ini dapat menjadi referensi bagi peneliti, mahasiswa, atau praktisi di bidang statistika ataupun kesehatan masyarakat yang membutuhkan pemahaman mengenai analisis *survival*.

3. Manfaat Program Studi Matematika

Penelitian ini dapat mendukung pengembangan keilmuan Program Studi Matematika, khususnya pada bidang statistika dan analisis *survival*, serta dapat dimanfaatkan sebagai referensi pembelajaran maupun bahan kajian pada mata kuliah statistika.

4. Manfaat bagi Pemilik Data

Hasil penelitian ini dapat menjadi sumber informasi bagi pemilik data atau institusi kesehatan terkait mengenai faktor-faktor yang berpengaruh terhadap ketahanan hidup pasien NSCLC, sehingga dapat digunakan sebagai bahan pertimbangan dalam evaluasi dan pengembangan layanan kesehatan.

5. Manfaat bagi Masyarakat

Penelitian ini diharapkan dapat meningkatkan pengetahuan masyarakat mengenai faktor-faktor yang memengaruhi ketahanan hidup pasien kanker paru-paru, sehingga dapat mendorong kesadaran terhadap pentingnya deteksi dini, pengobatan, dan pemantauan kesehatan secara berkelanjutan.

1.5 Batasan Masalah

Penelitian ini dibatasi pada data pasien NSCLC dengan periode observasi tahun 2016–2021. Evaluasi performa model dilakukan menggunakan *Akaike Information Criterion* (AIC), *calibration slope*, dan *concordance index* (C-index). Interpretasi parameter *hazard ratio* pada pemodelan *Cox Proportional Hazard* dilakukan berdasarkan model dengan performa yang lebih baik hasil perbandingan metode estimasi.

1.6 Definisi Istilah

1. Analisis *Survival* : Analisis yang mempelajari waktu hingga terjadinya suatu peristiwa
2. Data tersensor : Data yang waktu kejadiannya tidak diketahui secara pasti

3. *Proportional Hazard* : Asumsi bahwa *rasio hazard* antar individu konstan terhadap waktu
4. *Cox Proportional Hazard* : Model regresi yang menganalisis hubungan antara waktu kejadian dengan variabel penjelas
5. *Fungsi Hazard* : Peluang terjadinya peristiwa pada waktu tertentu dengan syarat individu masih bertahan hingga waktu tersebut
6. *Baseline Hazard* : Fungsi *hazard* saat semua variabel penjelas bernilai nol
7. *Hazard Ratio* : Perbandingan risiko antara dua kelompok berdasarkan variabel penjelas.
8. *Maximum Likelihood Estimation* : Metode estimasi parameter dengan memaksimalkan fungsi *likelihood*
9. *Firth's Penalized Likelihood* : Modifikasi metode *Maximum Likelihood* untuk mengurangi bias pada estimasi parameter dengan menambahkan penalti berbasis *Fisher Information* ke dalam fungsi *likelihood*
10. *Fisher Information* : Ukuran informasi yang terkandung dalam data tentang parameter model.
11. *Overfitting* : Kondisi ketika model terlalu menyesuaikan data hingga kehilangan kemampuan generalisasi

- 12. Bias : Penyimpangan antara nilai estimasi dan nilai parameter sebenarnya
- 13. *Akaike Criteria Information* : Ukuran untuk memilih model terbaik
- 14. *Non-Small Cell Lung Cancer* : Jenis kanker paru-paru
- 15. Histologi : Ilmu yang mempelajari struktur jaringan secara mikroskopis
- 16. Komorbiditas : Keberadaan satu atau lebih penyakit tambahan yang menyertai penyakit utama

BAB II

KAJIAN TEORI

2.1 Teori Pendukung

2.1.1 Analisis *Survival*

Analisis ketahanan hidup, atau *survival analysis*, merupakan metode statistik yang berfokus pada pengukuran durasi hingga terjadinya suatu kejadian tertentu (*event*), yang dikenal sebagai waktu ketahanan hidup (*survival time*). Waktu ini menggambarkan lamanya individu bertahan selama periode observasi. Dalam konteks analisis *survival*, peristiwa yang diamati bisa berupa kematian, kambuhnya penyakit setelah pengobatan atau perawatan, dan timbulnya penyakit baru (*recurrent event*). Namun, peristiwa tersebut tidak selalu bersifat negatif, peristiwa juga bisa merupakan hal positif seperti kesembuhan atau perbaikan kondisi setelah mengalami suatu penyakit (Kleinbaum & Klein, 2012).

Menurut Cox (2018), ada tiga hal penting yang perlu diperhatikan dalam menentukan waktu terjadinya kegagalan, yaitu: *start point* (titik awal pengamatan) harus jelas dan tidak menimbulkan ambiguitas. Kemudian skala waktu yang digunakan dalam pengukuran harus konsisten. Kemudian yang terakhir definisi dari peristiwa kegagalan (*failure event*) harus dirumuskan dengan jelas.

2.1.2 Fungsi *Survival* dan Fungsi *Hazard*

Fungsi *survival* dan fungsi *hazard* merupakan dua fungsi utama yang ada pada analisis ketahanan hidup (*analysis survival*). Fungsi *survival* merupakan dasar yang digunakan untuk menggambarkan probabilitas suatu individu dapat bertahan hidup terhadap suatu peristiwa dalam rentang waktu tertentu. Fungsi ini

dinotasikan sebagai $S(t)$, yang menyatakan peluang bahwa seseorang masih bertahan hidup setelah waktu ke- t .

Jika T merupakan variabel acak kontinu yang merepresentasikan waktu hingga seorang individu mengalami suatu *event*, maka bentuk umum fungsi *survival* dapat dituliskan sebagai berikut (Kleinbaum & Klein, 2012).

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \quad (2.1)$$

Keterangan:

$S(t)$: Fungsi *survival*

$P(T \leq t)$: Probabilitas *event* sudah terjadi pada atau sebelum waktu t

$F(t)$: Fungsi distribusi kumulatif

Selain itu, *survival function* juga dapat dinyatakan dalam bentuk fungsi kepadatan peluang, sebagaimana dijelaskan pada Persamaan (2.2).

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du \quad (2.2)$$

Keterangan:

$S(t)$: Fungsi *survival*

$P(T \geq t)$: Probabilitas Individu masih hidup pada waktu t

Fungsi *survival* ini digunakan untuk menentukan probabilitas waktu bertahan hidup individu dari waktu awal pengamatan hingga waktu ke- t yang merupakan waktu terjadinya suatu *event* (Collett, 2015).

Kemudian fungsi *hazard* adalah peluang sesaat per satuan waktu (laju) yang berfokus pada terjadinya suatu *event* dalam rentang waktu t hingga $t + \Delta t$, dengan syarat bahwa individu tersebut masih bertahan hingga waktu t . Karena definisinya menggunakan peluang dengan syarat, maka fungsi *hazard* dibangun

dari konsep peluang bersyarat. Fungsi *hazard* didefinisikan secara matematis dalam Persamaan (2.3).

$$\begin{aligned}
 h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < (t + \Delta t) | T \geq t)}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < (t + \Delta t) \cap (T \geq t))}{P(T \geq t) \Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < (t + \Delta t) \cap (T \geq t))}{S(t) \Delta t} \tag{2.3} \\
 &= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \\
 &= \frac{f(t)}{S(t)}
 \end{aligned}$$

2.1.3 *Survival Time* dan Data Tersensor

Dalam analisis *survival*, waktu ketahanan atau *survival time* merujuk pada lamanya periode pengamatan yang dihitung sebagai selang waktu antara titik awal pengamatan hingga terjadinya peristiwa yang menjadi fokus studi. Untuk menentukan waktu *survival*, diperlukan tiga komponen utama. Pertama adalah waktu awal pencatatan (*start point*), yaitu saat dimulainya pengamatan terhadap objek yang diteliti. Kedua, waktu akhir pencatatan (*end point*), yakni saat pengamatan dihentikan, yang berfungsi untuk menentukan apakah data dari individu tersebut termasuk tersensor atau tidak. Ketiga adalah skala waktu yang digunakan untuk mengukur durasi antara awal hingga akhir kejadian, yang dapat dinyatakan dalam satuan hari, minggu, atau tahun (Edina, 2020).

Dalam praktiknya, sering kali ditemui data tersensor, yaitu data yang mencatat informasi mengenai waktu *survival* namun tidak mencerminkan waktu kejadian yang sebenarnya. (Kleinbaum & Klein, 2012) mengemukakan bahwa

terdapat tiga alasan umum mengapa data tersensor dapat terjadi. Pertama, individu tidak mengalami peristiwa yang diamati sepanjang periode pencatatan. Kedua, individu tidak lagi dapat dilacak tanpa kejelasan hingga akhir pencatatan. Ketiga, individu keluar dari penelitian karena meninggal dunia atau karena alasan lain seperti efek samping dari pengobatan.

Dalam analisis *survival*, terdapat tiga jenis penyensoran. Salah satu yang paling umum adalah sensor kanan (*right censoring*) (Collett, 2015).

1. Sensor Kanan (*Right Censoring*)

Sensor kanan terjadi ketika individu yang diteliti belum mengalami peristiwa (*event*) hingga periode pengamatan berakhir, meskipun waktu awal pengamatannya diketahui secara lengkap. Contohnya, seorang responden diobservasi selama lima tahun sejak penelitian dimulai, tetapi pada tahun ketiga individu tersebut pindah ke negara lain sehingga tidak dapat lagi dilacak (*lost to follow-up*). Dalam kondisi ini, diketahui bahwa individu tersebut bertahan hidup setidaknya selama dua tahun, namun waktu pasti terjadinya *event* tidak diketahui. Oleh karena itu, data dari individu tersebut diklasifikasikan sebagai data tersensor kanan.

2. Sensor Kiri (*Left Censoring*)

Sensor kiri muncul ketika waktu awal terjadinya suatu peristiwa (*event*) tidak dapat diketahui, meskipun waktu kejadiannya dapat diamati sepenuhnya sebelum penelitian selesai. Misalnya, dalam penelitian mengenai pasien kanker, seorang pasien diketahui sudah menderita kanker pada pemeriksaan pertama, tetapi tidak ada informasi kapan penyakit tersebut mulai muncul. Dengan kata lain, peristiwa tersebut telah terjadi

sebelum periode pengamatan dimulai, namun waktu pastinya tidak diketahui. Oleh karena itu, data tersebut digolongkan sebagai data tersensor kiri, karena kejadian terjadi sebelum awal observasi dan tidak dapat ditentukan secara tepat.

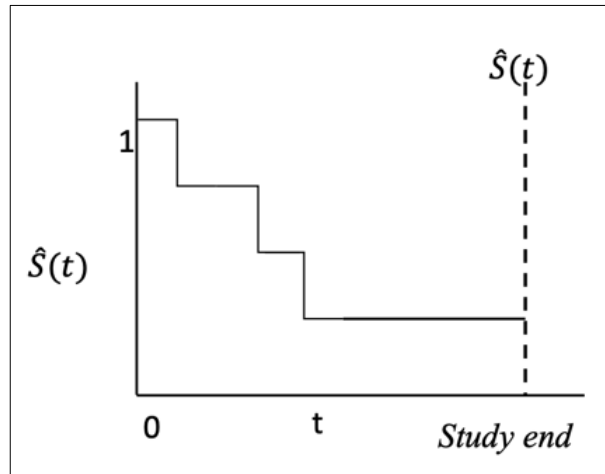
3. Sensor Interval (*Interval Censoring*)

Sensor interval terjadi ketika waktu terjadinya kejadian (*event*) tidak diketahui secara tepat, namun diketahui berada dalam suatu selang waktu tertentu (*interval*). Misal, catatan medis menunjukkan bahwa pada usia 45 tahun seorang pasien belum mengidap kanker. Kemudian pada usia 50 tahun, pasien tersebut menjalani pemeriksaan dan didiagnosis positif kanker. Karena waktu pasti terjadinya kanker tidak diketahui, tetapi diyakini terjadi di antara usia 45 dan 50 tahun, maka data pasien tersebut dikatakan mengalami sensor interval. Artinya, waktu kejadian berada dalam suatu rentang waktu yang diketahui, tetapi tidak dapat ditentukan secara pasti kapan tepatnya kejadian tersebut terjadi.

2.1.4 Kurva *Survival* Kaplan–Meier dan Pengujian Log-Rank

Kurva *survival* Kaplan–Meier menunjukkan hubungan antara probabilitas bertahan hidup dalam periode tertentu (hingga terjadinya *event*) dengan memperhitungkan waktu pada berbagai interval pengamatan (Altman, 1991). Kaplan–Meier juga mempertimbangkan data sensor, yaitu individu dengan status tidak mengalami *event* atau keluar dari pengamatan. Metode estimasi Kaplan–Meier digunakan untuk menganalisis serta membandingkan tingkat ketahanan hidup pasien di antara beberapa kelompok yang diamati dalam penelitian. Gambaran perbandingan antara dua kelompok melalui Kurva *Survival* Kaplan–

Meier dalam mengestimasi fungsi *survival* terhadap waktu ketahanan hidup ditunjukkan pada Gambar 2.1 berikut.



Gambar 2.1 Kurva *Survival* Kaplan–Meier (Kleinbaum & Klein, 2012)

Untuk mendapatkan kurva *survival* Kaplan–Meier, observasi dikelompokkan dalam beberapa pengamatan G . Estimator Kaplan–Meier diperoleh dari hasil pendugaan fungsi *survival* pada suatu kelompok pengamatan, di mana kejadian (*event*) telah diurutkan berdasarkan waktu terjadinya. Bentuk umum dari fungsi *survival* yang digunakan dalam penyusunan Kurva *survival* Kaplan–Meier dapat dinyatakan sebagai berikut.

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)}) \times \hat{P}_r(T > t_{(j)} | T \geq t_{(j)}) \quad (2.4)$$

dimana,

$$\hat{S}(t_{(j-1)}) = \prod_{i=1}^{j-1} \hat{P}_r(T > t_{(i)} | T \geq t_{(i)}) \quad (2.5)$$

Uji log-rank merupakan uji statistik nonparametrik yang digunakan untuk membandingkan perbedaan antara dua atau lebih kurva *survival* Kaplan–Meier pada kelompok yang berbeda. Uji ini tepat digunakan ketika data tidak

berdistribusi normal atau menunjukkan kemiringan ke arah kanan. Hipotesis yang digunakan dalam uji log-rank adalah sebagai berikut.

- H_0 : Tidak terdapat perbedaan kurva *survival* Kaplan–Meier antar kelompok yang berbeda
- H_1 : Terdapat perbedaan minimal satu pada kurva *survival* Kaplan–Meier antar kelompok yang berbeda.

Statistik uji yang digunakan dalam uji log-rank dirumuskan sebagai berikut (Kleinbaum & Klein, 2012)

$$\chi^2_{hitung} = \sum_{i=1}^G \frac{(O_i - E_i)^2}{E_i} \quad (2.6)$$

dimana,

$$O_i - E_i = \sum_{j=1}^h (m_{if} - e_{if}) \quad (2.7)$$

$$e_{if} = \left(\frac{n_{if}}{\sum_{i=1}^G n_{if}} \right) \left(\sum_{i=1}^G m_{if} \right) \quad (2.8)$$

Keterangan:

G : Banyaknya kelompok yang dibandingkan

i : Indeks kelompok, $i = 1, 2, \dots, G$

j : Indeks waktu kegagalan (event time), $j = 1, 2, \dots, h$

O_i : Nilai observasi individu kelompok ke- i

E_i : Nilai ekspektasi individu kelompok ke- i

m_{if} : Jumlah objek yang mengalami *event* dalam kelompok ke- i dan waktu ke- t

e_{if} : Nilai ekspektasi individu kelompok ke- i pada waktu t_j

n_{if} : Jumlah objek yang bertahan hidup dalam kelompok ke- i dan waktu ke- t

n : Banyaknya observasi

h : Banyaknya waktu kejadian (*event times*) dalam data

f : *Failure event*

Keputusan hipotesis nol H_0 ditolak jika nilai statistik $\chi^2_{hitung} > \chi^2_{\alpha;(G-1)}$ dengan taraf signifikansi $\alpha = 0,05$. Hal ini menunjukkan bahwa terdapat minimal satu kurva *survival* yang berbeda secara signifikan antar kelompok (Kleinbaum & Klein, 2012).

2.1.5 Asumsi *Proportional Hazard*

Asumsi *proportional hazard* merupakan prasyarat penting dalam model Cox, yang menyatakan bahwa rasio *hazard* antar kelompok tetap konstan sepanjang waktu. Untuk mengevaluasi apakah data memenuhi asumsi tersebut, dapat digunakan dua pendekatan, yaitu pendekatan grafik dan *goodness of fit* (GOF).

Pengujian asumsi *proportional hazard* secara visual dapat dilakukan dengan pendekatan grafik melalui penggunaan kurva Kaplan–Meier. Asumsi *proportional hazard* dianggap tidak terpenuhi apabila kurva kelangsungan hidup antar kelompok saling berpotongan. Sebaliknya, apabila kurva tersebut tidak berpotongan, maka asumsi *proportional hazard* dianggap terpenuhi (Dahlan, 2012).

Asumsi *proportional hazard* dapat diuji menggunakan pendekatan *goodness of fit* (GOF) melalui analisis residu Schoenfeld. Nilai residu ini dihitung

bagi individu yang mengalami *event* dan untuk masing-masing variabel prediktor yang digunakan dalam model.

Dalam penelitian ini, pengujian dilakukan dengan pendekatan statistik yang menghasilkan nilai *p-value*. Prosedur yang digunakan mengacu pada (Kleinbaum & Klein, 2012), dan melibatkan langkah-langkah sebagai berikut.

1. Perhitungan residu *Schoenfeld* dari model *Cox Proportional Hazard* dilakukan untuk setiap variabel prediktor, sehingga diperoleh nilai residu *Schoenfeld* pada masing-masing variabel prediktor j dan pada setiap kejadian f .
2. Menentukan variabel waktu *survival*, yaitu urutan waktu *survival* dari individu yang diurutkan dari yang terkecil hingga terbesar.
3. Pengujian korelasi dilakukan antara residu *Schoenfeld* parsial dengan *ranked survival time*.
4. Pengujian terhadap asumsi *proportional hazard* dilakukan dengan hipotesis sebagai berikut:

H_0 : $\rho = 0$ (tidak terdapat korelasi antara *RS* dan *RT*, artinya asumsi *proportional hazard* terpenuhi)

H_1 : $\rho \neq 0$ (terdapat korelasi antara *RS* dan *RT*, artinya asumsi *proportional hazard* tidak terpenuhi)

Asumsi *proportional hazard* terpenuhi ketika nilai statistik uji $t_{hitung} > t_{\frac{\alpha}{2}, (F-2)}$ atau $p - value < \alpha$ dengan $\alpha = 0,05$, artinya tidak terdapat korelasi antara residual *Schoenfeld* dengan *ranked-survival time*.

Rumus korelasi adalah sebagai berikut (Afranda dkk., 2015).

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{(n \sum (X)^2 - (\sum X)^2)(n \sum (Y)^2 - (\sum Y)^2)}} \quad (2.9)$$

2.1.6 Model Cox Proportional Hazard

Dalam analisis *survival*, keterkaitan antara waktu ketahanan hidup (*survival time*) dan variabel-variabel prediktor yang diduga berpengaruh dapat dianalisis melalui pendekatan regresi. Model Cox Proportional Hazard merupakan salah satu model yang paling sering diterapkan dalam analisis *survival*, karena mampu menjelaskan keterkaitan antara variabel terikat (Y) dan variabel bebas (X) melalui fungsi *hazard* yang bersifat eksponensial terhadap waktu. Tujuan utama penggunaan model *Cox Proportional Hazard* adalah untuk mengetahui pengaruh beberapa variabel terhadap peluang terjadinya suatu peristiwa (*event*) secara simultan (Cox, 1972). Model ini tergolong semiparametrik, karena tidak mensyaratkan asumsi khusus mengenai distribusi waktu *survival* yang mendasarinya (Kleinbaum & Klein, 2012). Adapun bentuk umum dari model *Cox Proportional Hazard* dapat dinyatakan sebagai berikut (Yasril & Kasjono, 2009).

$$h(t, X_i) = h_0(t) \exp \left[\sum_{j=1}^p \beta_j x_{ij} \right] \quad (2.10)$$

dimana,

$h(t, X_i)$: Fungsi *hazard* individu ke- i

$h_0(t)$: Fungsi *baseline hazard*

β_j : Koefisien regresi ke- j , $j = 1, 2, \dots, p$

x_{ij} : Variabel bebas observasi ke- i , $j = 1, 2, \dots, p$ dan $i = 1, 2, \dots, n$

Ketika parameter β_j bernilai nol ($\beta_j = 0$), Persamaan (2.10) akan berkurang menjadi fungsi *baseline hazard* $\hat{h}_0(t)$ yang merepresentasikan komponen dasar

dari fungsi *hazard*. Secara matematis, kondisi tersebut dapat dinyatakan sebagai berikut.

$$\begin{aligned}
 \hat{h}(t, X_i) &= \hat{h}_0(t) \exp \left[\sum_{j=1}^p \hat{\beta}_j x_{ij} \right] \\
 &= \hat{h}_0(t) \exp(0) \\
 &= \hat{h}_0(t)(1) \\
 &= \hat{h}_0(t)
 \end{aligned} \tag{2.11}$$

Dengan demikian, nilai *hazard* untuk individu ke-*i* sepenuhnya ditentukan oleh fungsi *baseline hazard*. Dalam regresi Cox, bentuk pasti dari $\hat{h}_0(t)$ tidak diketahui secara eksplisit. Namun, model ini tetap memungkinkan estimasi melalui *hazard ratio*, yaitu rasio antara dua fungsi *hazard* dari dua individu yang berbeda (Kleinbaum dan Klein, 2012). Misalkan individu pertama memiliki fungsi *hazard* $\hat{h}(t, X_i)$ dengan $x_i = [x_{i1}, x_{i2}, \dots, x_{ip}]'$ dan individu kedua memiliki fungsi *hazard* $\hat{h}(t, X_i^*)$ dengan $x_i^* = [x_{i1}^*, x_{i2}^*, \dots, x_{ip}^*]'$. Dengan demikian, diperoleh bentuk *hazard ratio* untuk variabel bebas yang bersifat kategorik dan dapat dinyatakan sebagai berikut.

$$\widehat{HR} = \frac{\hat{h}(t, X_i)}{\hat{h}(t, X_i^*)} = \exp \left[\sum_{j=1}^p \hat{\beta}_j (X_{ij} - X_{ij}^*) \right] \tag{2.12}$$

Untuk variabel bebas kontinu, nilai *hazard ratio* diinterpretasikan berdasarkan perubahan sebesar *c* satuan pada variabel tersebut. Sebagai ilustrasi, apabila individu pertama memiliki nilai *hazard* $\hat{h}(t, X_i + c)$ dan individu kedua memiliki nilai *hazard* $\hat{h}(t, X_i)$, maka bentuk umum *hazard ratio* untuk variabel kontinu dapat dirumuskan sebagai berikut (Hosmer dkk., 2008).

$$\begin{aligned}
\widehat{HR} &= \frac{\hat{h}(t, X_i + c)}{\hat{h}(t, X_i^*)} \\
&= \frac{\hat{h}_0(t) \exp[\hat{\beta}(X + c)]}{\hat{h}_0(t) \exp[\hat{\beta}X]} \\
&= \exp[\hat{\beta}(X + c) - \hat{\beta}X] \\
&= \exp[\hat{\beta}X + c\hat{\beta} - \hat{\beta}X] \\
&= \exp[c\hat{\beta}]
\end{aligned} \tag{2.13}$$

Suatu variabel bebas $(X_{i1}, X_{i2}, \dots, X_{ip})$ dikatakan memenuhi asumsi *proportional hazard* jika *hazard ratio* yang dihasilkan bersifat konstan sepanjang waktu.

2.1.7 Estimasi Parameter pada Model *Cox Proportional Hazard*

Misalkan terdapat data dari n individu, yang mencakup r waktu kejadian yang tidak tersensor dan $n - r$ individu yang mengalami penyensoran. Waktu kejadian diurutkan sebagai $t_{(1)} < t_{(2)} < \dots < t_{(j)} < \dots < t_{(n)}$, dengan $t_{(j)}$ menunjukkan waktu kejadian ke- j . Apabila vektor variabel bebas dari individu yang mengalami kejadian pada waktu $t_{(j)}$ dinyatakan dengan x_j , maka peluang terjadinya kejadian pada waktu tersebut dapat dituliskan sebagai berikut:

$$\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P[\text{individu dengan } x_j \text{ mengalami event pada } t_{(j)}]}{P[\text{semua event pada } t_{(j)}]} \\
&= \frac{h_i(t_{(j)})}{\sum_{l \in R(t_j)} h_l(t_{(j)})} \\
&= \frac{h_0(t) \exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{l \in R(t_{(j)})} h_0(t) \exp(\sum_{j=1}^p \beta_j x_{lj})}
\end{aligned} \tag{2.14}$$

$$\begin{aligned}
&= \frac{h_0(t) \exp(\sum_{j=1}^p \beta_j x_{ij})}{h_0(t) \sum_{l \in R(t_{(j)})} \exp(\sum_{j=1}^p \beta_j x_{lj})} \\
&= \frac{\exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{l \in R(t_{(j)})} \exp(\sum_{j=1}^p \beta_j x_{lj})}
\end{aligned}$$

Berdasarkan peluang bersyarat tersebut, fungsi *likelihood* untuk regresi *Cox Proportional Hazard* dapat dituliskan sebagai berikut (Afranda dkk., 2015).

$$L(\beta) = \prod_{i=1}^r \frac{\exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{l \in R(t_{(j)})} \exp(\sum_{j=1}^p \beta_j x_{lj})} \quad (2.15)$$

Keterangan:

$L(\beta)$: Fungsi *likelihood* untuk regresi *Cox Proportional Hazard*

β : Koefisien regresi

x_{ij} : Vektor variabel dari individu ke- i yang gagal pada waktu ke- j

$R(t_j)$: Himpunan individu yang bertahan pada waktu ke- j

x_{lj} : Vektor variabel individu dari *risk set* ke- l yang masih hidup dan merupakan elemen $R(t_j)$

1. *Maximum Likelihood Estimation* (MLE)

Estimasi *maximum likelihood* dari parameter β pada model *proportional hazard* dapat diperoleh dengan memaksimalkan fungsi *log-likelihood* tersebut melalui pendekatan metode numerik. Dari persamaan (2.15) diperoleh fungsi sebagai berikut.

$$\begin{aligned}
\ln L(\beta) &= \ln \prod_{i=1}^r \frac{\exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{l \in R(t_{(j)})} \exp(\sum_{j=1}^p \beta_j x_{lj})} \\
&= \sum_{i=1}^n \left[\ln \left(\exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right) - \ln \left(\sum_{l \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{lj} \right) \right) \right] \quad (2.16)
\end{aligned}$$

$$= \sum_{i=1}^n \left[\left(\sum_{j=1}^p \beta_j x_{ij} \right) - \ln \left(\sum_{l \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{lj} \right) \right) \right]$$

Turunan pertama dari $\ln L(\beta)$ terhadap β_j adalah sebagai berikut.

$$\begin{aligned} \frac{\partial \ln L(\beta)}{\partial \beta_j} &= \frac{\partial \left(\sum_{i=1}^n \left[\left(\sum_{j=1}^p \beta_j x_{ij} \right) - \ln \left(\sum_{l \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{lj} \right) \right) \right] \right)}{\partial \beta_j} \\ &= \sum_{i=1}^n \left[\sum_{j=1}^p x_{ij} - \frac{\sum_{l \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{lj} \right) \sum_{j=1}^p x_{lj}}{\sum_{l \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{lj} \right)} \right] \end{aligned} \quad (2.17)$$

Turunan pertama fungsi *log partial likelihood* terhadap parameter β disebut sebagai *score function* yang dinotasikan dengan $u(\beta)$. Estimasi parameter diperoleh dengan mencari solusi dari:

$$u(\beta) = \frac{\partial \ln L(\beta)}{\partial \beta_j} = 0 \quad (2.18)$$

Solusi dari persamaan tersebut tidak dapat diperoleh secara eksplisit sehingga estimasi parameter dilakukan menggunakan metode iteratif Newton–Raphson. Misalkan $I(\beta)$ merupakan matriks informasi Fisher yang diperoleh dari negatif turunan kedua fungsi *log partial likelihood*, yaitu:

$$- \frac{\partial^2 \log L(\beta)}{\partial \beta_j \partial \beta_k} \quad (2.19)$$

maka estimasi parameter pada iterasi ke-(s+1) diberikan oleh:

$$\hat{\beta}_{s+1} = \hat{\beta}_s + I^{-1}(\hat{\beta}_s)u(\hat{\beta}_s) \quad (2.20)$$

dengan $u(\hat{\beta}_s)$ merupakan vektor turunan pertama dari fungsi *log partial likelihood (score function)* yang dievaluasi pada iterasi ke-s, sedangkan $I^{-1}(\hat{\beta}_s)$ merupakan invers matriks informasi Fisher yang juga dievaluasi

pada iterasi ke- s . Nilai awal parameter biasanya ditentukan sebagai $\hat{\beta}_0 = 0$, kemudian proses iterasi dilakukan secara berulang hingga perubahan nilai estimasi parameter atau perubahan nilai *log partial likelihood* antar iterasi menjadi sangat kecil sehingga diperoleh estimasi parameter yang konvergen. Setelah proses iterasi konvergen, invers matriks informasi Fisher digunakan untuk menduga matriks varians-kovarians dari estimasi parameter, sedangkan akar kuadrat dari elemen diagonal matriks tersebut merupakan *standard error* dari masing-masing estimasi parameter (Collett, 2015).

2. Firth's *Penalized Likelihood*

Dalam penerapan model *Cox Proportional Hazard*, terdapat kondisi di mana estimasi parameter *maximum likelihood* tidak menghasilkan nilai parameter yang terbatas. Hal ini dikenal sebagai *monotone likelihood*, yaitu fenomena ketika nilai *log-partial likelihood* meningkat terus-menerus tanpa batas, namun estimasi parameter justru tidak konvergen dan cenderung menuju nilai tak hingga. Situasi ini sering muncul pada data dengan ukuran sampel kecil, jumlah kejadian (*event*) yang rendah, atau ketika terdapat variabel prediktor yang secara ekstrem memisahkan membedakan antara kelompok yang mengalami kejadian dan yang tidak. Dalam kasus seperti ini, prediktor menyebabkan *separation*, yakni kondisi di mana nilai prediktor mampu secara sempurna memisahkan status kejadian, sehingga menghasilkan estimasi parameter yang sangat besar atau tak hingga (Heinze & Schemper, 2001).

Untuk mengatasi permasalahan tersebut, Firth (1993) mengusulkan metode *penalized likelihood* yang bertujuan menghilangkan bias orde pertama $O(n^{-1})$ dalam estimasi *maximum likelihood*. Meskipun model Cox bersifat semi-parametrik, model ini dapat dituliskan ulang sebagai bagian dari keluarga eksponensial dengan parameter kanonik β , sehingga pendekatan Firth dapat diterapkan. Berdasarkan hal tersebut, Heinze & Schemper (2001) mengadaptasi metode Firth ke dalam model Cox dan mengembangkan fungsi skor yang dimodifikasi sebagai berikut:

$$U(\beta_r)^* = U(\beta_r) + a_r = 0, r = 1, 2, \dots, m, \quad (2.21)$$

dengan

$$a_r = \frac{1}{2} \text{trace} \left\{ I(\beta)^{-1} \left[\frac{\partial \log L}{\partial \beta_r} \right] \right\} \quad (2.22)$$

Keterangan:

$U(\beta_r)$: Fungsi skor dari *partial likelihood* biasa

$I(\beta)^{-1}$: Invers dari matriks informasi Fisher

$\frac{\partial \log L}{\partial \beta_r}$: Turunan *log-partial likelihood* terhadap parameter ke- r .

Modifikasi skor tersebut diturunkan dari bentuk *penalized log-likelihood* berikut:

$$\log L^*(\beta) = \log L(\beta) + \frac{1}{2} \log |I(\beta)| \quad (2.23)$$

Tambahan penalti $\frac{1}{2} \log |I(\beta)|$ dikenal sebagai *Jeffrey's invariant prior*.

Heinze & Schemper (2001) menunjukkan bahwa pendekatan ini secara efektif menghilangkan *monotone likelihood* dan menjamin estimasi parameter yang terbatas, selama masing-masing *risk set* memiliki paling

tidak satu kovariat dengan varians tidak nol, dan setiap kovariat memiliki varians tidak nol pada setidaknya satu *risk set*.

Karena solusi dari fungsi skor termodifikasi tidak dapat diperoleh secara eksplisit, estimasi parameter dilakukan menggunakan algoritma Newton–Raphson secara iteratif. Pada setiap iterasi, komponen koreksi *penalized* a_r diperbarui berdasarkan nilai parameter β saat itu hingga diperoleh estimasi parameter yang konvergen. Heinze & Schemper (2001) menyatakan bahwa tidak diperlukan modifikasi tambahan pada prosedur Newton–Raphson ketika koreksi Firth diterapkan pada regresi *Cox Proportional Hazard*. Varians dari estimator dapat dihitung dari matriks informasi *penalized* $I^*(\beta)$, dan digunakan untuk membentuk standar *error* dan interval kepercayaan.

3. Pengujian Parameter

Uji parameter dilakukan untuk mengetahui sejauh mana variabel independen berpengaruh terhadap variabel dependen (Afranda dkk., 2015). Setelah model terbentuk, pengujian dapat dilaksanakan baik secara simultan maupun secara parsial untuk menilai signifikansi masing-masing variabel.

a. Uji Serentak

Hipotesis yang digunakan dalam uji serentak adalah sebagai berikut:

H_0 : Tidak ada variabel yang berpengaruh signifikan.

H_1 : Minimal ada satu variabel yang berpengaruh signifikan.

Statistik uji yang digunakan adalah:

$$G^2 = -2 \ln \frac{L(\hat{\omega})}{L(\hat{\Omega})} \sim \chi_{a,p}^2 \quad (2.24)$$

Keterangan:

$L(\hat{\omega})$: Nilai *likelihood* untuk model tanpa menyertakan variabel prediktor

$L(\hat{\Omega})$: Nilai *likelihood* untuk model yang menyertakan semua variabel

Pengambilan keputusan dalam uji serentak *likelihood* ini adalah tolak H_0 jika $G_{hit}^2 > \chi_{a,p}^2$ atau $p - value < \alpha = 5\% = 0,05$ (Kleinbaum & Klein, 2012).

b. Uji parsial (*Wald Test*)

Hipotesis yang digunakan pada uji parsial adalah:

H_0 : Variabel ke- j tidak berpengaruh signifikan

H_1 : Variabel ke- j berpengaruh signifikan

Statistik uji yang digunakan adalah:

$$W^2 = \frac{(\hat{\beta}_j)^2}{(SE(\hat{\beta}_j))^2} \sim \chi_{a,1}^2 \quad (2.25)$$

dengan

$$SE(\hat{\beta}_j) = \sqrt{var(\hat{\beta}_j)} \quad (2.26)$$

Pengambilan keputusan dalam uji serentak *likelihood* ini adalah telah H_0 jika $W_{hit}^2 > \chi_{a,1}^2$ atau $p - value < \alpha = 5\% = 0,05$ (Kleinbaum & Klein, 2012).

2.1.8 Seleksi Model Terbaik

Proses pemilihan model bertujuan untuk menentukan variabel-variabel yang pantas dimasukkan ke dalam model regresi *Cox*. Beberapa metode yang dapat digunakan dalam proses ini antara lain seleksi maju, eliminasi mundur, dan metode bertahap. Model terbaik ditetapkan sebagai model dengan nilai *Akaike Information Criterion* (AIC) paling kecil (Collett, 2015)

Secara matematis, AIC dapat dinyatakan sebagai berikut

$$AIC = -2 \ln \hat{L} + 2P \quad (2.27)$$

Keterangan:

$\ln \hat{L}$: Nilai *log-likelihood* dari model regresi *Cox*

P : Banyaknya parameter dalam model regresi *Cox*

2.1.9 Evaluasi Model

Dalam evaluasi performa model prediksi, terdapat dua aspek utama: diskriminasi dan kalibrasi. Diskriminasi menggambarkan kemampuan model membedakan pasien dengan risiko tinggi dan rendah, model yang baik memberikan estimasi risiko yang sejalan dengan perbedaan tingkat kejadian. Kalibrasi menunjukkan ketepatan probabilitas yang diprediksi terhadap kejadian sebenarnya pada setiap tingkat risiko (Royston & Altman, 2013).

1. Kalibrasi

Kalibrasi model dinilai menggunakan *calibration slope*, yaitu kemiringan yang diperoleh dengan melakukan *re-fitting* model Cox PH menggunakan *linear predictor* ($\hat{\eta}_i$) dari model awal sebagai satu-satunya prediktor. Nilai kemiringan ($\hat{\beta}_{PI}$) menggambarkan tingkat kalibrasi, nilai 1 menunjukkan kalibrasi sempurna, nilai < 1

mengindikasikan *overfitting*, dan nilai > 1 menunjukkan *underfitting* (Adhikary & Shafiqur Rahman, 2021).

$$h(t|x_i) = h_0(t) \exp(\beta_{PI} \hat{\eta}_i) \quad (2.28)$$

2. Diskriminasi

Kemampuan diskriminatif model *survival* diukur menggunakan *concordance statistic*, *Gönen and Heller's C-statistic*, yang ekuivalen dengan AUC. Statistik ini menilai sejauh mana prediksi risiko selaras dengan urutan waktu kejadian yang sebenarnya. Nilai 0,5 menunjukkan tidak ada kemampuan diskriminasi, sedangkan nilai mendekati 1 menunjukkan diskriminasi yang semakin baik (Gönen & Heller, 2005).

$$C = \text{AUC} = \frac{2}{n(n-1)} \sum_{i < i'}^n \left[\frac{I(\hat{\beta}x_i > \hat{\beta}x_{i'})}{1 + \exp\{\hat{\beta}(x_{i'} - x_i)\}} + \frac{I(\hat{\beta}x_{i'} > \hat{\beta}x_i)}{1 + \exp\{\hat{\beta}(x_i - x_{i'})\}} \right] \quad (2.29)$$

2.1.10 Hazard Ratio

Nilai *hazard ratio* dalam model regresi *Cox Proportional Hazard* merepresentasikan perbandingan antara laju *hazard* individu dalam satu kelompok dengan laju *hazard* individu di kelompok lain (Kleinbaum & Klein, 2012). Dengan kata lain, *hazard ratio* menggambarkan perbandingan risiko terjadinya suatu kejadian (*event*) antara dua kelompok individu yang berbeda menurut variabel bebasnya. Persamaan umum fungsi *hazard* yang digunakan untuk menghitung *hazard ratio* adalah sebagai berikut (Afranda dkk., 2015):

$$H(t) = H_0(t)e^y \quad (2.30)$$

dengan

$$HR = \frac{H(t)^*}{H(t)} \quad (2.31)$$

Keterangan:

$H(t)$: *Hazard* pada waktu tertentu

$H_0(t)$: *Baseline hazard* pada waktu tertentu

e : Bilangan natural, $e = 2.714$

HR : *Hazard ratio*

$H(t)^*$: *Hazard* pada waktu tertentu untuk satu kelompok individu

2.1.11 Kanker Paru-paru

Secara global, kanker paru-paru termasuk salah satu jenis kanker dengan insidensi tertinggi, yakni sebanyak 2,5 juta kasus baru, atau sekitar 12,4% dari total kasus kanker, diikuti oleh kanker payudara, kolorektal, prostat, dan lambung. Selain itu, kanker paru-paru juga menempati posisi utama sebagai penyebab kematian terkait kanker. dengan 1,8 juta kematian, atau sekitar 18,7% dari seluruh kematian akibat kanker di dunia. Tingginya angka kejadian kanker paru-paru ini erat kaitannya dengan penggunaan tembakau yang masih tinggi, khususnya di Asia (WHO, 2024).

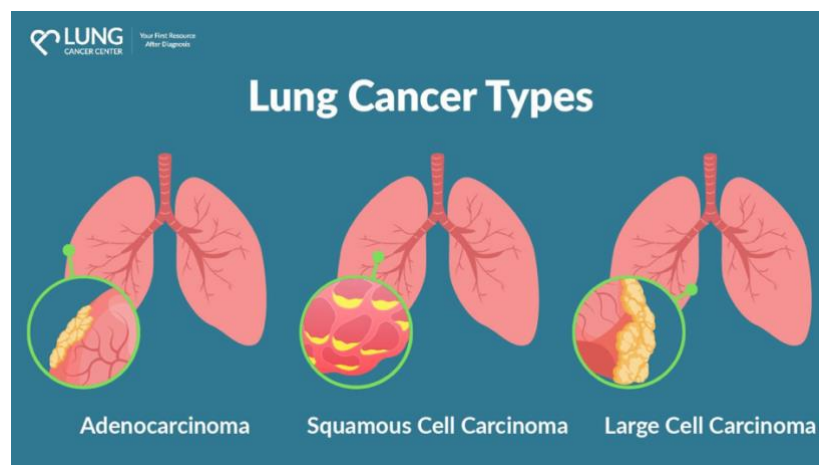
Kanker paru-paru umumnya dibagi menjadi dua kelompok utama, yaitu karsinoma sel kecil (*Small Cell Lung Cancer/SCLC*) dan karsinoma bukan sel kecil (*Non-Small Cell Lung Cancer/NSCLC*). Pada kelompok NSCLC, terdapat tiga subtipe utama yang meliputi adenokarsinoma, karsinoma sel skuamosa, dan karsinoma sel besar. Dari ketiga subtipe tersebut, adenokarsinoma merupakan jenis yang paling banyak ditemukan. Selain itu, adenokarsinoma dan karsinoma sel skuamosa secara bersama-sama dilaporkan menyumbang sekitar 60% dari total kasus kanker paru-paru (Robot dkk., 2021).

1. Usia

Usia merupakan salah satu faktor penting yang memengaruhi ketahanan hidup pasien kanker paru. Menurut (American Cancer Society, 2025), sebagian besar kasus kanker paru didiagnosis pada usia 65 tahun ke atas, dengan median usia diagnosis sekitar 70 tahun. Studi oleh (Tas dkk., 2013) juga menunjukkan bahwa semakin tua usia pasien saat diagnosis, semakin rendah tingkat kelangsungan hidupnya. Hal ini disebabkan oleh penurunan fungsi fisiologis dan meningkatnya kemungkinan adanya penyakit penyerta (komorbiditas) pada usia lanjut.

2. Histologi

Kanker paru bukan sel kecil (*Non-Small Cell Lung Cancer/NSCLC*) terdiri dari beberapa sub tipe histologis utama, yaitu sebagai berikut (Verta, 2019).



Gambar 2.2 Tipe Kanker Paru-paru (Verta, 2019)

- a. *Adenocarcinoma*, adenokarsinoma adalah tipe kanker paru yang paling umum, berkembang pada sel-sel kelenjar yang menghasilkan lendir, dan sering muncul di bagian luar paru-paru. Jenis ini biasanya

ditemukan pada non-perokok atau mantan perokok, serta menyumbang sekitar 40% dari kasus NSCLC.

- b. *Large Cell Carcinoma*, atau karsinoma sel besar bersifat tidak terdiferensiasi, dapat muncul di seluruh bagian paru-paru, tumbuh cepat, dan merupakan subtype yang lebih jarang, sekitar 10–15% dari kasus NSCLC.
- c. *Squamous*, atau karsinoma sel skuamosa berkembang pada sel pipih di saluran udara pusat paru-paru, sering dikaitkan dengan riwayat merokok, dan menyumbang 25–30% kasus.

3. Komorbiditas

Komorbiditas atau penyakit penyerta juga berperan besar dalam menentukan prognosis pasien kanker paru. Penelitian oleh (Wheeler dkk., 2023) menemukan bahwa pasien dengan ≥ 2 komorbiditas memiliki risiko kematian hampir dua kali lipat dibandingkan dengan pasien tanpa komorbiditas ($HR \approx 1,9$; $p < 0,05$). Penyakit penyerta seperti penyakit jantung, diabetes, dan gangguan paru kronis (COPD) dapat memperburuk kondisi klinis pasien, mengurangi toleransi terhadap pengobatan, dan mempercepat progresivitas penyakit.

4. Merokok

Kanker paru-paru sangat terkait dengan kebiasaan merokok, yang menjadi salah satu faktor risiko utamanya. Data dari Provinsi DKI Jakarta (2023) menunjukkan bahwa prevalensi merokok meningkat dari 21,25% pada 2022 menjadi 22,6%. Studi Zhang dkk. (2023) menunjukkan bahwa baik risiko genetik tinggi maupun kebiasaan merokok secara independen

meningkatkan risiko kejadian kanker paru, dengan *hazard ratio* (HR) sebesar 17,89 untuk perokok berat (>40 bungkus/tahun), dibandingkan dengan individu yang tidak pernah merokok.

2.2 Kajian Integrasi Topik dengan Al-Qur'an/Hadits

Dalam sejarah Islam, penyakit telah ada sejak zaman dahulu. Salah satu kisah yang menggambarkan hal ini adalah kisah Nabi Ayyub AS, seorang nabi yang diuji dengan penyakit berat hingga hampir seluruh tubuhnya menderita, kecuali mulutnya. Ujian tersebut berlangsung selama delapan belas tahun, namun Nabi Ayyub tetap sabar dan terus berdoa agar Allah menganugerahkan kesembuhan. Sebagaimana firman Allah SWT dalam surah Asy-Syu'ara ayat 80 (Kemenag RI, 2022):

“Dan apabila aku sakit, Dialah (Allah) yang menyembuhkan aku.”

Hal ini juga sejalan dengan sabda Rasulullah SAW:

“Allah tidak akan menurunkan penyakit melainkan menurunkan obatnya juga.”

Dari ayat dan hadits tersebut dapat dipahami bahwa setiap penyakit yang menimpa manusia pasti memiliki obatnya dengan izin Allah SWT. Nabi Ayyub AS menjadi teladan dalam hal kesabaran dan keimanan, setelah bertahun-tahun berjuang melawan penyakit, doanya pun dikabulkan dan ia sembuh dengan kondisi tubuh yang lebih kuat daripada sebelumnya. Hal ini sebagaimana ditegaskan dalam surah Al-Anbiya ayat 83 (Kemenag RI, 2022):

“Dan (ingatlah kisah) Ayyub, ketika dia berdoa kepada Tuhannya, ‘(Ya Tuhanku), sungguh aku telah ditimpa penyakit, padahal Engkau Tuhan Yang Maha Penyayang dari semua yang penyayang.’”

Oleh karena itu, manusia perlu berperan aktif dalam menjaga kesehatannya, baik fisik maupun moral, dengan menghindari perbuatan yang dapat membahayakan diri, seperti merokok. Salah satu penyakit berbahaya yang erat kaitannya dengan kebiasaan merokok adalah kanker paru-paru, yaitu salah satu jenis kanker yang menjadi penyebab utama kematian di seluruh dunia.

2.3 Kajian Topik dengan Teori Pendukung

Dalam bidang analisis data kesehatan, model *Cox Proportional Hazard* diterapkan pada studi mengenai *survival time* pasien kanker untuk mengidentifikasi variabel-variabel yang memengaruhi peluang bertahan hidup (Collett, 2015). Keunggulan model ini terletak pada sifatnya yang semi-parametrik, di mana tidak diperlukan asumsi eksplisit mengenai bentuk distribusi waktu ketahanan hidup.

Namun, pada data kesehatan yang kompleks seperti data pasien kanker sering ditemukan permasalahan seperti data tersensor, ukuran sampel kecil, atau jumlah kejadian yang tidak seimbang antara kategori variabel. Kondisi tersebut dapat menyebabkan estimasi parameter menjadi bias atau bahkan tak terhingga (*infinite estimates*) ketika model diestimasi menggunakan metode *Maximum Likelihood Estimation* (Heinze & Schemper, 2001).

Untuk mengatasi permasalahan tersebut, Firth (1993) mengusulkan pendekatan *Firth's Penalized Likelihood*, yang memodifikasi fungsi *likelihood* dengan menambahkan penalti berbasis *Jeffreys prior*. Metode ini terbukti dapat mengoreksi bias pada estimasi parameter dan menghasilkan nilai *hazard ratio* yang

lebih stabil dibandingkan dengan *maximum likelihood estimation* (Heinze & Schemper, 2001).

Selain itu, dalam menentukan model terbaik, *Akaike Information Criterion* (AIC) digunakan untuk mengevaluasi antara kecocokan model dan kompleksitas parameter. Dengan demikian, penerapan metode estimasi *Firth's Penalized Likelihood* dan estimasi *maximum likelihood* dalam model *Cox Proportional Hazard* diharapkan mampu memberikan hasil estimasi yang lebih akurat dalam menganalisis ketahanan hidup pasien kanker paru-paru jenis *Non-Small Lung Cancer*.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan metode studi literatur dengan pendekatan deskriptif kuantitatif. Studi literatur merupakan teknik pengumpulan data yang dilakukan dengan menelaah berbagai sumber pustaka atau teori dari beragam referensi yang relevan guna mendukung pelaksanaan dan penyelesaian permasalahan penelitian. Sementara itu, pendekatan deskriptif kuantitatif adalah metode yang dilakukan dengan mengumpulkan data yang dibutuhkan, kemudian menganalisisnya menggunakan metode statistik. Pendekatan ini berfokus pada analisis data dalam bentuk angka atau data numerik untuk memperoleh kesimpulan dari hasil penelitian.

3.2 Jenis Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi (Ríos-Sánchez dkk., 2024), yang memuat informasi pasien kanker paru-paru dari Departemen Onkologi Medis Rumah Sakit Universitas Puerta de Hierro Majadahonda, Madrid, Spanyol. Dataset tersebut mencakup 1.242 pasien, termasuk data diagnosis, pengobatan, serta informasi pribadi dan medis yang dicatat selama anamnesis.

Populasi dalam penelitian ini meliputi seluruh pasien yang terdiagnosis kanker paru-paru tipe NSCLC dan menjalani pengobatan di Rumah Sakit Universitas Puerta de Hierro Majadahonda. Sampel penelitian terdiri dari pasien yang menerima terapi mulai tahun 2016 hingga 2021 dan memenuhi kriteria inklusi

yang telah ditetapkan. Secara keseluruhan, jumlah sampel yang memenuhi kriteria adalah 312 pasien. Data lengkap dapat dilihat pada Lampiran 1.

3.3 Variabel Penelitian

Adapun uraian mengenai masing-masing variabel yang digunakan dalam penelitian ini disajikan pada Tabel 3.1 sebagai berikut.

Tabel 3.1 Variabel-variabel Penelitian		
Variabel	Keterangan	Deskripsi
t	Lama waktu bertahan hidup (hari)	Waktu bertahan hidup pasien sejak awal pengamatan hingga kejadian atau akhir pengamatan.
Status	Status kejadian	0 = Tidak meninggal (<i>censored</i>), pasien pulang, masih hidup, atau pindah pengobatan. 1 = Meninggal (<i>event</i>).
X_1	Kelompok Usia	1 = Usia 30 – 49 tahun 2 = Usia 49 – 64 tahun 3 = Usia 64 – 90 tahun
X_2	Jenis kelamin	1 = Laki-laki 0 = Perempuan
X_3	Status merokok	1 = Non Perokok 2 = Mantan Perokok 3 = Perokok Aktif
X_4	Stadium kanker	3 = Stadium III 4 = Stadium IV
X_5	Histologi kanker	1 = Adenokarsinoma 2 = Karsinoma Sel Skuamosa (KSS) 3 = <i>Large cell carcinoma</i>
X_6	Komorbiditas	1 = Ya, mempunyai penyakit penyerta lain 0 = Tidak memiliki komorbiditas

Sumber: Ríos-Sánchez dkk. (2024)

3.4 Langkah Penelitian

Langkah-langkah yang dilakukan dalam penelitian ini secara umum terdiri dari tahapan pemilihan data, pengolahan, analisis, hingga interpretasi hasil. Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:

1. Mengakses dan menyiapkan data sekunder pasien kanker paru, kemudian memilih data pasien dengan jenis *Non-Small Cell Lung Cancer* (NSCLC) yang terdiri atas tiga subtype utama, yaitu *Adenocarcinoma*, *Squamous*, dan *Large Cell Carcinoma*.
2. Mendeskripsikan karakteristik pasien dan menyusun data *survival*, yaitu waktu bertahan hidup dari diagnosis hingga kejadian (*event*) atau akhir pengamatan (*censoring*), serta mengklasifikasikan status pasien.
3. Mengilustrasikan kurva *survival* pasien NSCLC menggunakan metode Kaplan–Meier berdasarkan Persamaan (2.5). Selanjutnya dilakukan uji log-rank menggunakan statistik uji pada Persamaan (2.6) untuk mengetahui apakah terdapat perbedaan kurva *survival* antar kelompok berdasarkan masing-masing variabel kategorik.
4. Melakukan pengujian asumsi *proportional hazard* menggunakan pendekatan *goodness of fit* (GOF) melalui korelasi antara Schoenfeld residual dan *ranked survival time* sesuai Persamaan (2.9). Asumsi *proportional hazard* dinyatakan terpenuhi apabila tidak terdapat hubungan yang signifikan antara residual dan waktu *survival*.
5. Menyusun model *Cox Proportional Hazard* berdasarkan Persamaan (2.10) menggunakan variabel yang memenuhi asumsi *proportional hazard*,

kemudian mengestimasi parameter model serta menguji signifikansi parameter dengan dua metode sebagai berikut:

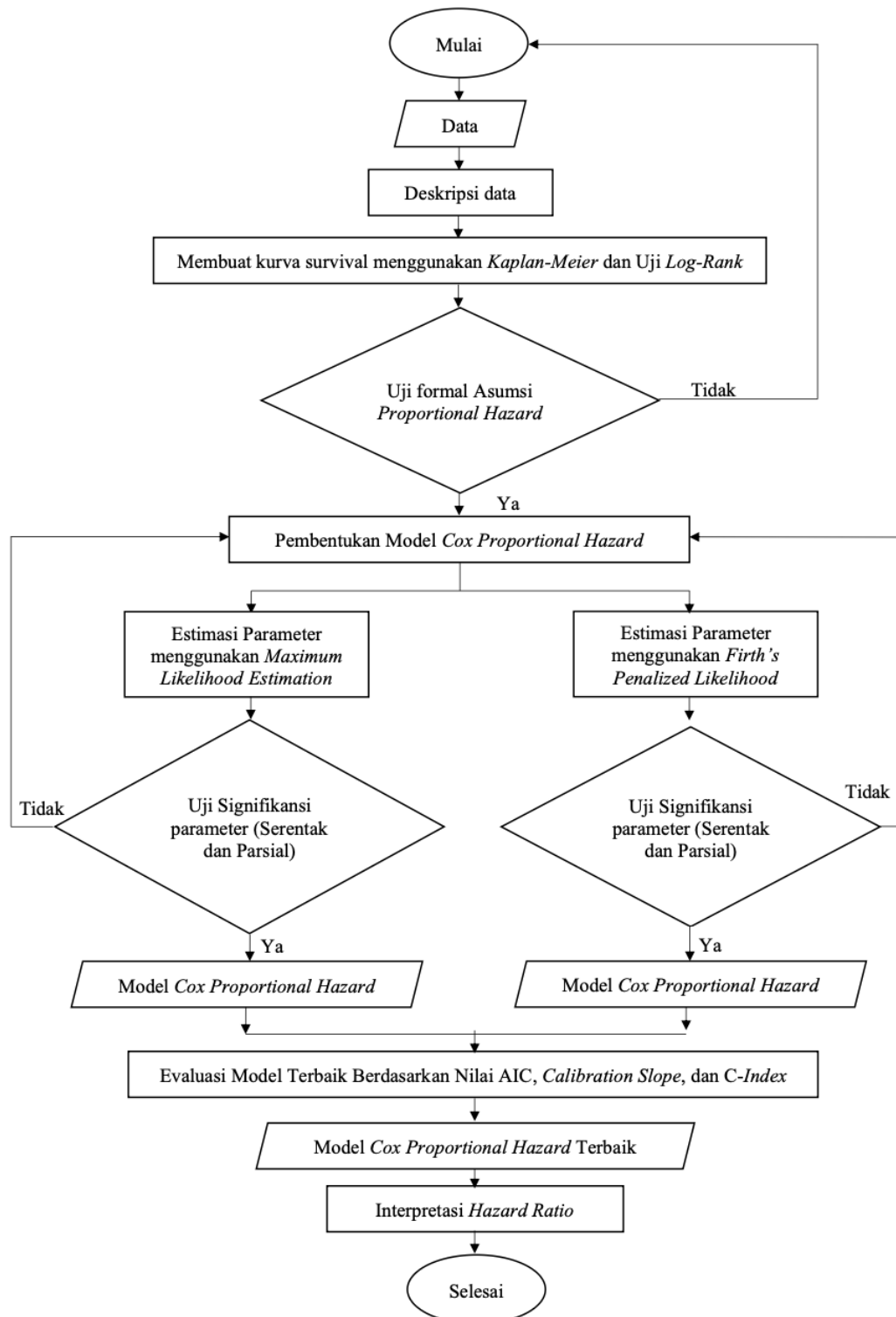
- a. *Maximum Likelihood Estimation* (MLE) sesuai Persamaan (2.16).

Selanjutnya dilakukan pengujian signifikansi parameter menggunakan *Wald test* berdasarkan Persamaan (2.24) dan Persamaan (2.25).

- b. Firth's Penalized Likelihood dengan memaksimalkan fungsi *penalized partial likelihood* sesuai Persamaan (2.23). Selanjutnya dilakukan pengujian signifikansi parameter menggunakan *penalized likelihood ratio test* berdasarkan Persamaan (2.24) dan Persamaan (2.25).

6. Melakukan evaluasi performa model *Cox Proportional Hazard* yang diperoleh dari metode MLE dan *Firth's Penalized Likelihood* dengan membandingkan nilai *Akaike Information Criterion* (AIC) berdasarkan Persamaan (2.27), *calibration slope* berdasarkan Persamaan (2.28), serta *concordance index* (C-index) berdasarkan Persamaan (2.29), kemudian menentukan model yang lebih baik berdasarkan hasil evaluasi tersebut.
7. Menginterpretasikan pengaruh variabel prediktor terhadap *hazard* menggunakan *hazard ratio* sebagaimana pada Persamaan (2.12) berdasarkan model yang lebih baik yang diperoleh.

3.5 Diagram Alir



Gambar 3.1 Diagram Alir Penelitian

BAB IV

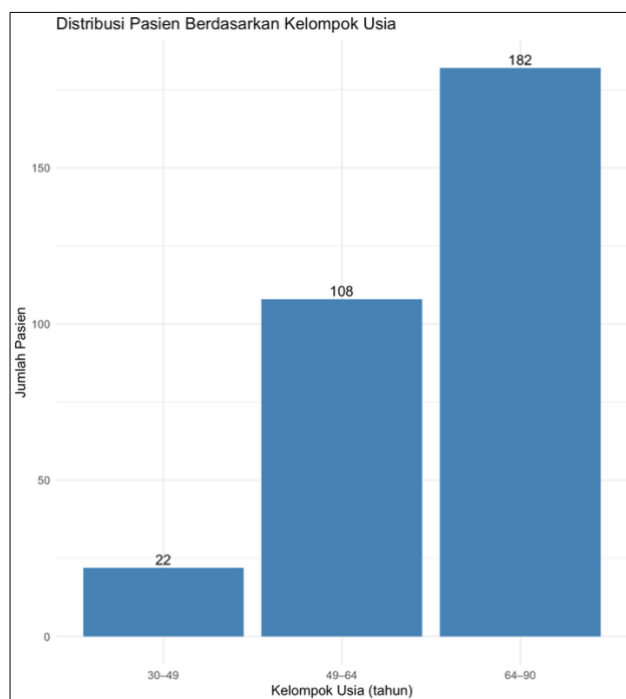
HASIL DAN PEMBAHASAN

4.1 Deskripsi Data

Pada bagian ini akan disajikan gambaran umum karakteristik pasien serta variabel-variabel yang diduga berhubungan dengan ketahanan hidup pasien kanker paru. Variabel waktu ketahanan hidup (t) yang memiliki skala rasio disajikan menggunakan ukuran pemusatan data. Sementara itu, variabel lain yang bersifat kategorik, baik nominal maupun ordinal, meliputi kelompok usia (X_1), jenis kelamin (X_2), status merokok (X_3), stadium kanker (X_4), histologi (X_5), komorbiditas (X_6), dan status pasien, disajikan dalam bentuk distribusi frekuensi dan persentase.

4.1.1 Karakter Pasien Berdasarkan Kelompok Usia

Karakter pasien berdasarkan kelompok usia (X_1) pasien kanker paru-paru jenis *Non-Small Lung Cancer* (NSCLC) dapat dilihat pada Gambar 4.1.

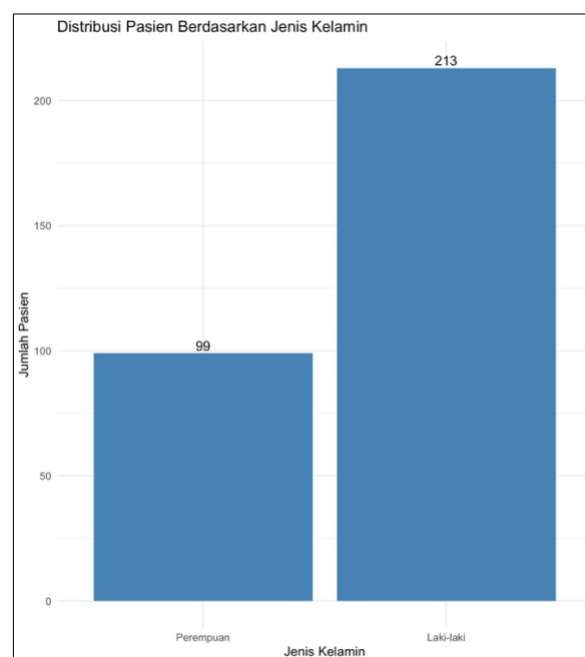


Gambar 4.1 Karakteristik Pasien Berdasarkan Kelompok Usia

Gambar 4.1 menunjukkan distribusi pasien kanker paru berdasarkan kelompok usia. Dari total 312 pasien, sebagian besar berada pada kelompok usia 64–90 tahun, yaitu sebanyak 182 pasien (58,3%). Kelompok usia 49–64 tahun berjumlah 108 pasien (34,6%), sedangkan kelompok usia 30–49 tahun merupakan kelompok dengan jumlah pasien paling sedikit, yaitu 22 pasien (7,1%). Hasil ini menunjukkan bahwa pasien kanker paru dalam penelitian ini didominasi oleh kelompok usia lanjut.

4.1.2 Karakteristik Pasien Berdasarkan Jenis Kelamin

Karakter pasien berdasarkan jenis kelamin (X_2) pasien kanker paru-paru jenis *Non-Small Lung Cancer* (NSCLC) dapat dilihat pada Gambar 4.2.



Gambar 4.2 Karakteristik Pasien Berdasarkan Jenis Kelamin

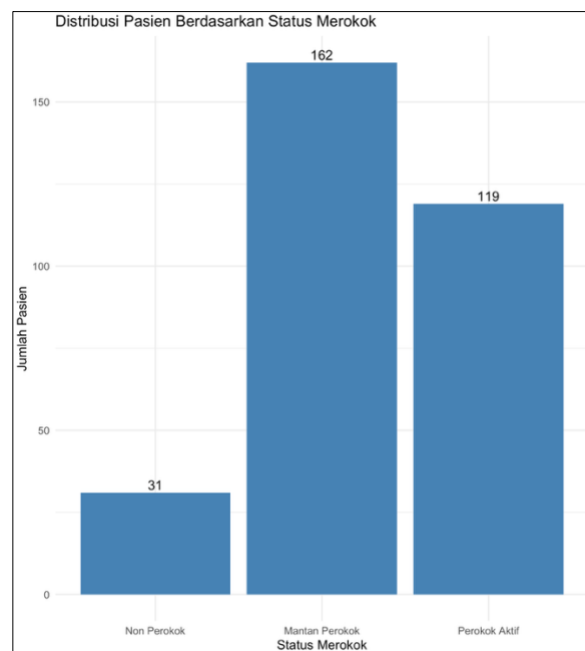
Gambar 4.2 menunjukkan distribusi pasien kanker paru berdasarkan jenis kelamin. Dari total 312 pasien, sebagian besar pasien berjenis kelamin laki-laki, yaitu sebanyak 213 pasien (68,3%), sedangkan pasien perempuan berjumlah 99

pasien (31,7%). Hal ini menunjukkan bahwa pasien kanker paru dalam penelitian ini didominasi oleh laki-laki.

4.1.3 Karakteristik Pasien Berdasarkan Status Merokok

Karakter pasien berdasarkan status merokok (X_3) pasien kanker paru-paru jenis *Non-Small Lung Cancer* (NSCLC) dapat dilihat pada Gambar 4.3.

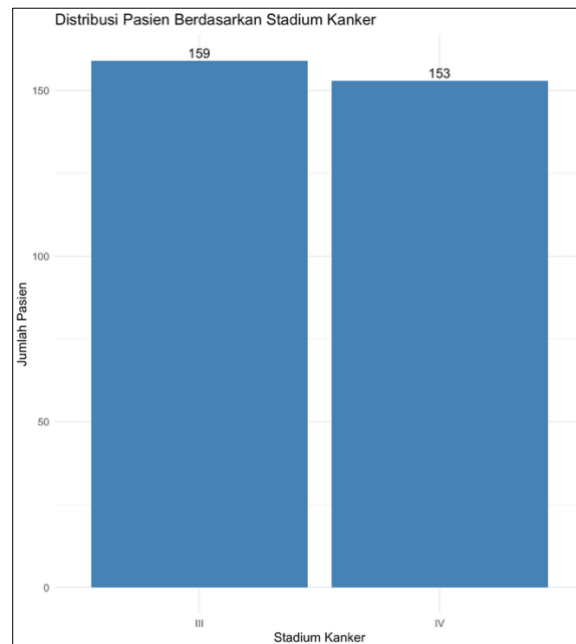
Gambar 4.3 menunjukkan distribusi pasien kanker paru berdasarkan status merokok. Dari total 312 pasien, sebagian besar merupakan mantan perokok, yaitu sebanyak 162 pasien (51,9%), diikuti oleh perokok aktif sebanyak 119 pasien (38,1%). Sementara itu, pasien yang tidak pernah merokok berjumlah 31 pasien (9,9%). Distribusi ini menunjukkan bahwa mayoritas pasien memiliki riwayat merokok.



Gambar 4.3 Karakteristik Pasien Berdasarkan Status Merokok

4.1.4 Karakteristik Pasien Berdasarkan Stadium Kanker

Karakter pasien berdasarkan stadium kanker (X_4) pasien kanker paru-paru jenis *Non-Small Lung Cancer* (NSCLC) dapat dilihat pada Gambar 4.4.

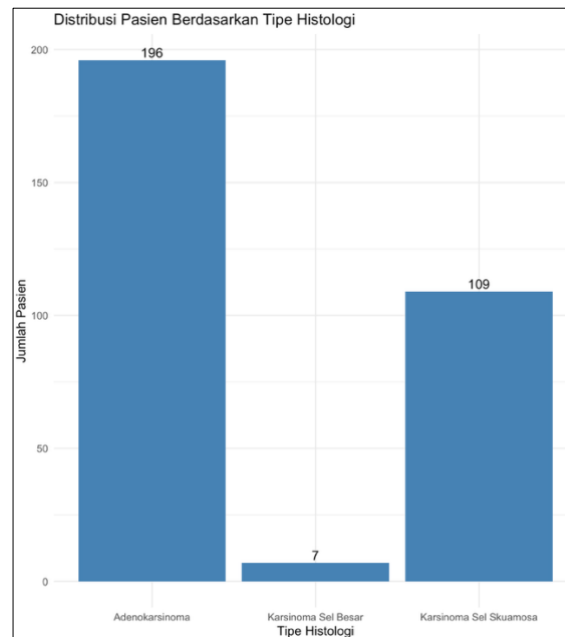


Gambar 4.4 Karakteristik Pasien Berdasarkan Stadium Kanker

Gambar 4.4 menunjukkan distribusi pasien kanker paru berdasarkan stadium kanker. Dari total 312 pasien, sebagian besar pasien berada pada stadium lanjut, yaitu stadium III sebanyak 159 pasien (51%) dan stadium IV sebanyak 153 pasien (49%). Proporsi tersebut menunjukkan bahwa komposisi pasien stadium III dan stadium IV dalam penelitian ini relatif seimbang, dengan selisih persentase yang tidak terlalu besar antara kedua kelompok.

4.1.5 Karakteristik Pasien Berdasarkan Histologi

Karakter pasien berdasarkan histologi (X_5) pasien kanker paru-paru jenis *Non-Small Lung Cancer* (NSCLC) dapat dilihat pada Gambar 4.5.

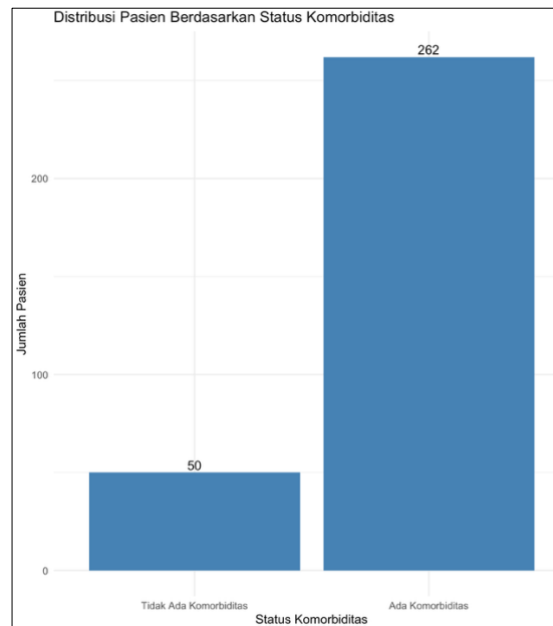


Gambar 4.5 Karakteristik Pasien Berdasarkan Histologi

Gambar 4.5 menunjukkan distribusi pasien kanker paru berdasarkan jenis histologi. Seluruh jenis histologi yang disajikan dalam penelitian ini termasuk dalam kelompok *non-small cell lung cancer* (NSCLC), yang terdiri dari adenokarsinoma, karsinoma sel besar, dan karsinoma skuamosa. Dari total 312 pasien, jenis histologi yang paling banyak ditemukan adalah adenokarsinoma sebanyak 196 pasien (62,8%), diikuti oleh karsinoma skuamosa sebanyak 109 pasien (34,9%), dan karsinoma sel besar sebanyak 7 pasien (2,2%). Hasil ini menunjukkan bahwa adenokarsinoma merupakan tipe histologi NSCLC yang dominan pada pasien kanker paru dalam penelitian ini.

4.1.6 Karakteristik Pasien Berdasarkan Komorbiditas

Karakter pasien berdasarkan komorbiditas (X_6) pasien kanker paru-paru jenis *Non-Small Lung Cancer* (NSCLC) dapat dilihat pada Gambar 4.6.



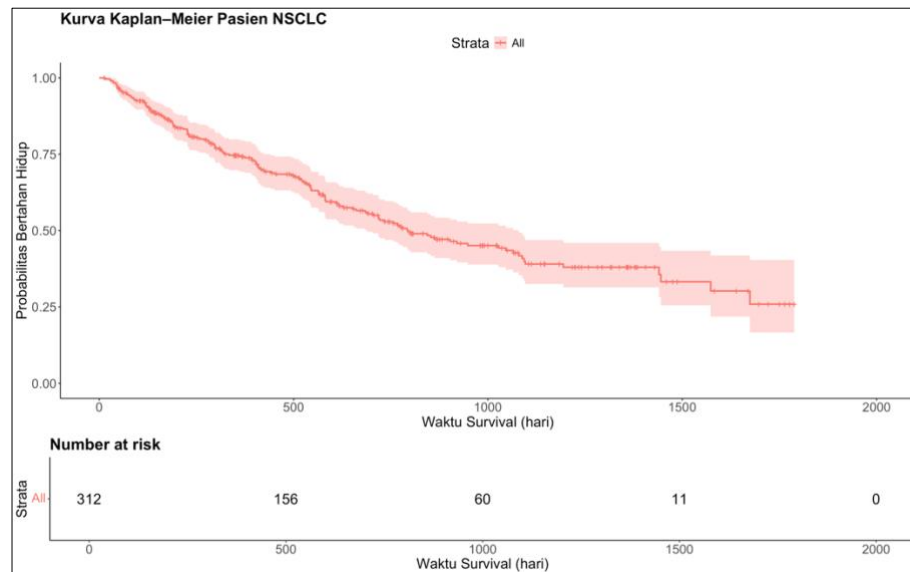
Gambar 4.6 Karakteristik Pasien Berdasarkan Komorbiditas

Gambar 4.6 menunjukkan distribusi pasien kanker paru berdasarkan status komorbiditas. Dari total 312 pasien, sebagian besar pasien memiliki komorbiditas, yaitu 262 pasien (84%), sedangkan pasien yang tidak memiliki komorbiditas berjumlah 50 pasien (16%). Hasil ini menunjukkan bahwa keberadaan penyakit penyerta merupakan kondisi yang umum pada pasien kanker paru dalam penelitian ini.

4.2 Kurva *Survival* dan Uji Log-Rank

4.2.1 Kurva *Survival* Pasien Kanker Paru-paru NSCLC

Analisis Kaplan–Meier digunakan untuk menggambarkan ketahanan hidup pasien kanker paru-paru jenis *Non-Small Cell Lung Cancer* (NSCLC). Pendugaan fungsi *survival* dilakukan menggunakan metode Kaplan–Meier sebagaimana dirumuskan pada Persamaan (2.4), yang memperhitungkan peluang bertahan hidup pada setiap waktu kejadian.



Gambar 4.7 Kurva *Survival* Kaplan–Meier Pasien Kanker NSCLC

Gambar 4.7 menunjukkan kurva *survival* Kaplan–Meier yang menggambarkan probabilitas bertahan hidup pasien kanker paru-paru jenis *Non-Small Cell Lung Cancer* (NSCLC) selama periode pengamatan. Sebagai ilustrasi perhitungan Kaplan–Meier, pada waktu kejadian pertama (hari ke-15) terdapat 309 ($n_1 = 309$) pasien yang masih berada dalam *risk set* dan 1 pasien mengalami *event*. Berdasarkan Persamaan (2.4), estimasi probabilitas bertahan hidup dihitung sebagai berikut:

$$\begin{aligned}
 \hat{S}(t_1) &= \hat{S}(t_0) \times P(T > t_1 | T \geq t_1) \\
 &= 1 \times \left(\frac{309 - 1}{309} \right) \\
 &= \frac{308}{309} \\
 &= 0.9968
 \end{aligned}$$

Nilai tersebut menunjukkan bahwa probabilitas pasien NSCLC untuk tetap bertahan hidup setelah waktu kejadian pertama adalah sebesar 99,68%. Jumlah pasien dalam *risk set* pada waktu kejadian pertama sebanyak 309 orang karena

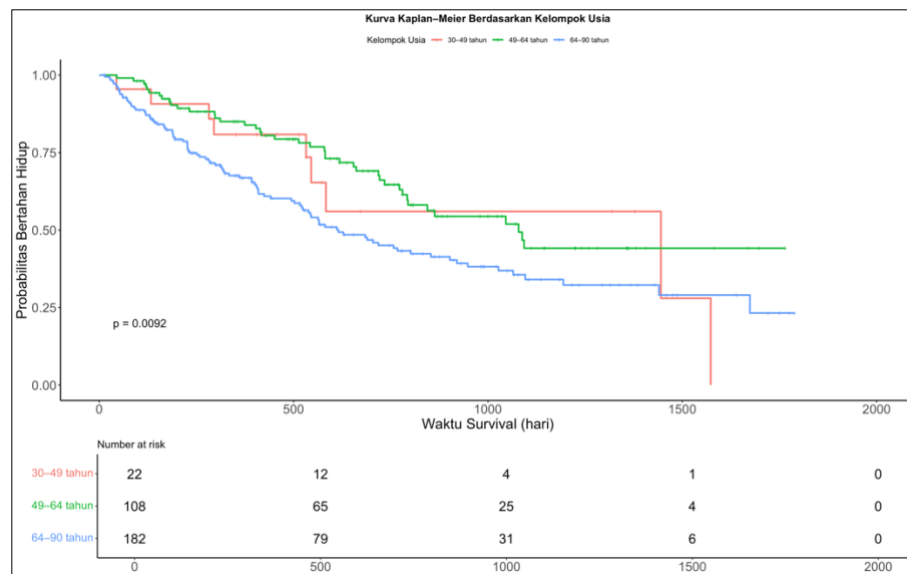
sebelum hari ke-15 terdapat beberapa observasi yang tersensor sehingga tidak lagi termasuk dalam *risk set* pada waktu tersebut.

Kurva mengalami penurunan secara bertahap seiring bertambahnya waktu *survival*. Terlihat bahwa penurunan probabilitas bertahan hidup terjadi cukup tajam pada periode awal hingga sekitar 500 hari pertama, yang mengindikasikan tingginya kejadian kematian pada fase awal setelah diagnosis. Setelah periode tersebut, kurva cenderung menurun lebih landai, menunjukkan bahwa risiko kematian relatif lebih stabil pada pasien yang mampu bertahan melewati fase awal penyakit.

Pada akhir masa pengamatan (sekitar 1.800 hari), probabilitas bertahan hidup berada di bawah 0,50, yang menandakan bahwa kurang dari separuh pasien masih bertahan hidup hingga waktu tersebut. Tabel *number at risk* memperlihatkan bahwa jumlah pasien yang masih berisiko semakin berkurang seiring waktu, sehingga estimasi *survival* pada waktu akhir pengamatan didasarkan pada jumlah subjek yang relatif sedikit.

Penurunan probabilitas bertahan hidup yang terus terjadi menunjukkan bahwa NSCLC merupakan penyakit yang bersifat progresif sehingga risiko kematian tetap meningkat seiring waktu, meskipun pasien telah menjalani pengobatan. Pengobatan pada pasien NSCLC umumnya bertujuan untuk memperlambat perkembangan penyakit dan memperpanjang ketahanan hidup pasien. Hal ini didukung oleh penelitian sebelumnya yang menyatakan bahwa pasien NSCLC tanpa pengobatan memiliki prognosis yang buruk dengan median *survival* sekitar 2–5 bulan sejak diagnosis, sedangkan pasien yang mendapatkan terapi menunjukkan ketahanan hidup yang lebih baik (Wao dkk., 2013).

4.2.2 Kurva *Survival* Usia dan Uji Log-Rank



Gambar 4.8 Kurva *Survival* Kaplan–Meier Faktor Usia

Pada Gambar 4.8 warna merah menunjukkan kurva *survival* pasien kanker paru-paru NSCLC yang tergolong dalam kelompok usia 30–49 tahun, warna hijau menunjukkan kelompok usia 49–64 tahun, dan warna biru menunjukkan kelompok usia 64–90 tahun. Terlihat bahwa kurva *survival* antar kelompok usia tidak saling terhimpit dan menunjukkan pola pemisahan yang jelas sepanjang waktu pengamatan.

Kelompok usia 64–90 tahun memiliki penurunan kurva *survival* yang lebih cepat dibandingkan kelompok usia lainnya, yang menunjukkan probabilitas bertahan hidup yang lebih rendah. Sebaliknya, kelompok usia 49–64 tahun menunjukkan probabilitas bertahan hidup yang lebih tinggi sepanjang periode pengamatan. Sementara itu, kelompok usia 30–49 tahun menunjukkan pola *survival* yang relatif terbaik pada awal pengamatan, namun mengalami penurunan pada akhir periode *follow-up*.

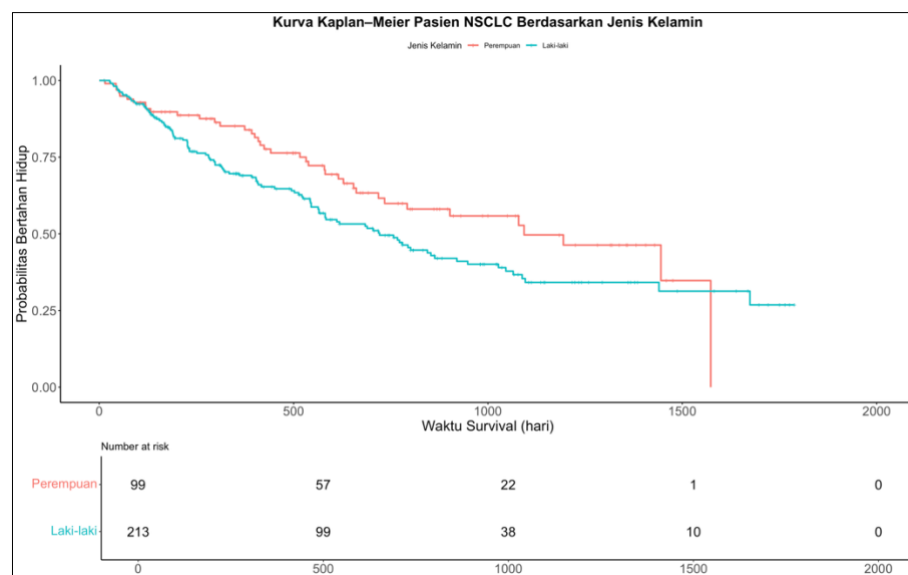
Untuk memperkuat kesimpulan adanya perbedaan probabilitas bertahan hidup antar kelompok usia, dilakukan uji log-rank . Hasil uji log-rank ditunjukkan pada Tabel 4.1.

Tabel 4.1 Uji Log-Rank Faktor Usia

Log-Rank	<i>df</i>	<i>p-value</i>
9.4	2	9×10^{-3}

Tabel 4.1 menunjukkan hasil uji log-rank untuk faktor usia dengan nilai statistik uji log-rank sebesar 9.4 dan *p-value* sebesar 9×10^{-3} . Dengan tingkat signifikansi 0,05, diperoleh keputusan tolak H_0 , yang berarti terdapat perbedaan yang signifikan secara statistik antara kurva *survival* pasien kanker paru-paru NSCLC berdasarkan kelompok usia. Dengan demikian, dapat disimpulkan bahwa kelompok usia berpengaruh terhadap probabilitas ketahanan hidup pasien kanker paru-paru NSCLC.

4.2.3 Kurva *Survival* Jenis Kelamin dan Uji Log-Rank



Gambar 4.9 Kurva *Survival* Kaplan–Meier Faktor Jenis Kelamin

Gambar 4.9 menunjukkan kurva *survival* Kaplan–Meier pasien kanker paru-paru NSCLC berdasarkan jenis kelamin. Kurva *survival* pasien perempuan

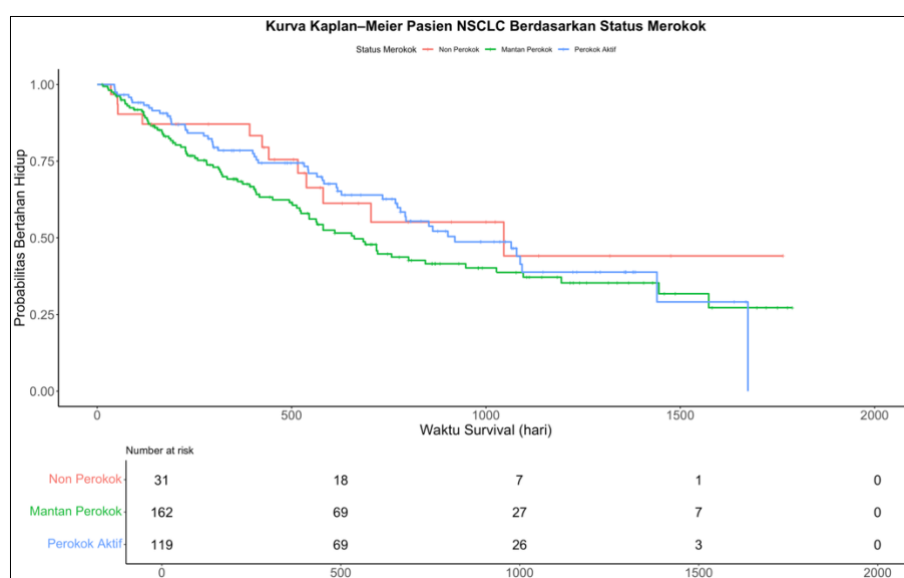
cenderung berada di atas kurva pasien laki-laki sepanjang waktu pengamatan, yang menunjukkan probabilitas bertahan hidup yang lebih baik pada pasien perempuan. Untuk menguji perbedaan kurva *survival* tersebut, dilakukan uji log-rank.

Tabel 4.2 Uji Log-Rank Faktor Jenis Kelamin

Log-Rank	<i>df</i>	<i>p-value</i>
4.1	1	4×10^{-2}

Hasil uji log-rank menunjukkan nilai statistik *chi-square* sebesar 4.1 dengan *p-value* sebesar 4×10^{-2} . Dengan taraf signifikansi 5%, dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara kurva *survival* pasien berdasarkan jenis kelamin. Dengan demikian, jenis kelamin berhubungan dengan perbedaan probabilitas ketahanan hidup pasien NSCLC, di mana pasien laki-laki cenderung mengalami *event* lebih cepat dibandingkan pasien perempuan.

4.2.4 Kurva *Survival* Status Merokok dan Uji Log-Rank



Gambar 4.10 Kurva *Survival* Kaplan–Meier Faktor Status Merokok

Gambar 4.10 menunjukkan kurva *survival* Kaplan–Meier pasien kanker paru-paru NSCLC berdasarkan status merokok. Terlihat adanya perbedaan pola

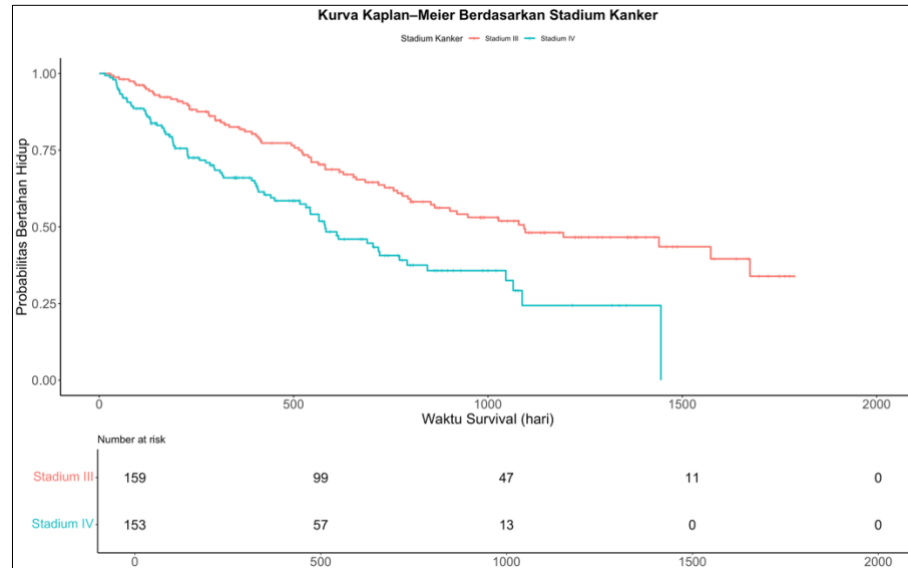
kurva *survival* antar kelompok, di mana pasien non-perokok cenderung memiliki probabilitas bertahan hidup yang lebih tinggi dibandingkan kelompok mantan perokok dan perokok aktif. Untuk menguji perbedaan kurva *survival* tersebut, dilakukan uji log-rank.

Tabel 4.3 Uji Log-Rank Faktor Status Merokok

Log-Rank	<i>df</i>	<i>p-value</i>
3.2	2	0.2

Hasil uji log-rank menunjukkan nilai statistik *chi-square* sebesar 3,2 dengan *p-value* sebesar 0,2. Dengan taraf signifikansi 5%, hasil ini menunjukkan bahwa terdapat perbedaan yang signifikan antara kurva *survival* pasien berdasarkan status merokok. Hal ini menunjukkan bahwa status merokok berhubungan dengan perbedaan probabilitas ketahanan hidup pada pasien NSCLC.

4.2.5 Kurva *Survival* Stadium Kanker dan Uji Log-Rank



Gambar 4.11 Kurva *Survival* Kaplan–Meier Faktor Stadium Kanker

Gambar 4.11 menunjukkan kurva *survival* Kaplan–Meier pasien kanker paru NSCLC berdasarkan stadium kanker. Terlihat adanya perbedaan pola kurva *survival* antara pasien stadium III dan stadium IV. Pasien dengan stadium III

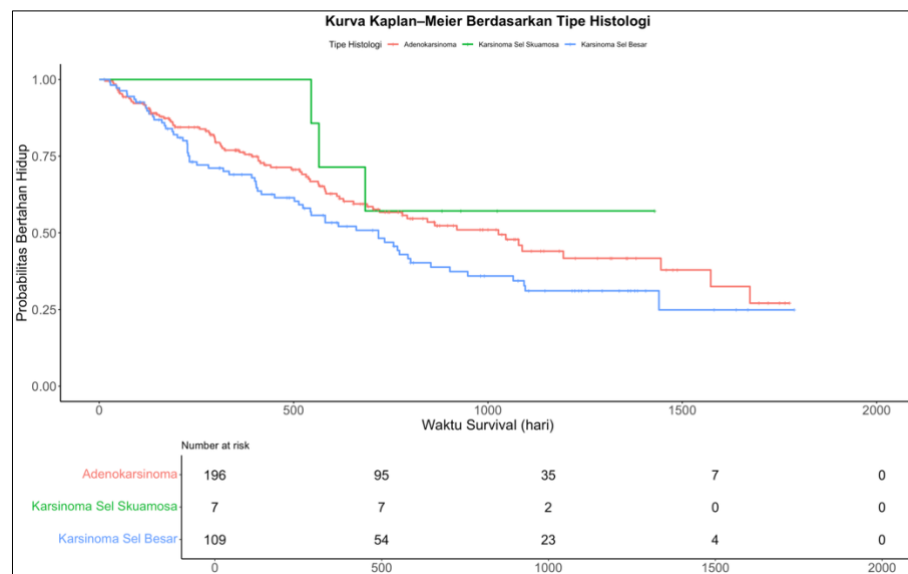
cenderung memiliki probabilitas bertahan hidup yang lebih tinggi dibandingkan pasien stadium IV. Kurva stadium IV menunjukkan penurunan yang lebih tajam, yang mengindikasikan risiko kematian yang lebih tinggi selama periode pengamatan. Untuk menguji perbedaan kurva *survival* tersebut, dilakukan uji log-rank.

Tabel 4.4 Uji Log-Rank Faktor Stadium Kanker

Log-Rank	<i>df</i>	<i>p-value</i>
16.2	1	6×10^{-5}

Hasil uji log-rank menunjukkan nilai statistik *chi-square* 16.2 dengan derajat bebas 2 dan *p-value* sebesar 6×10^{-5} . Dengan taraf signifikansi 5%, hasil ini menunjukkan bahwa terdapat perbedaan yang sangat signifikan antara kurva *survival* pasien berdasarkan stadium kanker, artinya stadium kanker berhubungan secara signifikan dengan perbedaan probabilitas ketahanan hidup pada pasien NSCLC, di mana pasien dengan stadium IV memiliki ketahanan hidup yang lebih rendah dibandingkan pasien stadium III.

4.2.6 Kurva *Survival* Histologi dan Uji Log-Rank



Gambar 4.12 Kurva *Survival* Kaplan–Meier Faktor Histologi

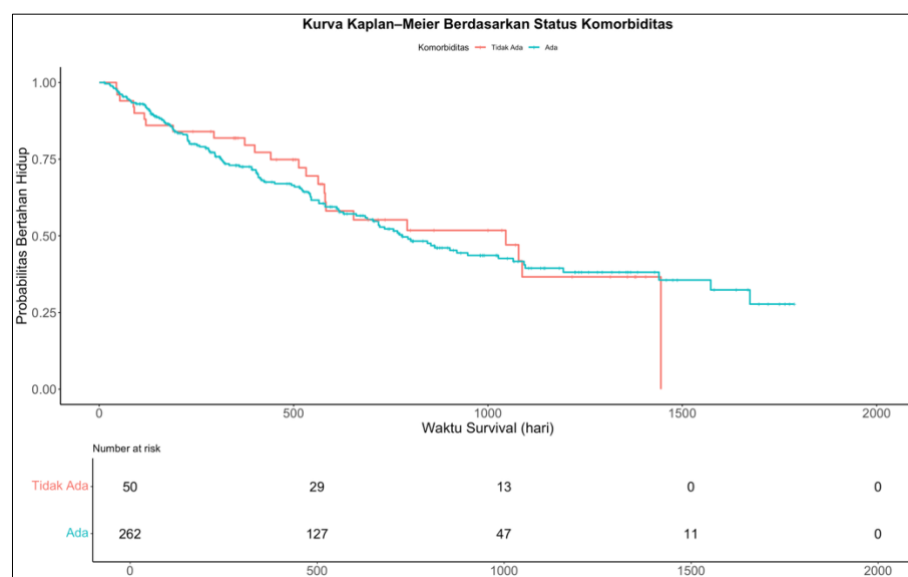
Gambar 4.12 menunjukkan kurva survival Kaplan–Meier pasien kanker paru-paru NSCLC berdasarkan tipe histologi. Secara visual, terlihat adanya perbedaan pola kurva *survival* antar kelompok histologi, di mana pasien dengan adenokarsinoma cenderung memiliki probabilitas bertahan hidup yang lebih tinggi dibandingkan pasien dengan tipe karsinoma sel skuamosa. Untuk menguji perbedaan kurva *survival* tersebut, dilakukan uji log-rank.

Tabel 4.5 Uji Log-Rank Faktor Histologi

<i>Log-Rank</i>	<i>df</i>	<i>p-value</i>
4.2	2	0.1

Hasil uji log-rank diperoleh nilai statistik *chi-square* sebesar 4.2 dengan *p-value* sebesar 0,1. Dengan taraf signifikansi 5%, hasil ini menunjukkan bahwa tidak terdapat perbedaan yang signifikan secara statistik antara kurva *survival* pasien berdasarkan tipe histologi. Meskipun demikian, perbedaan pola kurva secara visual tetap terlihat, yang mengindikasikan adanya kecenderungan perbedaan ketahanan hidup antar tipe histologi.

4.2.7 Kurva *Survival* Komorbiditas dan Uji Log-Rank



Gambar 4.13 Kurva *Survival* Kaplan–Meier Faktor Komorbiditas

Gambar 4.13 menunjukkan kurva *survival* Kaplan–Meier pasien kanker paru-paru NSCLC berdasarkan status komorbiditas. Secara visual, kurva *survival* pasien dengan komorbiditas dan tanpa komorbiditas terlihat relatif berdekatan sepanjang periode pengamatan, yang menunjukkan pola ketahanan hidup yang serupa antara kedua kelompok. Untuk menguji perbedaan kurva *survival* tersebut, dilakukan uji log-rank .

Tabel 4.6 Uji Log-Rank Faktor Komorbiditas

Log-Rank	<i>df</i>	<i>p-value</i>
0.1	1	0.8

Hasil uji log-rank menunjukkan nilai statistik *chi-square* sebesar 0.1 dengan *p-value* sebesar 0,8. Dengan taraf signifikansi 5%, hasil ini menunjukkan bahwa tidak terdapat perbedaan yang signifikan secara statistik antara kurva *survival* pasien dengan dan tanpa komorbiditas. Hal ini mengindikasikan bahwa secara univariat, status komorbiditas belum menunjukkan perbedaan probabilitas ketahanan hidup yang signifikan pada pasien kanker paru-paru NSCLC.

4.3 Uji Asumsi Proportional Hazard

Pengujian asumsi *proportional hazard* dapat dilakukan melalui uji visual dan uji formal. Uji visual dilakukan dengan menggunakan pendekatan kurva Kaplan–Meier, sebagaimana telah disajikan sebelumnya. Berdasarkan hasil pengamatan kurva tersebut, tidak terdapat variabel yang memenuhi asumsi *proportional hazard* karena garis *survival* antar kelompok saling berpotongan. Selanjutnya, pengujian asumsi *proportional hazard* dilakukan dengan menggunakan pendekatan uji *Goodness of Fit*. Uji ini bertujuan untuk memberikan dasar pengambilan keputusan yang lebih objektif. Melalui pengujian tersebut, dapat

diidentifikasi faktor-faktor yang diduga memengaruhi ketahanan hidup pasien kanker NSCLC yang memenuhi asumsi *proportional hazard*, serta faktor-faktor yang tidak memenuhi asumsi tersebut. Hasil uji *Goodness of Fit* untuk seluruh faktor yang diduga berpengaruh terhadap ketahanan hidup pasien kanker NSCLC disajikan sebagai berikut.

Tabel 4.7 Hasil Uji *Goodness of Fit*

Variabel	<i>p-value</i>	Keputusan
Usia (X_1)	0.089	Terima H_0
Jenis Kelamin (X_2)	0.363	Terima H_0
Kebiasaan Merokok (X_3)	0.218	Terima H_0
Stadium Kanker (X_4)	0.496	Terima H_0
Hasil Histologi (X_5)	0.538	Terima H_0
Ada Komorbiditas (X_6)	0.586	Terima H_0

Berdasarkan hasil pengujian pada Tabel 4.7, seluruh variabel memiliki nilai *p-value* lebih besar dari 0,05. Dengan taraf signifikansi 5%, keputusan yang diperoleh adalah gagal menolak H_0 untuk setiap variabel maupun secara global. Hal ini berarti tidak terdapat bukti yang cukup untuk menyatakan adanya pelanggaran asumsi *proportional hazard*. Dengan demikian, dapat disimpulkan bahwa asumsi *proportional hazard* terpenuhi dan model regresi *Cox* layak digunakan untuk analisis lebih lanjut.

4.4 Pemodelan Regresi *Cox Proportional Hazard* Pasien Kanker NSCLC

4.4.1 Estimasi Parameter dengan *Maximum Likelihood Estimation*

Setelah uji asumsi *proportional hazard* selesai dilakukan, tahap berikutnya adalah membangun model regresi *Cox Proportional Hazard* dengan estimasi *Maximum Likelihood* menggunakan fungsi *log partial likelihood* pada Persamaan

(2.16). Estimasi parameter dilakukan secara iteratif menggunakan metode Newton–Raphson dan mencapai kondisi konvergen pada iterasi ke-4.

Tabel 4.8 Hasil Uji Asumsi *Proportional Hazard*

Variabel	$coef(\beta)$	$Wald$	sig	$Exp(\beta)$
Kelompok_Usia(1) (X_1)				
Kelompok_Usia(2) (X_1)	-0.05019	0.017424	0.895	0.95105
Kelompok_Usia(3) (X_1)	0.44876	1.44	0.230	1.56637
Laki-Laki (X_2)	0.28167	1.887876	0.169	1.32534
Perempuan (X_2)				
Mantan Perokok (X_3)	0.35634	1.147041	0.284	1.42809
Perokok Aktif (X_3)	0.07370	0.047089	0.828	1.07648
Tidak Merokok (X_3)				
III (X_4)				
IV (X_4)	0.88705	24.098281	0.00000091	2.42796
<i>Large Cell Carcinoma</i> (X_5)	-0.46023	0.602176	0.438	0.63114
<i>Squamous</i> (X_5)	0.37184	4.056196	0.044	1.45040
<i>Adenocarcinoma</i> (X_5)				
Ada Komorbid (X_6)	-0.13608	0.316969	0.573	0.87278
Tidak Komorbid (X_6)				

Dalam regresi *Cox Proportional Hazard*, variabel kategorik dengan k kategori akan direpresentasikan ke dalam bentuk variabel *dummy* sebanyak $k - 1$. Satu kategori ditetapkan sebagai kategori referensi (*baseline category*) dan tidak ditampilkan dalam output model karena berfungsi sebagai dasar pembandingan. Oleh karena itu, koefisien dan nilai *hazard ratio* yang muncul pada kategori lainnya merupakan perbandingan terhadap kategori referensi tersebut. Berdasarkan *output* regresi *Cox Proportional Hazard* pada Tabel 4.8, model pertama dapat dituliskan dengan Persamaan (2.10) sebagai berikut:

$$\begin{aligned}
h(t | X) = h_0(t) \cdot \exp(& - 0.050 \cdot \text{Kelompok Usia}_2 + 0.449 \\
& \cdot \text{Kelompok Usia}_3 + 0.282 \cdot \text{Jenis Kelamin}_1 + 0.356 \\
& \cdot \text{Merokok_Mantan Perokok} + 0.074 \cdot \text{Merokok_Aktif} \\
& + 0.887 \cdot \text{StadiumIV} - 0.460 \cdot \text{Histologi_LargeCell} + 0.372 \\
& \cdot \text{Histologi_Squamous} - 0.136 \cdot \text{AdaKomorbiditas1})
\end{aligned}$$

Setelah model regresi *Cox Proportional Hazard* diperoleh, tahap selanjutnya adalah melakukan pengujian parameter model. Pengujian ini dilakukan dalam dua tahap, yaitu secara simultan menggunakan *Likelihood Ratio Test* dan secara parsial menggunakan *Wald Test*. Pengujian simultan dilakukan berdasarkan hipotesis dan Persamaan 2.24).

Statistik uji G merupakan *Likelihood Ratio Test*, di mana nilai G mengikuti distribusi χ^2 dengan derajat kebebasan v , yaitu banyaknya parameter yang diuji tanpa termasuk β_0 . Hipotesis nol (H_0) ditolak jika $G > \chi^2_{v,\alpha}$ atau jika $p\text{-value}$ (signifikansi) $< \alpha(5\%)$. Daerah kritis ditentukan sebagai semua nilai G yang lebih besar dari $\chi^2_{v,\alpha}$ atau $p\text{-value}$ lebih kecil dari 0.05, dengan derajat kebebasan v .

Tabel 4.9 Hasil *Likelihood Ratio Test*

Uji Serentak	G^2	df	$p\text{-value (Sig)}$	Keputusan H_0
	41.08	9	5×10^{-6}	Tolak H_0

Tabel 4.9 menunjukkan hasil *Likelihood Ratio Test* untuk model regresi *Cox Proportional Hazard*. Dari tabel tersebut didapatkan nilai G^2 sebesar 41.08 dengan derajat kebebasan 9, sehingga memberikan keputusan Tolak H_0 secara serentak. Hal ini menunjukkan bahwa koefisien variabel yang dimasukkan signifikan secara bersama-sama terhadap model regresi *Cox*. Nilai signifikansi ($p\text{-value}$) yang diperoleh sebesar 5×10^{-6} sehingga memberikan kesimpulan bahwa

minimal terdapat satu variabel bebas yang secara simultan mempengaruhi risiko kematian pasien.

Setelah dilakukan pengujian secara serentak menggunakan *Likelihood Ratio Test*, langkah berikutnya adalah melakukan pengujian parameter secara parsial dengan menggunakan *Wald Test*. Pengujian parsial dilakukan sesuai dengan hipotesis dan Persamaan (2.19).

Tabel 4.10 Hasil Uji Parsial

Variabel	<i>Wald</i>	<i>sig</i>	Keputusan
Kelompok_Usia(2)	0.017424	0.895	Terima H_0
Kelompok_Usia(3)	1.44	0.230	Terima H_0
Laki-Laki	1.887876	0.169	Terima H_0
Mantan Perokok	1.147041	0.284	Terima H_0
Perokok Aktif	0.047089	0.828	Terima H_0
IV	24.098281	9.14×10^{-7}	Tolak H_0
<i>Large Cell Carcinoma</i>	0.602176	0.438	Terima H_0
<i>Squamous</i>	4.056196	0.044	Tolak H_0
Ada Komorbid	0.316969	0.573	Terima H_0

Berdasarkan Tabel 4.10, dapat diketahui bahwa variabel yang berpengaruh secara signifikan ($p \leq 0,05$) adalah Stadium Kanker IV dan Histologi *Squamous*, sedangkan variabel lainnya seperti Kelompok Usia, Jenis Kelamin, Kebiasaan Merokok, Histologi *Large Cell*, dan Komorbiditas tidak menunjukkan pengaruh yang signifikan. Karena terdapat beberapa variabel yang tidak signifikan, variabel-variabel tersebut selanjutnya dieliminasi dari model awal. Variabel yang terbukti signifikan pada model pertama kemudian digunakan kembali untuk membentuk model regresi *Cox Proportional Hazard* yang baru, sehingga diperoleh hasil estimasi model kedua sebagai berikut.

Tabel 4.11 Hasil Regresi *Cox Proportional Hazard* Kedua

Variabel	$coef(\beta)$	Wald	sig	$Exp(\beta)$
IV	0.8314	21.566736	3.29×10^{-6}	2.2966
<i>Large Cell Carcinoma</i>	-0.3236	0.301401	0.58267	0.7235
<i>Squamous</i>	0.5437	9.498724	0.00206	1.7224

Berdasarkan *output* regresi yang diperoleh pada Tabel 4.11, model kedua dapat dituliskan dengan Persamaan (2.10) adalah sebagai berikut:

$$\begin{aligned}
 h(t | X) = & h_0(t) \cdot \exp(0.8314 \cdot \text{Stadium IV} - 0.3236 \\
 & \cdot \text{Histologi_Large Cell Carcinoma} + 0.5437 \\
 & \cdot \text{Histologi_Squamous})
 \end{aligned}$$

Setelah membentuk model kedua, langkah berikutnya adalah melakukan pengujian parameter secara serentak dan parsial kembali. Hipotesis, statistik uji, dan daerah kritis yang digunakan sama seperti pada pengujian parameter model pertama. Hasil uji *likelihood ratio* untuk model kedua diperoleh sebagai berikut.

Tabel 4.12 Hasil *Likelihood Ratio Test* Model Kedua

Uji Serentak	G^2	df	p-value (Sig)	Keputusan H_0
	25.82	3	1×10^{-5}	Tolak H_0

Tabel 4.12 menunjukkan hasil *Likelihood Ratio Test* untuk model regresi *Cox Proportional Hazard* kedua. Dari tabel tersebut diperoleh nilai G^2 sebesar 25,82 dengan derajat kebebasan 3, sehingga memberikan keputusan Tolak H_0 secara serentak. Hal ini menunjukkan bahwa koefisien variabel yang dimasukkan dalam model signifikan secara bersama-sama. Selain itu, nilai signifikansi yang diperoleh sebesar 1×10^{-5} , sehingga dapat disimpulkan bahwa minimal terdapat satu variabel bebas yang secara simultan mempengaruhi risiko kematian pasien.

Setelah melakukan pengujian secara serentak menggunakan *Likelihood Ratio Test* di atas, langkah berikutnya adalah melakukan pengujian parameter secara parsial dengan menggunakan *Wald Test*. Hipotesis, statistik uji, dan daerah kritis yang digunakan sama seperti pada pengujian model pertama. Hasil uji *Wald* untuk model kedua diperoleh sebagai berikut.

Tabel 4.13 Hasil Uji Parsial Model Kedua

Variabel	<i>Wald</i>	<i>sig</i>	Keputusan
IV	21.566736	3.29×10^{-6}	Tolak H_0
<i>Large Cell Carcinoma</i>	0.301401	0.58267	Terima H_0
<i>Squamous</i>	9.498724	0.00206	Tolak H_0

Berdasarkan Tabel 4.13, dapat diketahui bahwa variabel yang berpengaruh secara signifikan ($p \leq 0,05$) adalah Stadium Kanker IV dan Histologi *Squamous*, sehingga dapat disimpulkan bahwa variabel yang berpengaruh signifikan terhadap risiko kematian adalah Stadium IV dan Histologi *Squamous*. Namun, variabel Histologi tetap dipertahankan dalam model sebagai satu kesatuan variabel kategorik. Berdasarkan hasil pengujian simultan dan parsial yang telah dilakukan, dapat disimpulkan bahwa model kedua layak digunakan karena variabel-variabel independen dalam persamaan model terbukti berpengaruh signifikan terhadap variabel dependen. Oleh sebab itu, persamaan model tersebut selanjutnya dapat diinterpretasikan.

4.4.2 Estimasi Parameter dengan *Firth's Penalized Likelihood*

Setelah model *Cox Proportional Hazard* pertama diestimasi menggunakan *Maximum Likelihood Estimation*, dilakukan pendekatan alternatif dengan *Firth's Penalized Likelihood*. Metode ini mempertahankan struktur model *Cox* yang sama, namun menambahkan penalti pada *likelihood*. Model *Cox Proportional Hazard* dengan estimasi *Firth's Penalized Likelihood* dibangun menggunakan

variabel independen yang sama seperti pada estimasi *Maximum Likelihood* berdasarkan fungsi *penalized log-likelihood* pada Persamaan (2.23). Estimasi parameter dilakukan secara iteratif menggunakan metode Newton–Raphson dan mencapai kondisi konvergen pada iterasi ke-8. Hasil estimasi model dengan FPL ditampilkan pada Tabel 4.14.

Tabel 4.14 Hasil Regresi CoxPH dengan Estimasi *Firth's Penalized Likelihood*

Variabel	<i>Coef</i> (β)	<i>exp</i> (β)	p-value	Keputusan
Kelompok_Usia(2)	-0.0857	0.9179	0.8193	Terima H_0
Kelompok_Usia(3)	0.4056	1.5001	0.2482	Terima H_0
Laki-Laki	0.2737	1.3149	0.1735	Terima H_0
Mantan Perokok	0.3281	1.3884	0.3024	Terima H_0
Perokok Aktif	0.0489	1.0501	0.8831	Terima H_0
Stadium IV	0.8851	2.4233	9.52×10^{-7}	Tolak H_0
<i>Large Cell Carcinoma</i>	-0.3129	0.7314	0.5525	Terima H_0
<i>Squamous</i>	0.3709	1.4491	0.0445	Tolak H_0
Ada Komorbid	-0.1519	0.8591	0.5316	Terima H_0

Setelah model regresi *Cox Proportional Hazard* dengan Estimasi FPL diperoleh, selanjutnya adalah melakukan pengujian parameter model dilakukan menggunakan *Penalized Likelihood Ratio Test* yang digunakan untuk menguji secara serentak apakah minimal ada satu variabel yang berpengaruh signifikan terhadap *hazard* secara keseluruhan. Untuk pengujian serentak, hipotesis yang digunakan adalah sebagai berikut.

H_0 : Tidak ada variabel yang berpengaruh signifikan.

H_1 : Minimal ada satu variabel yang berpengaruh signifikan.

Tabel 4.15 Hasil *Penalized Likelihood Ratio Test*

Uji Serentak	G^2	df	$p - value$ (<i>Sig</i>)	Keputusan H_0
<i>Penalized Likelihood Ratio Test</i>	40.09609	9	7.30×10^{-6}	Tolak H_0

Hasil *Penalized Likelihood Ratio Test* untuk model *Cox Proportional Hazard* dengan Estimasi FPL ditampilkan pada Tabel yaitu sebesar 40,09609 dengan derajat kebebasan 3 dan $p\text{-value } 7.30 \times 10^{-6}$, sehingga H_0 ditolak.

Selanjutnya dilakukan pengujian parameter secara parsial menggunakan *Profile Penalized Likelihood*. Pengujian ini bertujuan untuk mengetahui pengaruh masing-masing variabel independen terhadap *hazard*. Berdasarkan hasil pada Tabel 4.14, dapat diketahui bahwa variabel yang berpengaruh signifikan ($p \leq 0,05$) adalah Stadium IV dan Histologi Squamous, sedangkan variabel lainnya tidak menunjukkan pengaruh yang signifikan terhadap waktu *survival* pasien.

Selanjutnya, model *Cox Proportional Hazard* dengan estimasi *Firth's Penalized Likelihood* dibentuk kembali dan kemudian model yang dihasilkan diuji kembali untuk memastikan signifikansi model secara keseluruhan. Pengujian serentak menggunakan *Penalized Likelihood Ratio Test* menghasilkan nilai statistik uji G^2 , derajat kebebasan, dan $p\text{-value}$ sebagai berikut:

Tabel 4.16 *Penalized Likelihood Ratio Test Model Kedua*

Uji Serentak	G^2	df	$p - value (Sig)$	Keputusan H_0
<i>Penalized Likelihood Ratio Test</i>	25.48867	3	1.22×10^{-5}	Tolak H_0

Berdasarkan hasil tersebut, diperoleh $p\text{-value} < 0.05$ sehingga H_0 ditolak. Hal ini menunjukkan bahwa model kedua dengan estimasi FPL signifikan secara keseluruhan. Selanjutnya, pengujian parsial menunjukkan bahwa kedua variabel, yaitu Stadium IV dan Histologi *Squamous*, tetap signifikan terhadap *hazard*.

Tabel 4.17 Uji Parsial Model Kedua

Variabel	<i>Coef</i> (β)	<i>exp</i> (β)	p-value	Keputusan
Stadium IV	0.8295	2.2922	3.33×10^{-6}	Tolak H_0
<i>Large Cell Carcinoma</i>	-0.1772	0.8376	0.7392	Terima H_0
<i>Squamous</i>	0.5441	1.7230	0.00227	Tolak H_0

Berdasarkan Tabel 4.17, dapat diketahui bahwa variabel yang berpengaruh secara signifikan ($p \leq 0,05$) adalah Stadium Kanker IV dan Histologi *Squamous*, sehingga dapat disimpulkan bahwa variabel yang berpengaruh signifikan terhadap risiko kematian adalah Stadium IV dan Histologi *Squamous*. Berdasarkan hasil estimasi pada Tabel 4.17, model *Cox Proportional Hazard* dengan Estimasi FPL dapat dituliskan sebagai berikut: $h(t | X) = h_0(t) \cdot \exp(0.8295 \cdot \text{Stadium IV} - 0.1772 \cdot \text{Histologi Large Cell} + 0.5441 \cdot \text{Histologi Squamous})$.

4.5 Perbandingan Model *Cox Proportional Hazard*

Pada subbab ini dilakukan perbandingan hasil estimasi model *Cox Proportional Hazard* menggunakan dua pendekatan, yaitu *Maximum Likelihood Estimation* dan *Firth's Penalized Likelihood*. Perbandingan ini dilakukan untuk menilai perbedaan koefisien, signifikansi, serta kecocokan model yang diukur menggunakan *Akaike Information Criterion* (AIC) berdasarkan Persamaan (2.24), kalibrasi model menggunakan *calibration slope* berdasarkan Persamaan (2.25), serta kemampuan diskriminasi model menggunakan *Area Under the Curve* (AUC)/*concordance index* berdasarkan Persamaan (2.26). Hasil perbandingan ditampilkan pada Tabel 4.18.

Tabel 4.18 Perbandingan Model CoxPH dengan MLE dan FPL

Variabel	MLE			FPL		
	<i>Coef</i> (β)	<i>HR</i>	<i>p - value</i>	<i>Coef</i> (β)	<i>HR</i>	<i>p - value</i>
Stadium						
Stadium III	(Kategori Referensi)					
Stadium IV	0.8314	2.2966	3.42×10^{-6}	0.829524	2.292228	3.3×10^{-6}
Histologi						
<i>Adenocarcinoma</i>	(Kategori Referensi)					
<i>Large Cell Carcinoma</i>	-0.3236	0.7235	0.58267	-0.177199	0.837613	0.739
<i>Squamous</i>	0.5437	1.7224	0.00206	0.544050	1.722971	0.002
	<i>AIC</i> = 1459.943			<i>AIC</i> = 1445.332		

Berdasarkan Tabel 4.16 perbandingan antara estimasi MLE dan FPL, terlihat bahwa kedua metode menghasilkan koefisien (β), *hazard ratio* (*HR*), dan *p-value* yang hampir sama untuk masing-masing variabel. Untuk variabel Stadium IV, koefisien β sebesar 0.831 pada MLE dan 0,830 pada FPL, dengan *HR* masing-masing 2,297 dan 2,292, serta *p-value* sangat signifikan ($< 0,001$).

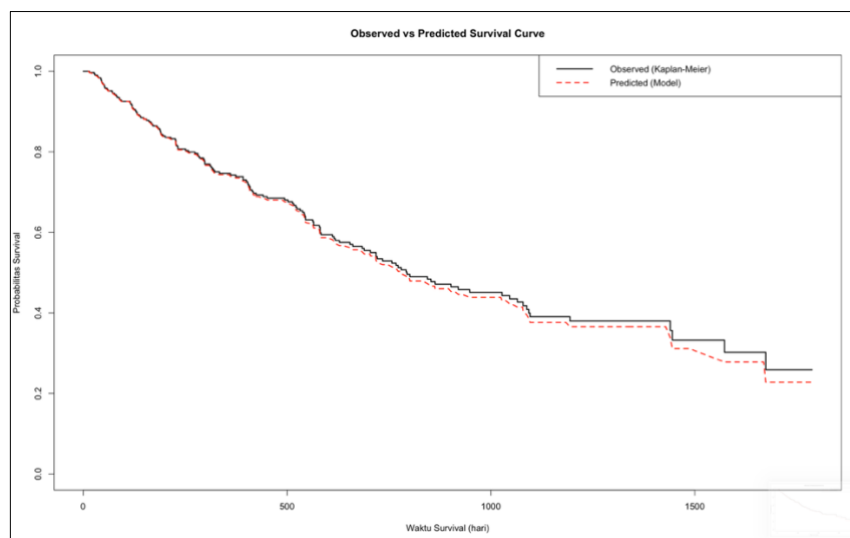
Sedangkan untuk variabel Histologi *Squamous*, kedua metode juga menunjukkan hasil yang serupa, dengan β sekitar 0,544, *HR* sekitar 1,72, dan *p-value* signifikan ($p < 0,01$). Ini mengindikasikan bahwa pasien dengan histologi *Squamous* memiliki risiko kematian sekitar 1,7 kali lebih tinggi dibandingkan *baseline* (*Adenocarcinoma*), baik menggunakan MLE maupun FPL. Nilai kriteria model, *AIC* Firth (1445,332) lebih rendah dibandingkan *AIC* MLE (1459,943), yang mengindikasikan bahwa model FPL memberikan *fit* yang lebih baik terhadap data.

Selain membandingkan kecocokan model berdasarkan nilai *AIC*, evaluasi performa prediksi model juga dilakukan melalui pengukuran kalibrasi dan diskriminasi. Kalibrasi model digunakan untuk menilai kesesuaian antara probabilitas *survival* hasil prediksi model dengan data observasi aktual, sedangkan diskriminasi digunakan untuk menilai kemampuan model dalam membedakan pasien dengan risiko kematian yang lebih tinggi dan lebih rendah.

Tabel 4.19 Perbandingan *Calibration Slope* dan *C-Index* Model CoxPH

Model	<i>Calibration Slope</i>	<i>C-index</i>
MLE	1.000	0.622
Firth	0.485	0.622

Berdasarkan hasil evaluasi performa model pada Tabel 4.19, model *Cox Proportional Hazard* dengan estimasi MLE menghasilkan nilai *calibration slope* sebesar 1 yang menunjukkan bahwa probabilitas *survival* hasil prediksi model memiliki kesesuaian yang baik dengan data *survival* observasi. Sementara itu, model *Cox Proportional Hazard* dengan estimasi FPL menghasilkan nilai *calibration slope* sebesar 0.485. Nilai tersebut mengindikasikan bahwa model masih memiliki kecenderungan *overfitting*, yaitu kondisi ketika prediksi risiko yang dihasilkan model cenderung lebih ekstrem dibandingkan pola *survival* aktual pada data observasi.

Gambar 4.14 Kurva *Survival* Observasi dan Prediksi Model *Cox Proportional Hazard*

Berdasarkan Gambar 4.14, kurva *survival* hasil prediksi model *Cox Proportional Hazard* dengan estimasi *Maximum Likelihood Estimation* (MLE) menunjukkan pola yang sangat mendekati kurva *survival* empiris Kaplan–Meier. Hal ini menunjukkan bahwa model memiliki kemampuan kalibrasi yang baik dalam

merepresentasikan pola *survival* pasien NSCLC. Perbedaan antara kurva observasi dan kurva prediksi terlihat relatif kecil pada sebagian besar interval waktu pengamatan. Hasil visual tersebut juga sesuai dengan nilai *calibration slope* sebesar 1 yang menunjukkan bahwa probabilitas *survival* hasil prediksi model MLE memiliki kesesuaian yang baik dengan data observasi aktual.

Dari aspek diskriminasi, kedua model menunjukkan kemampuan yang relatif sama dalam membedakan pasien dengan risiko kematian yang lebih tinggi dan lebih rendah. Hal ini ditunjukkan oleh nilai *concordance index* (C-index) sebesar 0,622 pada kedua model. Nilai tersebut menunjukkan bahwa model memiliki kemampuan diskriminasi yang cukup baik dalam memprediksi urutan risiko kematian pasien NSCLC.

Secara umum, hasil perbandingan menunjukkan bahwa metode MLE dan FPL memberikan hasil estimasi yang relatif konsisten pada model *Cox Proportional Hazard*. Model Firth menunjukkan kecocokan model yang lebih baik berdasarkan nilai AIC yang lebih rendah, sedangkan model MLE menunjukkan performa kalibrasi yang lebih baik berdasarkan nilai *calibration slope* dan kesesuaian visual antara kurva *survival* prediksi dan kurva Kaplan–Meier observasi. Sementara itu, kedua model memiliki kemampuan diskriminasi yang relatif sama dalam membedakan risiko kematian pasien NSCLC.

4.6 *Hazard Ratio*

Setelah dilakukan perbandingan model *Cox Proportional Hazard* menggunakan estimasi Maximum Likelihood Estimation (MLE) dan *Firth's Penalized Likelihood*, model dengan estimasi *Maximum Likelihood Estimation*

(MLE) dipilih sebagai model yang lebih baik berdasarkan hasil evaluasi performa kalibrasi dan kesesuaian visual kurva *survival* prediksi terhadap kurva Kaplan–Meier observasi. Selanjutnya dilakukan interpretasi model menggunakan *hazard ratio* (HR). *Hazard ratio* digunakan untuk mengetahui besarnya risiko kematian antar kelompok pada setiap variabel yang berpengaruh terhadap ketahanan hidup pasien kanker paru-paru. Nilai *hazard ratio* sebesar 1 ($HR = 1$) menunjukkan tidak adanya pengaruh antara paparan dan kejadian. Nilai *hazard ratio* yang lebih besar dari 1 ($HR > 1$) menunjukkan peningkatan risiko kejadian pada kelompok yang terpapar, sedangkan nilai yang lebih kecil dari 1 ($HR < 1$) menunjukkan risiko yang lebih rendah dibandingkan kategori referensi (Kleinbaum & Klein, 2012). Model regresi *Cox Proportional Hazard* yang lebih baik diperoleh menggunakan estimasi *Maximum Likelihood Estimation* (MLE) adalah sebagai berikut.

$$h(t | X) = h_0(t) \cdot \exp(0.8314 \cdot \text{Stadium IV} - 0.3236 \\ \cdot \text{Histologi_Large Cell} + 0.5437 \cdot \text{Histologi_Squamous})$$

Hazard ratio dari variabel tersebut dapat diinterpretasikan menggunakan $\text{Exp}(\beta)$ pada Tabel 4.18.

1. *Hazard ratio* yang diperoleh menunjukkan bahwa Stadium IV memiliki nilai $\text{Exp}(\beta)$ sebesar 2,2966. Karena $2,2966 \approx 2,3 > 1$, dapat diartikan bahwa pasien dengan Stadium IV memiliki risiko kematian sekitar 2.3 kali lebih tinggi dibandingkan pasien dengan Stadium III sebagai kategori referensi.
2. *Hazard ratio* yang diperoleh menunjukkan bahwa histologi Squamous memiliki nilai $\text{Exp}(\beta)$ sebesar 1,7224. Karena $1,7224 \approx 1,7 > 1$, dapat diartikan bahwa pasien dengan histologi *Squamous* memiliki risiko

kematian sekitar 1,7 kali lebih tinggi dibandingkan pasien dengan *Adenocarcinoma* sebagai kategori referensi.

4.7 Integrasi Al-Quran

Integrasi nilai-nilai Al-Qur'an dalam penelitian mengenai ketahanan hidup pasien yang menderita penyakit dapat dikaitkan dengan konsep waktu dan kejadian yang menjadi inti dari analisis *survival*. Dalam Islam dijelaskan bahwa setiap peristiwa yang terjadi di dunia berada dalam ketentuan Allah SWT. Hal ini sebagaimana firman Allah SWT dalam Surah Al-Hadid ayat 22 (Kemenag RI, 2022):

“Tidak ada bencana apa pun yang menimpa di bumi dan tidak pula pada dirimu sendiri, melainkan telah tertulis dalam Kitab (Lauhulmahfuz) sebelum Kami mewujudkannya. Sesungguhnya yang demikian itu mudah bagi Allah.”

Ayat tersebut menunjukkan bahwa segala sesuatu yang terjadi telah berada dalam ketetapan Allah SWT. Meskipun demikian, manusia tetap diberi akal untuk meneliti dan memahami berbagai peristiwa yang terjadi di sekitarnya.

Selain itu, penelitian ini juga membandingkan dua metode estimasi, yaitu *Maximum Likelihood Estimation* dan *Firth's Penalized Likelihood*, untuk memperoleh model yang paling sesuai. Upaya memilih metode yang paling baik ini sejalan dengan firman Allah SWT dalam Surah Az-Zumar ayat 18 yang menyatakan bahwa orang-orang yang berakal adalah mereka yang mendengarkan berbagai perkataan kemudian mengikuti yang paling baik di antaranya.

Dalam sejarah Islam, kesabaran dalam menghadapi penyakit dicontohkan oleh Nabi Ayyub AS yang diuji dengan penyakit berat dalam waktu yang lama

namun tetap bersabar dan berdoa kepada Allah SWT. Hal ini sejalan dengan firman Allah SWT dalam Surah Asy-Syu'ara ayat 80 yang berbunyi (Kemenag RI, 2022):

“Dan apabila aku sakit, Dialah (Allah) yang menyembuhkan aku”.

Selain itu, Rasulullah SAW juga bersabda:

“Allah tidak menurunkan suatu penyakit melainkan Dia juga menurunkan obatnya”.

Oleh karena itu, manusia dianjurkan untuk menjaga kesehatan dan menghindari kebiasaan yang dapat membahayakan tubuh, seperti merokok. Salah satu penyakit berbahaya yang berkaitan dengan kebiasaan merokok adalah kanker paru-paru, yang merupakan salah satu penyebab utama kematian di dunia.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut:

1. Perbandingan model *Cox Proportional Hazard* menggunakan metode estimasi Maximum Likelihood Estimation (MLE) dan *Firth's Penalized Likelihood* menunjukkan bahwa model Firth menghasilkan nilai *Akaike Information Criterion* (AIC) yang lebih rendah dibandingkan model MLE, yaitu sebesar 1445,332 pada model Firth dan 1459,943 pada model MLE. Namun demikian, model MLE menunjukkan performa kalibrasi yang lebih baik berdasarkan nilai *calibration slope* sebesar 1 serta kesesuaian visual antara kurva *survival* prediksi dan kurva Kaplan–Meier observasi. Selain itu, kedua model *memiliki* kemampuan diskriminasi yang relatif sama dengan nilai *concordance index* (C-index) sebesar 0,622. Oleh karena itu, model Cox Proportional Hazard dengan estimasi *Maximum Likelihood Estimation* (MLE) dipilih sebagai model yang lebih baik dalam penelitian ini.
2. Model terbaik yang didapatkan dapat dituliskan sebagai berikut: $h(t | X) = h_0(t) \cdot \exp(0.8314 \times Stadium\ I - 0.3236 \cdot Histologi_Large\ Cell + 0.5437 \cdot Histologi_Squamous)$. Berdasarkan model *Cox Proportional Hazard* dengan estimasi *Maximum Likelihood Estimation* (MLE) yang diperoleh, variabel yang berpengaruh signifikan terhadap ketahanan hidup pasien NSCLC adalah stadium kanker dan jenis histologi. Nilai *hazard ratio* menunjukkan bahwa pasien dengan Stadium IV memiliki risiko kematian

sekitar 2,3 kali lebih tinggi dibandingkan pasien dengan Stadium III, sedangkan pasien dengan histologi *Squamous* memiliki risiko kematian sekitar 1,7 kali lebih tinggi dibandingkan pasien dengan *Adenocarcinoma* sebagai kategori referensi.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, penelitian selanjutnya diharapkan dapat mengembangkan analisis faktor-faktor yang memengaruhi ketahanan hidup pasien NSCLC dengan mempertimbangkan karakteristik klinis maupun demografis lainnya. Selain itu, hasil penelitian ini diharapkan dapat menjadi referensi dalam penerapan analisis *survival*, khususnya model *Cox Proportional Hazard*, pada berbagai kasus kesehatan yang melibatkan data waktu *survival*.

DAFTAR PUSTAKA

- Adhikary, A. C., & Shafiqur Rahman, M. (2021). Firth's Penalized Method In Cox Proportional Hazard Framework For Developing Predictive Models For Sparse Or Heavily Censored Survival Data. *Journal Of Statistical Computation and Simulation*, 91(3), 445–463. <https://doi.org/10.1080/00949655.2020.1817924>
- Afranda, O., Wuryandari, T., & Ispriyanti, D. (2015). Analisis Regresi Kegagalan Proporsional dari Cox Pada Data Waktu Tunggu Sarjana dengan Sensor Tipe I (Studi Kasus di Fakultas Sains dan Matematika Universitas Diponegoro). *Jurnal Gaussian*, 4(3), 621–630. <https://doi.org/doi.org/10.14710/j.gauss.4.3.621-630>
- Altman, D. G. (1991). Categorising Continuous Variables. *British Journal of Cancer*, 64(5), 975–982. <https://doi.org/10.1038/bjc.1991.441>
- American Cancer Society. (2025). *Lung Cancer Statistics | How Common Is Lung Cancer?* <https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html>
- Collett, D. (2015). *Modelling Survival Data in Medical Research* (3rd Ed.). Chapman and Hall/CRC.
- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of The Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202.
- Cox, D. R. (2018). *Analysis of Survival Data*. Chapman and Hall/CRC.
- Dahlan, M. S. (2012). Analisis *Survival*: Dasar-Dasar Teori & Aplikasi dengan Program SPSS. *Jakarta: Epidemiologi Indonesia*.
- Edina, A. I. (2020). *Estimasi Fungsi Baseline Hazard pada Model Cox dengan Metode Maximum Penalized Likelihood* [Institut Teknologi Sepuluh Nopember]. <http://repository.its.ac.id/id/eprint/81090>
- Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 80(1), 27–38.
- Gönen, M., & Heller, G. (2005). Concordance Probability and Discriminatory Power in Proportional Hazards Regression. *Biometrika*, 92(4), 965–970.
- Harrell, F. E. (2015). *Regression Modeling Strategies*. Springer.

- Heinze, G. (1999). The Application of Firth's Procedure To Cox and Logistic Regression (Technical Report 10). Dalam *Technical Report 10*. Department of Medical Computer Sciences, Section of Clinical Biometrics.
- Heinze, G., & Schemper, M. (2001). A Solution to The Problem of Monotone Likelihood in Cox Regression. *Biometrics*, 57(1), 114–119.
- Hosmer, D. W., Lemeshow, S., & May, S. (2008). Applied Survival Analysis. *Wiley Series in Probability and Statistics*, 60.
- Jakarta, D. K. P. D. K. I. (2023). *Profil Kesehatan Provinsi DKI Jakarta Tahun 2023*. Dinas Kesehatan Provinsi DKI Jakarta. <https://goodstats.id/article/merokok-sampai-tua-menelaah-data-seputar-perokok-di-indonesia-xlpep>
- Kemenag RI. (2022). *Qur'an Kemenag*. <https://quran.kemenag.go.id/>
- Kleinbaum, D. G., & Klein, M. (2012). Introduction to Survival Analysis. *Survival Analysis: A Self-Learning Text*, 1–54.
- Monica, Y. S. A., & Purhadi, P. (2016). Analisis Faktor yang Memengaruhi Laju Kesembuhan Pasien Tuberkulosis Paru di RSUD Dr. Soetomo Tahun 2015 Menggunakan Regresi Weibull dan Regresi Cox Proportional Hazard. *Jurnal Sains dan Seni ITS*, 5(2), 133245.
- Ríos-Sánchez, B., Viguera, G., García, A., Gómez-Bravo, D., Menasalvas, E., Torrente, M., Parejo, C., Aisopos, F., Vogiatzis, D., Purohit, D., Provencio, M., Vidal, M.-E., & Rodríguez-González, A. (2024). A Lung Cancer Diagnosis and Treatment Dataset With Geno- and Phenotypical Characteristics of The Patient. *Data in Brief*, 57, 111167. <https://doi.org/https://doi.org/10.1016/j.dib.2024.111167>
- Robot, R. Y., Durry, M. F., & Kairupan, C. F. (2021). Morfologi, Patogenesis, dan Imunoterapi Kanker Paru Tipe Adenokarsinoma. *Medical Scope Journal*, 3(1), 74–82.
- Royston, P., & Altman, D. G. (2013). External Validation of A Cox Prognostic Model: Principles and Methods. *BMC Medical Research Methodology*, 13(1), 33.

- Tas, F., Ciftci, R., Kilic, L., & Karabulut, S. (2013). Age Is A Prognostic Factor Affecting Survival in Lung Cancer Patients. *Oncology Letters*, 6(5), 1507–1513.
- Verta, J. (2019). *Different Types of Lung Cancer and Treatment*. Lung Cancer Center. <https://www.lungcancercenter.com/lung-cancer/types/>
- Wao, H., Mhaskar, R., Kumar, A., Miladinovic, B., & Djulbegovic, B. (2013). Survival of Patients With Non-Small Cell Lung Cancer Without Treatment: A Systematic Review and Meta-Analysis. *Systematic Reviews*, 2(1), 10. <https://doi.org/10.1186/2046-4053-2-10>
- Wheeler, M., Karanth, S. D., Mehta, H. J., Yang, D., Aduse-Poku, L., Washington, C., Hong, Y.-R., Zhang, D., Gould, M. K., & Braithwaite, D. (2023). Survival Differences by Comorbidity Burden Among Patients With Stage I/II Non-Small-Cell Lung Cancer After Thoracoscopic Resection. *Cancers*, 15(7), 2075.
- WHO. (2024, Februari 1). *Global Cancer Burden Growing, Amidst Mounting Need for Services*. <https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services>
- Yasril, & Kasjono, H. S. (2009). *Analisis Multivariat untuk Penelitian Kesehatan*. Mitra Cendekia Press, Jogjakarta.
- Zhang, J., Wang, Y., Hua, T., Wei, X., Jiang, X., Ji, M., Ma, Z., Huang, Y., Wang, H., Du, L., Zhu, M., Xu, L., Wu, W., & Ma, H. (2023). Association of Psychological Distress, Smoking and Genetic Risk With The Incidence of Lung Cancer: A Large Prospective Population-Based Cohort Study. *Frontiers in Oncology*, 13. <https://doi.org/10.3389/fonc.2023.1133668>

LAMPIRAN

Lampiran 1. Data Pasien Kanker Paru-Paru Tahun 2016 – 2020

ID	t	x_1	x_2	x_3	x_4	x_5	x_6
1	232	3	1	2	3	2	1
2	281	3	1	2	3	2	1
3	90	3	1	3	3	2	0
4	128	2	1	2	3	2	1
5	334	3	1	2	3	2	1
6	201	2	0	2	3	2	1
7	116	3	1	1	3	2	0
8	298	3	1	3	3	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
312	13	3	1	2	4	1	1

Lampiran 2. Script RStudio

```
# LOAD PACKAGE
library(readr)
library(dplyr)
library(ggplot2)
library(survival)
library(survminer)

# BACA DATA
KankerData <- read_delim(
  "Downloads/KankerData.csv",
  delim = ";",
  escape_double = FALSE,
  trim_ws = TRUE
)

# PILIH & RENAME VARIABEL
Data_Kanker <- KankerData %>%
  select(
    Gender, AgeGroup, smokinghabit,
    backgroundind, comorind, comorbidity,
    deceasedind, decdate, diagnosisdate,
    inistage, histology
  ) %>%
  rename(
    Jenis_Kelamin = Gender,
    Kelompok_Usia = AgeGroup,
    Kebiasaan_Merokok = smokinghabit,
    Riwayat_Kanker_Keluarga = backgroundind,
    Ada_Komorbidity = comorind,
    Jenis_Komorbidity = comorbidity,
    Meninggal = deceasedind,
    Tanggal_Kematian = decdate,
    Tanggal_Diagnosis = diagnosisdate,
    Stadium_Awal = inistage,
    Hasil_Histologi = histology
  )

# RECODE VARIABEL
Data_Kanker <- Data_Kanker %>%
  mutate(
    Jenis_Kelamin = ifelse(Jenis_Kelamin == "Male", 1, 0),
    Ada_Komorbidity = ifelse(Ada_Komorbidity == "Yes", 1, 0),
    Meninggal = ifelse(Meninggal == "Yes", 1, 0)
  )
Data_Kanker$Kelompok_Usia <- recode(
  Data_Kanker$Kelompok_Usia,
  "[30 - 49]" = 1,
  "[49 - 64]" = 2,
  "[64 - 90]" = 3
)
```



```

Data_Kanker <- Data_Kanker %>%
  mutate(
    Kebiasaan_Merokok = factor(
      Kebiasaan_Merokok,
      levels = c(
        "No Smoker (<100 cigarettes/life time)",
        "Former Smoker (> 1 year)",
        "Active Smoker"
      )
    ),
    Stadium_Awal = factor(
      Stadium_Awal,
      levels = c("IA", "IB", "IIIA", "IIIB", "IIIC", "IVA",
        "IVB")
    )
  )

# KONVERSI FORMAT TANGGAL
Data_Kanker <- Data_Kanker %>%
  mutate(
    Tanggal_Diagnosis = as.Date(Tanggal_Diagnosis, format =
"%d/%m/%Y"),
    Tanggal_Kematian = as.Date(Tanggal_Kematian, format =
"%d/%m/%Y")
  )

# FILTER DATA
Data_Kanker <- Data_Kanker %>%
  filter(
    Hasil_Histologi %in% c(
      "Adenocarcinoma",
      "Squamous",
      "Large cell carcinoma"
    ),
    Tanggal_Diagnosis >= as.Date("2016-01-01"),
    Tanggal_Diagnosis <= as.Date("2020-12-31"),
    Stadium_Awal %in% c("IIIA", "IIIB", "IIIC", "IVA", "IVB")
  )

Data_Kanker <- Data_Kanker %>%
  mutate(
    Stadium_Kanker_Label = case_when(
      Stadium_Awal %in% c("IIIA", "IIIB", "IIIC") ~ "III",
      Stadium_Awal %in% c("IVA", "IVB") ~ "IV"
    ),
    Stadium_Kanker_Label = factor(
      Stadium_Kanker_Label,
      levels = c("III", "IV")
    )
  )

# STATUS EVENT & WAKTU SURVIVAL

```

```

tanggal_akhir <- as.Date("2020-12-31")

Data_Kanker <- Data_Kanker %>%
  mutate(
    Status_Event = ifelse(
      Meninggal == 1 &
      !is.na(Tanggal_Kematian) &
      Tanggal_Kematian <= tanggal_akhir,
      1, 0
    ),
    Waktu_Survival_Hari = ifelse(
      Status_Event == 1,
      as.numeric(Tanggal_Kematian - Tanggal_Diagnosis),
      as.numeric(tanggal_akhir - Tanggal_Diagnosis)
    )
  )

# barplot
# 1. Kelompok Usia
ggplot(Data_Kanker, aes(x = factor(Kelompok_Usia))) +
  geom_bar(fill = "steelblue") +
  geom_text(
    stat = "count",
    aes(label = after_stat(count)),
    vjust = -0.3,
    size = 4
  ) +
  scale_x_discrete(
    labels = c(
      "1" = "30-49",
      "2" = "49-64",
      "3" = "64-90"
    )
  ) +
  labs(
    x = "Kelompok Usia (tahun)",
    y = "Jumlah Pasien",
    title = "Distribusi Pasien Berdasarkan Kelompok Usia"
  ) +
  theme_minimal()

# 4.2
# OBJEK SURVIVAL
surv_object <- Surv(
  time = Data_Kanker$Waktu_Survival_Hari,
  event = Data_Kanker$Status_Event
)

# KURVA KAPLAN-MEIER
km_all <- survfit(surv_object ~ 1, data = Data_Kanker)

```

```

ggsurvplot(
  km_all,
  data = Data_Kanker,
  risk.table = TRUE,
  risk.table.fontsize = 6.5,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Bertahan Hidup",
  title = "Kurva Kaplan-Meier Pasien NSCLC",
  ggtheme = theme_classic() +
    theme(
      text = element_text(size = 18),
      plot.title = element_text(size = 20, face = "bold"),
      axis.title = element_text(size = 18),
      axis.text = element_text(size = 16)
    )
)

# Kaplan Meier & Log-Rank
km_usia <- survfit(
  surv_object ~ factor(Kelompok_Usia),
  data = Data_Kanker
)

ggsurvplot(
  km_usia,
  data = Data_Kanker,
  conf.int = FALSE,
  censor.shape = "+",
  censor.size = 3,
  risk.table = TRUE,
  risk.table.height = 0.25,
  risk.table.fontsize = 6,
  pval = TRUE,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Bertahan Hidup",
  title = "Kurva Kaplan-Meier Berdasarkan Kelompok Usia",
  legend.title = "Kelompok Usia",
  legend.labs = c("30-49 tahun", "49-64 tahun", "64-90 tahun"),
  ggtheme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.7),
      axis.text = element_text(size = 16),
      axis.title = element_text(size = 18),
      plot.title = element_text(hjust = 0.5, face = "bold")
    ),
  tables.theme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.6),
      axis.title = element_blank(),
      axis.text.x = element_text(size = 16),    # <- memperbesar
angka hari
      axis.text.y = element_text(size = 14)
    )
)

```

```

    )
  )

logrank_usia <- survdiff(
  surv_object ~ factor(Kelompok_Usia),
  data = Data_Kanker
)

logrank_usia

# KM berdasarkan jenis kelamin
km_gender <- survfit(
  surv_object ~ factor(Jenis_Kelamin),
  data = Data_Kanker
)

ggsurvplot(
  km_usia,
  data = Data_Kanker,
  conf.int = FALSE,
  censor.shape = "+",
  censor.size = 3,
  risk.table = TRUE,
  risk.table.height = 0.25,
  risk.table.fontsize = 6,
  pval = TRUE,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Bertahan Hidup",
  title = "Kurva Kaplan-Meier Berdasarkan Kelompok Usia",
  legend.title = "Kelompok Usia",
  legend.labs = c("30-49 tahun", "49-64 tahun", "64-90 tahun"),
  ggtheme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.7),
      axis.text = element_text(size = 16),
      axis.title = element_text(size = 18),
      plot.title = element_text(size = 18, hjust = 0.5, face =
"bold")
    ),
  tables.theme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.6),
      axis.title = element_blank(),
      axis.text.x = element_text(size = 16),
      axis.text.y = element_text(size = 16)
    )
)

logrank_usia <- survdiff(
  surv_object ~ factor(Kelompok_Usia),
  data = Data_Kanker
)

```

```

logrank_usia

# KM berdasarkan jenis kelamin
km_gender <- survfit(
  surv_object ~ factor(Jenis_Kelamin),
  data = Data_Kanker
)

ggsurvplot(
  km_gender,
  data = Data_Kanker,
  conf.int = FALSE,
  censor.shape = "+",
  censor.size = 3,
  risk.table = TRUE,
  risk.table.height = 0.25,
  risk.table.fontsize = 6,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Bertahan Hidup",
  title = "Kurva Kaplan-Meier Pasien NSCLC Berdasarkan Jenis
Kelamin",
  legend.title = "Jenis Kelamin",
  legend.labs = c("Perempuan", "Laki-laki"),
  ggtheme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.7),
      axis.text = element_text(size = 16),
      axis.title = element_text(size = 18),
      plot.title = element_text(size = 18, hjust = 0.5, face =
"bold")
    ),
  tables.theme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.6),
      axis.title = element_blank(),
      axis.text.x = element_text(size = 16),
      axis.text.y = element_text(size = 16)
    )
)

logrank_gender <- survdiff(
  surv_object ~ factor(Jenis_Kelamin),
  data = Data_Kanker
)

logrank_gender

# KM berdasarkan status merokok
km_merokok <- survfit(
  surv_object ~ Kebiasaan_Merokok,
  data = Data_Kanker
)

```

```

)

ggsurvplot(
  km_merokok,
  data = Data_Kanker,
  conf.int = FALSE,
  censor.shape = "+",
  censor.size = 3,
  risk.table = TRUE,
  risk.table.height = 0.25,
  risk.table.fontsize = 6,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Bertahan Hidup",
  title = "Kurva Kaplan-Meier Pasien NSCLC Berdasarkan Status
Merokok",
  legend.title = "Status Merokok",
  legend.labs = c("Non Perokok", "Mantan Perokok", "Perokok
Aktif"),
  ggtheme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.7),
      axis.text = element_text(size = 16),
      axis.title = element_text(size = 18),
      plot.title = element_text(size = 18, hjust = 0.5, face =
"bold")
    ),
  tables.theme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.6),
      axis.title = element_blank(),
      axis.text.x = element_text(size = 16),
      axis.text.y = element_text(size = 16)
    )
)

logrank_merokok <- survdiff(
  surv_object ~ Kebiasaan_Merokok,
  data = Data_Kanker
)

logrank_merokok

# Kaplan-Meier Stadium
km_stadium <- survfit(
  surv_object ~ Stadium_Kanker_Label,
  data = Data_Kanker
)

ggsurvplot(
  km_stadium,
  data = Data_Kanker,
  conf.int = FALSE,

```

```

    censor.shape = "+",
    censor.size = 3,
    risk.table = TRUE,
    risk.table.height = 0.25,
    risk.table.fontsize = 6,
    xlab = "Waktu Survival (hari)",
    ylab = "Probabilitas Bertahan Hidup",
    title = "Kurva Kaplan-Meier Berdasarkan Stadium Kanker",
    legend.title = "Stadium Kanker",
    legend.labs = c("Stadium III", "Stadium IV"),
    ggtheme = theme_classic() +
      theme(
        axis.line = element_line(size = 0.7),
        axis.text = element_text(size = 16),
        axis.title = element_text(size = 18),
        plot.title = element_text(size = 18, hjust = 0.5, face =
"bold")
      ),
    tables.theme = theme_classic() +
      theme(
        axis.line = element_line(size = 0.6),
        axis.title = element_blank(),
        axis.text.x = element_text(size = 16),
        axis.text.y = element_text(size = 16)
      )
  )

# Log-rank Stadium
logrank_stadium <- survdiff(
  surv_object ~ Stadium_Kanker_Label,
  data = Data_Kanker
)

logrank_stadium

# Kaplan-Meier Histologi
km_histologi <- survfit(
  surv_object ~ Hasil_Histologi,
  data = Data_Kanker
)

ggsurvplot(
  km_histologi,
  data = Data_Kanker,
  conf.int = FALSE,
  censor.shape = "+",
  censor.size = 3,
  risk.table = TRUE,
  risk.table.height = 0.25,
  risk.table.fontsize = 6,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Bertahan Hidup",

```

```

title = "Kurva Kaplan-Meier Berdasarkan Tipe Histologi",
legend.title = "Tipe Histologi",
legend.labs = c(
  "Adenokarsinoma",
  "Karsinoma Sel Skuamosa",
  "Karsinoma Sel Besar"
),
ggtheme = theme_classic() +
  theme(
    axis.line = element_line(size = 0.7),
    axis.text = element_text(size = 16),
    axis.title = element_text(size = 18),
    plot.title = element_text(size = 18, hjust = 0.5, face =
"bold")
  ),
  tables.theme = theme_classic() +
    theme(
      axis.line = element_line(size = 0.6),
      axis.title = element_blank(),
      axis.text.x = element_text(size = 16),
      axis.text.y = element_text(size = 16)
    )
)

# Log-rank Histologi
logrank_histologi <- survdiff(
  surv_object ~ Hasil_Histologi,
  data = Data_Kanker
)

logrank_histologi

# Kaplan-Meier Komorbiditas
km_komorbid <- survfit(
  surv_object ~ factor(Ada_Komorbiditas),
  data = Data_Kanker
)

ggsurvplot(
  km_komorbid,
  data = Data_Kanker,
  conf.int = FALSE,
  censor.shape = "+",
  censor.size = 3,
  risk.table = TRUE,
  risk.table.height = 0.25,
  risk.table.fontsize = 6,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Bertahan Hidup",
  title = "Kurva Kaplan-Meier Berdasarkan Status Komorbiditas",
  legend.title = "Komorbiditas",
  legend.labs = c("Tidak Ada", "Ada"),

```



```

ggtheme = theme_classic() +
  theme(
    axis.line = element_line(size = 0.7),
    axis.text = element_text(size = 16),
    axis.title = element_text(size = 18),
    plot.title = element_text(size = 18, hjust = 0.5, face =
"bold")
  ),
  tables.theme = theme_classic() +
  theme(
    axis.line = element_line(size = 0.6),
    axis.title = element_blank(),
    axis.text.x = element_text(size = 16),
    axis.text.y = element_text(size = 16)
  )
)

# Log-rank Komorbiditas
logrank_komorbid <- survdiff(
  surv_object ~ factor(Ada_Komorbiditas),
  data = Data_Kanker
)

logrank_komorbid

cox_model <- coxph(
  surv_object ~ factor(Kelompok_Usia) +
    factor(Jenis_Kelamin) +
    Kebiasaan_Merokok +
    Stadium_Kanker_Label +
    Hasil_Histologi +
    factor(Ada_Komorbiditas),
  data = Data_Kanker
)

# Uji Asumsi Proportional Hazard
ph_test <- cox.zph(cox_model)
ph_test

# Plot residual Schoenfeld
ggcoxzph(ph_test)
cox_model <- coxph(
  Surv(Waktu_Survival_Hari, Status_Event) ~
    factor(Kelompok_Usia) +
    factor(Jenis_Kelamin) +
    Kebiasaan_Merokok +
    Stadium_Kanker_Label +
    Hasil_Histologi +
    factor(Ada_Komorbiditas),
  data = Data_Kanker
)

```

```

summary(cox_model)

cox_model2 <- coxph(
  Surv(Waktu_Survival_Hari, Status_Event) ~
    Stadium_Kanker_Label +
    Hasil_Histologi,
  data = Data_Kanker
)

summary(cox_model2)

library(coxphf)
cox_firth <- coxphf(
  Surv(Waktu_Survival_Hari, Status_Event) ~
    factor(Kelompok_Usia) +
    factor(Jenis_Kelamin) +
    Kebiasaan_Merokok +
    Stadium_Kanker_Label +
    Hasil_Histologi +
    factor(Ada_Komorbidity),
  data = Data_Kanker,
  pl = TRUE,
  maxit = 1000,
  maxstep = 0.1
)

summary(cox_firth)

cox_firth2 <- coxphf(
  Surv(Waktu_Survival_Hari, Status_Event) ~
    Stadium_Kanker_Label +
    Hasil_Histologi,
  data = Data_Kanker,
  pl = TRUE,
  maxit = 1000,
  maxstep = 0.1
)

summary(cox_firth2)

AIC(cox_firth2)
AIC(cox_model2)

# Calibration Slope
lp <- predict(cox_model2, type = "lp")

cal_model <- coxph(
  Surv(Waktu_Survival_Hari, Status_Event) ~ lp,
  data = Data_Kanker
)

```

```

summary(cal_model)

lp_firth <- predict(cox_firth2)

cal_firth <- coxph(
  Surv(Waktu_Survival_Hari, Status_Event) ~ lp_firth,
  data = Data_Kanker
)

summary(cal_firth)

# C-index
summary(cox_model2)
summary(cal_firth)

km_obs <- survfit(
  Surv(Waktu_Survival_Hari, Status_Event) ~ 1,
  data = Data_Kanker
)

pred_all <- survfit(
  cox_model2,
  newdata = Data_Kanker
)

pred_mean <- rowMeans(pred_all$surv)

plot(
  km_obs,
  conf.int = FALSE,
  col = "black",
  lwd = 2,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Survival",
  main = "Observed vs Predicted Survival Curve"
)

lines(
  pred_all$time,
  pred_mean,
  col = "red",
  lwd = 2,
  lty = 2
)

legend(
  "topright",
  legend = c(
    "Observed (Kaplan-Meier)",
    "Predicted (Model)"
  ),
  col = c("black", "red"),

```

```

    lwd = 2,
    lty = c(1,2)
  )

lp_firth <- predict(cox_firth2)

Data_Kanker$risk_group <- ifelse(
  lp_firth > median(lp_firth),
  "Risiko Tinggi",
  "Risiko Rendah"
)

km_firth <- survfit(
  Surv(Waktu_Survival_Hari, Status_Event) ~ risk_group,
  data = Data_Kanker
)

ggsurvplot(
  km_firth,
  data = Data_Kanker,
  risk.table = TRUE,
  conf.int = FALSE,
  pval = TRUE,
  xlab = "Waktu Survival (hari)",
  ylab = "Probabilitas Survival",
  title = "Kurva Survival Berdasarkan Risk Score Model Firth",
  legend.title = "Kelompok Risiko",
  legend.labs = c("Risiko Rendah", "Risiko Tinggi"),
  ggtheme = theme_classic()
)

# Cek Iterasi
cox_model2$iter
cox_firth$iter

```

RIWAYAT HIDUP



Aisya Fatimatuz Zahra lahir di Pasuruan pada tanggal 29 Juli 2005. Penulis menempuh pendidikan dasar di MI NU Miftahul Huda dan lulus pada tahun 2017, kemudian melanjutkan pendidikan menengah pertama di MTs Negeri 3 Pasuruan dan lulus pada tahun 2020. Setelah itu, penulis menempuh pendidikan menengah atas di MA Negeri 1 Pasuruan dan lulus pada tahun 2022. Pada tahun 2022, penulis melanjutkan pendidikan di Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang dengan fokus peminatan pada bidang statistika. Selama masa perkuliahan, penulis aktif dalam kegiatan organisasi kemahasiswaan melalui Himpunan Mahasiswa Program Studi Matematika serta terlibat dalam kegiatan pengembangan bahasa Inggris melalui Mathematics English Club sebagai ketua komunitas. Pengalaman tersebut menjadi bagian dari perjalanan penulis selama menempuh pendidikan di perguruan tinggi hingga terselesaikannya tugas akhir ini.

**BUKTI KONSULTASI SKRIPSI**

Nama : Aisya Fatimatuz Zahra
NIM : 220601110004
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Analisis *Survival Cox Proportional Hazard* dengan
Estimasi *Maximum Likelihood* dan *Firth's Penalized Likelihood*
Pembimbing I : Ria Dhea Layla Nur Karisma, M.Si
Pembimbing II : Mohammad Nafie Jauhari, M.Si

No	Tanggal	Hal	Tanda Tangan
1.	13 Juni 2025	Konsultasi Topik	1.
2.	10 September 2025	Konsultasi Bab I	2.
3.	18 September 2025	Konsultasi Bab I, II, III	3.
4.	25 September 2025	Konsultasi Bab I, II, III	4.
5.	1 Oktober 2025	Konsultasi Bab I, II, III	5.
6.	8 Oktober 2025	Konsultasi Bab I, II, III	6.
7.	15 Oktober 2025	Konsultasi Bab I, II, III	7.
8.	22 Oktober 2025	Konsultasi Bab I, II, III	8.
9.	29 Oktober 2025	Konsultasi Bab I, II, III	9.
10.	30 Oktober 2025	Konsultasi Kajian Agama Bab I dan II	10.
11.	31 Oktober 2025	ACC Kajian Agama Bab I dan II	11.
12.	5 November 2025	ACC Bab I, II, dan III	12.
13.	19 November 2025	Seminar Proposal	13.



14.	9 Desember 2025	Konsultasi Revisi Seminar Proposal	14.
15.	3 Februari 2026	Konsultasi Bab IV dan V	15.
16.	12 Februari 2026	Konsultasi Bab IV dan V	16.
17.	24 Februari 2026	Konsultasi Bab IV dan V	17.
18.	3 Maret 2026	Konsultasi Bab IV dan V	18.
19.	10 Maret 2026	Konsultasi Bab IV dan V	19.
20.	2 April 2026	ACC Kajian Agama Bab IV	20.
21.	7 April 2026	ACC Bab IV dan V	21.
22.	17 April 2026	Seminar Hasil	22.
23.	5 Mei 2026	Konsultasi Revisi Seminar Hasil	23.
24.	5 Juni 2026	ACC Keseluruhan	24.

Malang, 5 Juni 2026

Mengetahui,
Kepala Program Studi Matematika



Dr. Fachrur Rozi, M.Si
NIP. 19800527 200801 1 012