

Log-Rank Test as A Continuation of The Kaplan-Meier Method in Survival Analysis Using R Language (Case Study: D-penicillamine Treatment on the Probability of Survival in Primary Biliary Cirrhosis (PBC) Patients at the Mayo Clinic)

Log-Rank Test Sebagai Lanjutan Dari Metode Kaplan-Meier Dalam Analisis Survival Menggunakan Bahasa R (Studi Kasus: Treatment D-penicillamine terhadap Peluang Tahan Hidup Pasien Primary Biliary Cirrhosis (PBC) di Mayo Clinic)

Ferennia Putri¹, BAP Martadiputra², Lukman³

^{1,2,3} Program Studi Matematika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan, Universitas Pendidikan Indonesia, Indonesia

Email: ferenniaputri25@upi.edu, bambangavip@upi.edu, lukman12@upi.edu

Abstract

From 1974 to 1984, Mayo Clinic conducted a D-penicillamine trial to improve the chances of survival of PBC patients. Mayo Clinic conducted a randomized placebo-controlled trial on patients to compare D-penicillamine with D-penicillamine. Mayo Clinic considered other variables besides the time of death or censored time in its study, namely the variables of Ascites, Hepatomegaly, Spider or the presence of esophageal varices, and other PBC symptom variables. Therefore, the author conducted this study to find out D-penicillamine as a drug treatment for PBC by only looking at the variable of time of death or censored time and the variable of drug treatment received by patients during the ten-year data collection period and completing the study with statistical science. Based on that, the author used survival analysis with the Kaplan-Meier and Log-Rank methods to find D-penicillamine as a PBC drug treatment. This study shows how to complete the analysis using the R language. The study results showed that D-penicillamine treatment was not good enough to increase the survival chances of PBC patients.

Keywords: D-penicillamine, Kaplan-Meier, Log-Rank Test, Cirrhosis



Abstrak

Pada tahun 1974 hingga 1984, Mayo Clinic melakukan uji D-penicillamine dalam meningkatkan peluang tahan hidup pasien pengidap PBC. Mayo Clinic melakukan uji coba acak terkontrol placebo pada pasien sebagai pembanding untuk D-penicillamine. Mayo Clinic memperhatikan variabel lain selain waktu kematian atau pun waktu tersensor pada penelitiannya, yaitu variabel gejala Asites, Hepatomegali, Spider atau adanya varises esofagus, dan variabel gejala PBC lainnya. Oleh karena itu, penulis melakukan penelitian ini dengan tujuan mengetahui D-penicillamine sebagai treatment obat untuk penyakit PBC dengan hanya melihat variabel waktu kematian atau waktu tersensor dan variabel treatment obat yang didapatkan pasien selama masa pengumpulan data rentang sepuluh tahun tersebut serta menyelesaikan penelitian dengan ilmu statistika. Berdasarkan hal itu, penulis menggunakan analisis survival dengan metode Kaplan-Meier dan Log-Rank untuk mengetahui D-penicillamine sebagai treatment obat untuk penyakit PBC. Penelitian ini menunjukkan cara penyelesaian analisis tersebut menggunakan bahasa R. Hasil dari penelitian menunjukkan bahwa treatment D-penicillamine tidak cukup baik dalam meningkatkan peluang tahan hidup pasien pengidap PBC.

Kata Kunci: D-penicillamine, Kaplan-Meier, Log-Rank Test, Cirrhosis

1. PENDAHULUAN

Menurut Budhiarta (2017), sirosis bilier primer atau Primary Biliary Cirrhosis (PBC) adalah suatu penyakit hati autoimun kronis yang dapat berdampak pada saluran empedu kecil di hati. Penyakit ini bersifat langka dan progresif yang seiring berjalannya waktu dapat menyebabkan kerusakan hati yang signifikan. PBC disebabkan oleh respon sistem kekebalan tubuh atau autoimun yang keliru menyerang dan merusak sel-sel saluran empedu di hati. Hal tersebut mengakibatkan saluran empedu mengalami peradangan, penumpukan cairan empedu, dan kerusakan bertahap. Meskipun tidak ada obat yang menyembuhkan secara total penyakit PBC, treatment yang tepat dapat membantu memperlambat perkembangan penyakit dan meningkatkan kualitas hidup penderita. Oleh karena permasalahan tersebut, dapat dilakukan analisis apakah suatu treatment baik untuk mencegah perkembangan penyakit PBC pada pasien. Analisis yang dapat digunakan pada permasalahan ini adalah analisis survival atau ketahanan hidup. Peristiwa seperti kematian suatu makhluk hidup, kekambuhan suatu penyakit, penyembuhan suatu penyakit, dan peristiwa-peristiwa lainnya dapat diketahui dan analisis dengan perkembangan ilmu pengetahuan di abad ini. Analisis survival atau ketahanan hidup merupakan metode yang bertujuan untuk menyelesaikan permasalahan tersebut (Gayatri, 2005).

Analisis survival merupakan analisis data yang diperoleh dari catatan waktu yang dicapai suatu objek sampai terjadinya peristiwa khusus (failure event) (Suhartini, dkk., 2018). Catatan waktu tersebut dapat berupa tahun, bulan, hari, jam, atau menit yang diukur sejak pengamatan dimulai hingga muncul kejadian atau peristiwa (event) (Gayatri, 2005). Analisis survival memiliki tiga faktor penting yaitu waktu ketahanan, kejadian, dan sensor (Maruddani, dkk., 2021). Waktu ketahanan merupakan waktu yang dimiliki objek dari awal penelitian dimulai sampai terjadi event pada objek tersebut. Event atau kejadian yang dimaksud dapat berupa kematian, kesembuhan, kekambuhan, kerusakan alat, insiden penyakit, dan sebagainya. Sementara itu, sensor adalah keadaan ketika objek mengalami beberapa keadaan di antaranya, objek tidak mengalami kejadian selama penelitian, loss to follow up yaitu objek menghilang selama masa penelitian, Withdraw yaitu objek mengundurkan diri dari penelitian, dan objek terpaksa diberhentikan dari penelitian (Maruddani, dkk., 2021).

Data tahan hidup pada analisis survival merupakan variabel random (kontinu) yang dapat membentuk suatu distribusi yang terdiri atas fungsi tahan hidup (survival) (Maruddani, dkk., 2021).

Fungsi tahan hidup atau fungsi survival merupakan peluang atau probabilitas tahan hidup suatu objek sampai waktu tertentu. Fungsi survival dapat diestimasi dengan menggunakan dua metode, yakni metode parametrik dan non parametrik. Metode parametrik dapat digunakan dengan mengasumsikan distribusi populasinya terlebih dahulu, sedangkan metode non parametrik tidak bergantung pada asumsi distribusi populasinya. Salah satu metode non parametrik yang biasa digunakan adalah metode Kaplan-Meier (Firsawan dkk., 2022).

Kaplan-Meier atau yang bisa disebut juga Product-limit diperkenalkan pertama kali oleh Kaplan dan Meier pada tahun 1958 (Muhajir & Palupi, 2017). Metode Kaplan-Meier merupakan hasil modifikasi dari fungsi tahan hidup yang digunakan untuk menangani masalah data tak lengkap (Lawless, 2003). Penanganan data tersensor bisa dilakukan dengan Metode Kaplan-Meier dengan perhitungan yang relatif sederhana. Selain itu, metode Kaplan-Meier dapat digunakan pada data sampel berukuran kecil (Cook, 2008). Kelebihan metode Kaplan-Meier yakni dapat memberikan proporsi ketahanan hidup yang pasti sebab metode ini menggunakan waktu ketahanan hidup secara tepat sehingga diperoleh data yang lebih akurat bukan berdasarkan interval. Analisis Kaplan-Meier digunakan untuk mengestimasi fungsi survival kemudian dari estimasi fungsi survival dapat dibentuk kurva survival Kaplan-Meier (Suhartini, dkk., 2018). Metode Kaplan-Meier dapat digunakan untuk membandingkan dua atau lebih data tahan hidup berdasarkan kurva estimasi fungsi survival. Salah satu pengujian atau perbandingan antara dua kurva survival atau lebih untuk data yang memiliki data tersensor adalah Log-Rank.

Studi kasus pada penelitian ini cocok untuk menggunakan metode Kaplan-Meier. Hal tersebut disebabkan oleh metode Kaplan-Meier yang bersifat fleksibel daripada analisis survival lainnya karena analisis survival yang bersifat parametrik harus mengidentifikasi data masuk ke dalam suatu distribusi (Khoiri, 2018). Kaplan-Meier merupakan analisis statistik yang bersifat non parametrik sehingga tidak menggunakan asumsi pada parameter populasi datasetnya. Selain itu, data nyata yang ada pada lapangan umumnya bersifat data tidak lengkap atau terdapat data tersensor. Metode Kaplan-Meier cocok digunakan pada data penyakit contohnya seperti data ketahanan penyakit PBC pada pasien yang data sajiannya pada umumnya bersifat data tidak lengkap atau terdapat data tersensor (pada istilah analisis survival).

Cara kerja pada studi kasus penelitian ini adalah mengklasifikasi pasien dengan macam treatment pemberian obat yang berbeda. Pasien PBC dibagi menjadi tiga golongan yaitu pasien yang diberi obat Placebo, D-penicillamine, dan kelompok pasien yang tidak diketahui jenis obat yang diberikan atau NA. Hal tersebut dilakukan karena penulis ingin mengetahui apakah obat D-penicillamine baik terhadap peluang tahan hidup pasien pengidap penyakit PBC dengan melihat peluang tahan hidup dari ketiga kelompok pasien. Tiga data sampel tersebut masing-masing dianalisis ketahanan hidup pasien dari penyakit menggunakan Kaplan-Meier dengan melihat waktu ketahanan, waktu kejadian, dan waktu data tersensor dari para pasien. Setelah mengetahui ketahanan penyakit dari masing-masing treatment obat yang berbeda, waktu ketahanan masing-masing dibandingkan oleh metode Log-Rank. Penyelesaian seluruh analisis dari metode Kaplan-Meier hingga metode Log-Rank akan dilakukan dengan bantuan program dalam bahasa R. R merupakan bahasa pemrograman dan lingkungan komputasi yang digunakan untuk analisis statistik dan visualisasi data (Rosadi, 2016). Lingkungan pengembangan R yang populer adalah RStudio, yang menyediakan antarmuka yang nyaman untuk bekerja dengan R. R sering digunakan dalam berbagai bidang, termasuk ilmu sosial, ilmu alam, ekonomi, kedokteran, dan khususnya untuk analisis survival yang akan dilakukan pada penelitian ini (Kosourova, E. dalam datacamp.com. 2022).

Oleh karena hal-hal yang telah disampaikan di atas, penulis tertarik untuk melakukan penelitian mengenai penggunaan metode Kaplan-Meier kemudian dilanjutkan dengan membandingkan waktu ketahanan masing-masing treatment dengan Log-Rank test menggunakan bahasa pemrograman R pada artikel yang berjudul “Log-Rank Test Sebagai Lanjutan dari Metode Kaplan-Meier dalam

Analisis Survival Menggunakan Bahasa R (Studi Kasus: Treatment D-penicillamine terhadap Peluang Tahan Hidup Pasien Primary Biliary Cirrhosis (PBC) di Mayo Clinic)".

2. TINJAUAN PUSTAKA

2.1 Ketahanan Hidup (Survival)

Maruddani, dkk., (2021) menyatakan bahwa ketahanan hidup (*survival*) merupakan probabilitas atau peluang yang dimiliki suatu objek terus beroperasi tanpa adanya kejadian (*event*) untuk waktu yang ditentukan di bawah kondisi yang disyaratkan. *Event* dapat berupa kematian, kesembuhan, kekambuhan, kerusakan alat atau bahan, insiden penyakit, pemulihan, dan sebagainya. Analisis data tahan hidup (*survival data analysis*) adalah alat yang digunakan untuk mendeskripsikan data tahan hidup. Data tahan hidup (*survival data*) adalah data waktu terjadinya suatu peristiwa dari awal pengamatan sampai dengan waktu berakhirnya pengamatan (Maruddani, dkk., 2021).

Berdasarkan pendapat Maruddani, dkk. (2021) waktu tahan hidup (T) adalah variabel random non-negatif yang mewakili ketahanan hidup dari individu-individu atau objek-objek dari suatu populasi. Variabel random (T) merupakan variabel random yang bersifat kontinu dalam selang $[0, \infty)$ atau ketahanan hidup pada waktu t dengan $t > 0$. Waktu tahan hidup (T) akan membentuk suatu distribusi yang disebut dengan distribusi waktu hidup.

Penentuan data tahan hidup memiliki tiga faktor yang harus diperhatikan:

- a. Waktu awal pencatatan (*time origin* atau *start-point*).
- b. Waktu akhir pencatatan (*event time* atau *end-point*).
- c. Skala pengukuran.

2.2 Tipe Data Survival

Data tahan hidup dengan data lengkap (*complete data*) untuk setiap objek tidak bisa selalu tersedia pada pengamatan (Maruddani, dkk., 2021). Data lengkap mengharuskan setiap objek diamati sejak awal masuk ke dalam pengamatan sampai dengan objek tersebut mengalami *event* yang telah ditentukan. Kenyataannya, terdapat kemungkinan objek-objek tersebut keluar dari pengamatan sebelum ada *event*, objek tidak mengalami *event* selama masa pengamatan, atau kemungkinan lainnya yang menjadikan objek tersebut tidak bisa diamati hingga akhir pengamatan atau terjadinya *event* pada objek tersebut. Oleh karena itu, tidak setiap objek dapat dihitung data tahan hidupnya. Kondisi seperti ini membutuhkan tipe data tersensor, yaitu objek yang tidak dapat ditentukan data tahan hidupnya sehingga terdapat dua tipe data pada data *survival* yakni data lengkap dan data tersensor (Kleinbaum & Klein, 2012).

2.2.1 Data Lengkap

Data lengkap merupakan keadaan setiap objek masuk pada waktu pengamatan yang sama diamati sampai dengan objek tersebut mengalami *event* yang telah ditentukan. Kondisi data lengkap terjadi dengan objek sebanyak n atau dapat ditulis sebagai $x_1, x_2, x_3, \dots, x_n$ kemudian masing-masing objek tersebut diamati sampai mengalami *event* sehingga dimiliki data tahan hidup yakni $t_1, t_2, t_3, \dots, t_n$ (Kleinbaum & Klein, 2012).

2.2.2 Data Tersensor

Data tersensor terjadi apabila objek tidak mengalami kejadian selama penelitian, *loss to follow up* yaitu objek menghilang selama masa penelitian, *Withdraw* yaitu objek mengundurkan diri dari penelitian, atau objek terpaksa diberhentikan dari penelitian. (Kleinbaum & Klein, 2012).

2.3 Fungsi Survival

Berdasarkan Maruddani, dkk., (2021), data tahan hidup merupakan variabel random yang dapat membentuk suatu distribusi. Distribusi data tahan hidup terdiri dari tiga fungsi, yaitu fungsi tahan hidup (survival), fungsi densitas probabilitas, dan fungsi kegagalan (hazard). Jika T adalah variabel random yang menyatakan data tahan hidup maka distribusi dari T dapat dinyatakan dalam tiga buah fungsi yang ekuivalen seperti yang telah disebutkan. Penelitian ini mengacu pada metode Kaplan-Meier yang menggunakan fungsi tahan hidup (survival) sehingga fungsi yang akan difokuskan adalah fungsi tahan hidup. Fungsi ini dinotasikan dengan $S(t)$, dan didefinisikan sebagai probabilitas suatu objek tidak mengalami event lebih dari waktu t , dengan $t > 0$ (Maruddani, dkk., 2021).

2.4 Metode Kaplan-Meier

Berdasarkan pernyataan Maruddani, dkk., (2021), misalkan terdapat data n objek dengan data tahan hidup yang telah diurutkan dari waktu *event* terkecil ke waktu *event* terbesar yakni $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$, maka fungsi survival pada saat $t_{(i)}$:

$$\hat{S}(t_{(i)}) = \frac{n-i}{n} = 1 - \frac{i}{n} \quad (2.1)$$

dengan $n - i$ adalah jumlah objek yang tidak mengalami *event* atau *survive* lebih dari waktu $t_{(i)}$. Untuk data lengkap, setiap objek masih hidup pada saat awal studi, dan tidak ada yang masih hidup lebih dari waktu $t_{(n)}$ maka

$$\hat{S}(t_{(0)}) = 1 \text{ dan } \hat{S}(t_{(n)}) = 0$$

Metode di atas hanya dapat diaplikasikan jika seluruh objek mengalami *event*.

Metode Kaplan-Meier dapat dilakukan untuk data yang memuat data tersensor yakni kondisi ketika terdapat objek yang tidak mengalami *event* sampai dengan waktu akhir penelitian.

Berikut merupakan ilustrasi dari metode Kaplan-Meier. Simbol atau keterangan yang harus dipahami pada ilustrasi:

t_0 = awal dari masa penelitian

t_1 = titik terjadinya *event* untuk pertama kali

t_2 = titik terjadinya *event* untuk kedua kali

t_3 = titik terjadinya *event* untuk ketiga kali

t_i = titik terjadinya *event* untuk ke- i kali, dengan $i = 0,1,2,3, \dots$

R_i = himpunan yang berisi objek yang beresiko mengalami *event* di titik waktu t_i , dengan $i = 0,1,2, \dots$

Y_i = banyaknya anggota dari R_i , dengan $i = 0,1,2, \dots$

m = banyaknya sampel pada penelitian

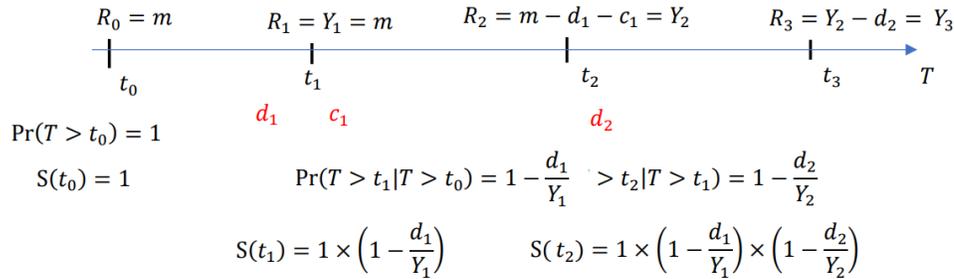
d_i = banyaknya objek yang mengalami kejadian pada waktu t_i

c_i = banyaknya objek tersensor pada waktu t_i

$\Pr(T > t_i)$ = peluang tahan hidup suatu objek setelah sampai dan melewati t_i

$\Pr(T > t_{i+1}|T > t_i)$ = peluang bersyarat tahan hidup suatu objek setelah melewati t_{i+1} dan t_i

$S(t_i) = \Pr(T > t_i)$ = fungsi tahan hidup atau survival pada t_i



Gambar 2.1 Ilustrasi Kaplan-Meier

Misalkan dalam suatu penelitian terdapat m objek sampel yang akan diketahui fungsi survivalnya. Perhatikan bahwa,

$$\begin{aligned} R_0 &= m = Y_0 \\ \Pr(T > t_0) &= 1 \\ S(t_0) &= 1 \end{aligned}$$

Abdullah (2020) menyatakan hal tersebut mengartikan bahwa himpunan yang berisi objek-objek yang beresiko mengalami *event* pada t_0 memiliki m objek atau setara dengan banyaknya sampel karena belum ada objek yang mengalami *event* sehingga probabilitas atau peluang tahan hidup setiap objek saat t_0 adalah 1. Pada saat keadaan t_1 , himpunan yang berisi objek-objek yang beresiko mengalami *event* memiliki $Y_1 = m$. Akan tetapi, pada saat t_1 terdapat objek yang mengalami *event* sebanyak d_1 dan c_1 objek yang hilang dari masa penelitian atau disebut data tersensor. Perlu diingat bahwa apabila terdapat objek yang terindikasi sebagai data tersensor maka titik waktu pada objek tersebut tidak didefinisikan taksiran dari fungsi survival sehingga $\Pr(T > t_1|T > t_0) = 1 - \frac{d_1}{Y_1}$ (c_1 tidak mempengaruhi taksiran atau peluang tahan hidup) dan $S(t_1) = 1 \times \left(1 - \frac{d_1}{Y_1}\right)$.

Begitu pun seterusnya untuk $S(t_2)$, $S(t_3)$, hingga $S(t_n)$ dengan n adalah banyaknya kejadian. Oleh karena itu, dapat disimpulkan bahwa fungsi survival dalam me :
determined dengan rumus,

$$S(t_i) = \begin{cases} 1, & \text{jika } i = 0 \\ \prod_1^i \left[1 - \frac{d_i}{Y_i}\right], & \text{jika } i \geq 1 \end{cases} \quad (2.2)$$

2.5 Log-Rank Test

Menurut Ruslianah, I.N. (2023), Uji *Log-Rank* digunakan untuk mendeskripsikan kurva yang dibentuk dari kategori yang ada pada setiap variabel bebas dengan kondisi yang berbeda. Statistik Uji *Log-Rank* merupakan jumlah observasi kegagalan dikurangi ekspektasi kegagalan bersyarat yang dihitung pada setiap waktu kegagalan atau selisih antara observasi dengan ekspektasi kegagalan dalam satu kelompok atau grup.

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ferennia Putri, BAP Martadiputra , Lukman

- *Log-Rank Test* tiga grup

Misalkan $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(k)}$ merupakan data waktu tahan hidup pada tiga sampel grup. O_1, O_2 , dan O_3 adalah jumlah observasi pada masing-masing grup, dan E_1, E_2 , dan E_3 merupakan ekspektasi jumlah kegagalan pada masing-masing grup (Kleinbaum & Klein, 2012). Statistik uji dari *Log-Rank Test*:

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} \quad (2.3)$$

Statistik uji di atas berdistribusi *Chi-Square* dengan derajat bebas 2. Perhitungan E_1, E_2 , dan E_3 dilakukan dengan mengambil d_t yang merupakan jumlah yang gagal pada waktu t . n_{1t}, n_{2t} , dan n_{3t} jumlah yang berisiko gagal lebih dari t dalam grup 1, 2, dan 3. Ekspektasi kegagalan untuk kelompok 1, 2, dan 3 adalah

$$e_{1t} = \frac{n_{1t}}{n_{1t} + n_{2t} + n_{3t}} d_t \quad (2.4)$$

$$e_{2t} = \frac{n_{2t}}{n_{1t} + n_{2t} + n_{3t}} d_t \quad (2.5)$$

$$e_{3t} = \frac{n_{3t}}{n_{1t} + n_{2t} + n_{3t}} d_t \quad (2.6)$$

Sehingga ekspektasi kegagalan total dalam tiga grup atau kelompok adalah

$$E_1 = \sum_t e_{1t} \quad (2.7)$$

$$E_2 = \sum_t e_{2t} \quad (2.8)$$

$$E_3 = \sum_t e_{3t} \quad (2.9)$$

Hipotesis:

Antara grup 1, grup 2, dan grup 3

$H_0 : S_i(t) = S_j(t), i \neq j, i, j = 1, 2, 3$: tidak ada perbedaan nilai fungsi survival antara kelompok yang berbeda pada suatu variable bebas.

$H_1 : S_i(t) \neq S_j(t), i \neq j, i, j = 1, 2, 3$: terdapat perbedaan nilai fungsi survival antara kelompok yang berbeda pada suatu variable bebas.

Kriteria keputusan:

H_0 ditolak pada taraf signifikansi α jika

$$\text{nilai } \chi^2_{\text{hitung}} > \chi^2_{\text{tabel}} \text{ atau } p\text{value} \leq \alpha$$

H_0 diterima pada taraf signifikansi α jika

$$\text{nilai } \chi^2_{\text{hitung}} \leq \chi^2_{\text{tabel}} \text{ atau } p\text{value} > \alpha$$

2.6 Bahasa R untuk Metode Kaplan-Meier dan Log-Rank Test

Berdasarkan pernyataan Schutte, D. dalam *datacamp.com*. (2019), Kaplan-Meier merupakan fungsi yang terdapat pada *package* 'survival' atau bisa ditulis dengan *code*:

```
> library(survival)
```

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ferennia Putri, BAP Martadiputra , Lukman

Menurut Revou (2023), *Library* adalah kumpulan kode yang menyediakan fungsionalitas tertentu untuk digunakan dalam aplikasi. Berdasarkan hal tersebut, semua fungsi-fungsi yang ada dalam analisis survival, termasuk metode Kaplan-Meier, terdapat pada *Library* atau *Package* tersebut. Penginputan data yang dilakukan dalam penelitian ini adalah dengan menginput data berformat Excel. *Code* yang digunakan:

```
> namadata <- read_excel("lokasi file dalam device").
```

Apabila melihat data tersebut di dalam RStudio maka perlu menulis *code*:

```
> View(namadata)           (untuk melihat data dalam jendela Editor)
> namadata                 (untuk melihat data dalam jendela Console)
```

Berdasarkan *R Documentation*, kurva analisis survival dalam R dapat dilakukan dengan fungsi *Surv*, *plot*, dan *survfit* yang terdapat pada *Library* atau *package* 'survival'. *Surv* adalah fungsi untuk membuat objek survival atau sebagai variabel respon dalam model survival tersebut. *Survfit* adalah fungsi untuk membuat kurva survival dari rumus (misalnya Kaplan-Meier), model Cox yang dipasang sebelumnya, atau model waktu kegagalan dipercepat yang dipasang sebelumnya. *Plot* adalah fungsi untuk membuat grafik. Berikut adalah *code* yang didapatkan dari video pembelajaran analisis survival Ramya Rachmawati pada tahun 2020 dan telah dimodifikasi oleh penulis di bagian *Plot* serta pendefinisian *Summary Survfit*:

```
> namakurva <- survfit(Surv(waktu,status)~grup, data = namadata)
> plot(namakurva, conf.int=T/F, mark.time=T/F, pch=numeric,
      col="color code", lty=numeric, lwd=numeric, cex=numeric,
      log=T/F, xscale=numeric, yscale=numeric, xlim, ylim, xmax, fun,
      xlab="namasumbu-x", ylab="namasumbu-y", xaxs="r", conf.times,
      firstx, ymin, main = "judul plot", las = numeric,
      font.main=numeric, col.main = "color code", cex.main = numeric,
      font.axis = numeric, font.lab = numeric, col.lab = "color code",
      cex.lab=numeric, cex.axis= numeric)
```

Perhitungan dari hasil metode Kaplan-Meier pada R dapat dilakukan dengan *code*:

```
> summary(survfit(Surv(waktu,status)~grup, data = namadata))
```

Log-Rank Test pada R dapat dilakukan dengan fungsi *survdiff* dan *Surv*. *Survdiff* adalah fungsi untuk menguji apakah ada perbedaan antara dua atau lebih kurva survival. *Code* yang digunakan:

```
> survdiff(Surv(waktu, status)~grup, data=namadata)
```

Variabel yang diisi pada *code* di atas terdapat waktu yang didefinisikan dengan waktu terjadinya peristiwa *event* dan waktu penyensoran, status yang merupakan status objek apakah mengalami *event* atau sensor, grup mendefinisikan kelompok, dan data merupakan data yang digunakan (Schutte, D. dalam *datacamp.com*. 2019).

2.7 Fungsi D-penicillamine dan Placebo terhadap Penyakit Primary Biliary Cirrhosis (PBC)

Seperti yang telah disampaikan pada Pendahuluan, penyakit PBC merupakan keadaan hati yang mengalami kerusakan secara bertahap dan pengerasan akibat penumpukan jaringan fibrosa. Pada tahun 1974 sampai 1984, *treatment* yang diberikan oleh *Mayo Clinic* kepada pasien PBC salah

satunya adalah *treatment D-penicillamine*. *D-penicillamine* adalah obat *cuprurctic* yang dikenal karena kemanjurannya dalam mengobati penyakit Wilson (Sternlieb, 1964; Deiss, 1971 dalam Gong, Y., dkk., 2004). Sirosis bilier primer juga dikaitkan dengan peningkatan kadar tembaga di hati yang mengakibatkan hati mengeras. Oleh karena itu, terdapat keterkaitan antara *D-penicillamine* pada sirosis bilier primer dengan kemampuannya untuk menginduksi *cuprurctic*.

Mayo Clinic juga pada saat itu memberi *treatment* obat *placebo* pada pasien PBC sebagai pembandingan obat *D-penicillamine*. *Placebo* sering disebut sebagai obat kosong sebab kandungannya yang tidak berbahan aktif sehingga tidak mempengaruhi kesehatan sama sekali. Bentuk *placebo* dapat berupa tablet, kapsul, atau cairan suntik. Akan tetapi, faktanya kandungan dalam *placebo* hanyalah tepung, gula, atau larutan garam, bahkan mungkin hanya air putih (Adrian, K., 2021). *Placebo* sering digunakan dalam uji klinis obat atau vaksin untuk membantu peneliti memahami dan mengevaluasi pengaruh obat atau vaksin yang sedang diteliti dengan lebih baik. Meski tidak mengandung bahan aktif, beberapa relawan yang mengonsumsi *placebo* bisa merasa bahwa penyakit atau gejala yang mereka rasakan membaik. Fenomena ini disebut *placebo effect* atau efek plasebo (Adrian, K., 2021).

Dataset yang berhasil dikumpulkan dan diobservasi oleh *Mayo Clinic* pada rentang 10 tahun tersebut menyatakan bahwa terdapat pasien yang tidak diketahui jenis *treatment* obat yang diberikan. Hal tersebut terjadi karena pasien tidak berpartisipasi dalam uji klinis, tetapi menyetujui pengukuran dasar dicatat dan diikuti untuk kelangsungan hidup. Kemungkinan dari data *treatment* pasien yang bersifat NA (*treatment* yang tidak diketahui) tersebut diberikan *treatment D-penicillamine*, diberi jenis obat lain selain *D-penicillamine*, atau bahkan tidak diberi obat PBC apa pun (hanya kontrol gizi). Meskipun begitu, nutrisi yang seimbang baik dari segi kalori, karbohidrat, protein dan lemak, akan membawa pengaruh yang baik untuk memperbaiki kerusakan sel hati (Sutadi, 2003). Pada tingkat tertentu, kerusakan sel hati masih bisa diperbaiki dengan cara memproduksi sel hati baru yang sehat. Nutrisi sangat dibutuhkan pada pasien PBC untuk meningkatkan regenerasi jaringan hati, mencegah kerusakan lebih lanjut atau meningkatkan fungsi jaringan hati yang tersisa, mencegah penurunan berat badan atau meningkatkan berat badan bila kurang, dan mencegah komplikasi lebih lanjut (Purba dan Sinurat, 2018).

2.8 RMSE (Root Mean Square Error)

Menurut Maulid (2022), RMSE (*Root Mean Square Error*) adalah besar tingkat kesalahan atau error dari hasil prediksi atau peramalan. Semakin kecil nilai RMSE yang diperoleh maka hasil prediksi atau peramalan tersebut semakin akurat. RMSE diperoleh dari akar rata-rata kuadrat dari selisih antara nilai yang diprediksi oleh model persamaan dan nilai yang sebenarnya.

RMSE diperoleh dengan rumus:

$$RMSE = \varepsilon = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}} \quad (2.10)$$

Keterangan:

y_i = nilai sebenarnya

\hat{y}_i = nilai prediksi

n = jumlah sampel yang diuji

2.9 SEM (Structural Equation Model)

Structural Equation Model (SEM) adalah teknik statistik yang bertujuan untuk mengetahui dan menguji hubungan antar variabel pada suatu model (Ginting, 2009). SEM dapat melakukan uji

kecocokan (*Fit Indices*) sebab SEM dapat memperoleh ukuran yang digunakan untuk mengukur seberapa baik model sesuai dengan data yang diamati.

Menurut Robi, dkk. (2017) terdapat beberapa kriteria atau indeks yang umum digunakan untuk menilai kebaikan model pada SEM yaitu sebagai berikut:

1. *Root Mean Square Error of Approximation (RMSEA)*

Browne dan Cudeck pada tahun 1993 memberi aturan yang telah menjadi standar umum pada SEM untuk indeks RMSEA. Panduan yang tepat menurut mereka adalah sebagai berikut:

- *Close fit* (Sangat Akurat): $RMSEA \leq 0.05$
- *Good fit* (Akurat): $0.05 < RMSEA \leq 0.08$
- *Mediocre fit* (Cukup Akurat): $0.08 < RMSEA \leq 0.10$
- *Poor fit* (Kurang Akurat): $RMSEA > 0.10$

2. *Comparative Fit Index (CFI)*

CFI diperkenalkan oleh Karl G. Jöreskog dan Dag Sörbom pada tahun 1978. Sebagai kontribusi dalam pengembangan aplikasi SEM, mereka menyimpulkan bahwa CFI mengindikasikan model sesuai atau cocok dengan indeks lebih dari 0.95.

3. *Tucker-Lewis Index (TLI)*

TLI dicetus oleh LR Tucker dan C. Lewis pada tahun 1973. Mereka turut andil dalam pengembangan analisis metode SEM. TLI adalah salah satu indikator SEM yang mengindikasikan model sesuai atau cocok dengan indeks lebih dari 0.95.

4. *Standardized Root Mean Square Residual (SRMR)*

Pada tahun 1995, SRMR dikemukakan pertama kali oleh Peter M. Bentler. Peter seorang profesor psikologi dan metode statistik yang berpengaruh dalam pengembangan dan aplikasi SEM. Beliau menyatakan bahwa nilai SRMR yang rendah menunjukkan bahwa model sesuai yakni senilai kurang dari 0.08.

3. METODE PENELITIAN

3.1 Deskripsi Masalah

Penyakit Sirosis Bilier Primer atau *Primary Biliary Cirrhosis (PBC)* adalah penyakit dengan kondisi hati dalam keadaan penumpukan jaringan parut atau fibrosa sehingga hati mengeras dan tidak bisa berfungsi dengan baik (Budhiarta, 2017). Penyakit PBC tidak bisa disembuhkan secara total dengan *treatment* obat apa pun kecuali dilakukan transplantasi organ hati. Salah satu cara pasien pengidap penyakit PBC agar dapat menjalani hidup dengan kualitas lebih baik adalah dengan mengonsumsi obat yang dapat memperlambat perkembangan penyakit tersebut (Budhiarta, 2017).

Pada tahun 1974 hingga 1984, *Mayo Clinic* mengumpulkan data pasien pengidap PBC untuk melakukan uji coba *D-penicillamine* terhadap pasien pengidap PBC. Sebanyak 424 pasien PBC yang dirujuk ke *Mayo Clinic* selama interval sepuluh tahun tersebut memenuhi kriteria kelayakan uji coba acak terkontrol *placebo* terhadap obat *D-penicillamine*. 312 kasus pertama dalam kumpulan data berpartisipasi dalam uji coba acak dan sebagian besar berisi data lengkap. 312 pasien tersebut

diberi *treatment* obat *D-penicillamine* atau *placebo* secara acak. Sebanyak 112 kasus tambahan tidak berpartisipasi dalam uji klinis, tetapi menyetujui pengukuran dasar dicatat dan diikuti untuk kelangsungan hidup sehingga pasien tersebut tidak memiliki keterangan mengenai *treatment* obat yang mereka dapatkan pada dataset atau dikategorikan sebagai NA (*treatment* yang tidak diketahui). Enam dari kasus tersebut mangkir segera setelah diagnosis atau tidak bersedia untuk mengikuti uji coba dan pendataan, jadi data di sini adalah 106 kasus tambahan serta 312 peserta yang diacak.

Uji acak terkendali atau uji acak terkontrol (*Randomized Controlled Trial, RCT*) yang dilakukan oleh *Mayo Clinic* pada masa itu bertujuan untuk mengetahui *D-penicillamine* sebagai *treatment* obat untuk penyakit PBC dengan mempertimbangkan juga variabel-variabel lain seperti adanya gejala Asites, Hepatomegali, *Spider* atau adanya varises esofagus, dan variabel gejala PBC lainnya. Oleh karena itu, penulis melakukan penelitian ini dengan tujuan mengetahui *D-penicillamine* sebagai *treatment* obat untuk penyakit PBC dengan hanya melihat variabel waktu kematian atau waktu tersensor dan variabel *treatment* obat yang didapatkan pasien selama masa pengumpulan data rentang sepuluh tahun tersebut serta menyelesaikan penelitian dengan ilmu statistika. Berdasarkan hal itu, penulis menggunakan analisis survival dengan metode *Kaplan-Meier* dan *Log-Rank* untuk mengetahui *D-penicillamine* sebagai *treatment* obat untuk penyakit PBC.

3.2 Jenis dan Sumber Data

Jenis data yang digunakan adalah data sekunder yang bersumber dari rekam medis pasien rawat inap penderita PBC yang dirujuk untuk dirawat di *Mayo Clinic*. *Mayo Clinic* berlokasi di Amerika Serikat yang berpusat di wilayah Arizona, Florida, dan Minnesota. Selama perawatan tersebut, *Mayo Clinic* memberi *treatment* yang berbeda pada pasien-pasien guna menguji *D-penicillamine* untuk pasien PBC. Sebanyak 418 pasien pengidap PBC dibagi menjadi tiga kelompok yaitu pasien yang diberi obat *placebo*, *D-penicillamine*, dan NA (tidak diketahui *treatment* obat yang diberikan). *Mayo Clinic* mendata banyak pasien yang menerima transplantasi hati, banyak pasien yang meninggal akibat penyakit ini, dan banyak pasien yang masih bertahan hingga melebihi lama waktu pengumpulan data berakhir (data tersensor).

3.3 Variabel Penelitian

Terdapat tiga variabel dalam penelitian ini untuk mendapatkan estimasi fungsi survival agar memperoleh kurva Kaplan-Meier, yaitu waktu terjadinya *event*, waktu terjadinya data tersensor, dan pengelompokan pasien berdasarkan pemberian jenis obat. Waktu terjadinya *event* diindikasikan dengan kematian pasien. Waktu terjadinya data tersensor diindikasikan dengan pasien yang tidak mengalami kematian selama masa pengamatan, *loss to follow up* yaitu pasien menghilang selama masa pengamatan, *Withdraw* yaitu pasien mengundurkan diri dari pengamatan, atau pasien terpaksa diberhentikan dari pengamatan. Selain itu, pengunduran diri pasien dari pengamatan yang disebabkan oleh penerimaan transplantasi hati dikategorikan sebagai data tersensor pada penelitian ini. Pasien dikelompokkan berdasarkan pemberian jenis obat, yaitu pasien yang diberi obat *placebo*, *D-penicillamine*, dan NA (*treatment* yang tidak diketahui).

Variabel bebas yang digunakan dalam penelitian ini adalah jenis obat yang diberikan kepada pasien karena obat dapat mempengaruhi tumbuh kembang penyakit PBC pada pasien menjadi lebih lambat atau tidak. Variabel terikat yang digunakan dalam penelitian ini adalah waktu terjadinya kejadian dan waktu terjadinya data tersensor. Berikut kategori variabel terikat:

0 = pasien mengalami data tersensor (C = *Censoring*, CL = *Censoring due to tx* (transplantasi hati))

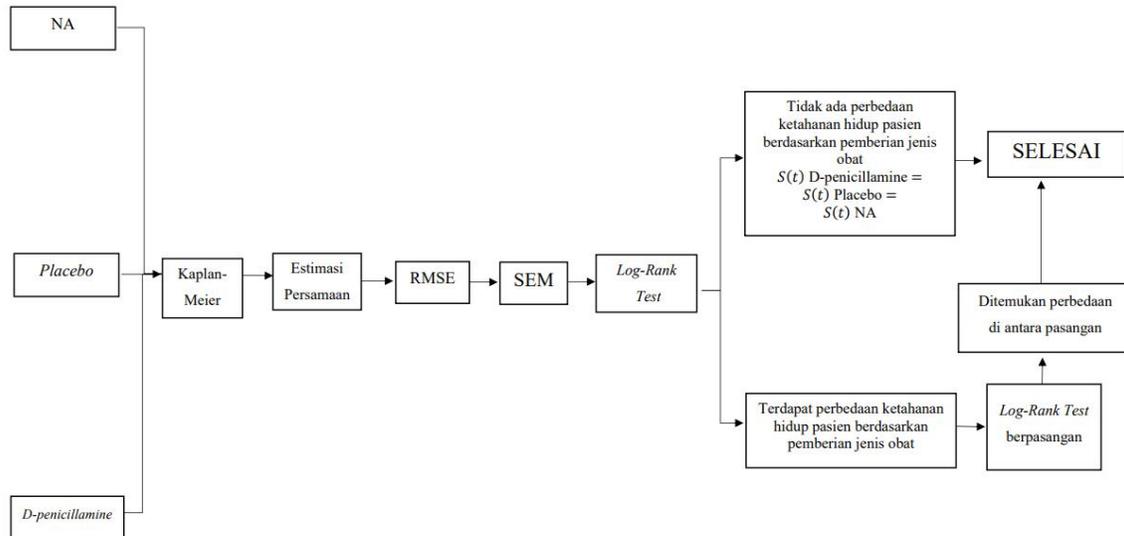
1 = pasien mengalami kejadian atau *event* yang diindikasikan dengan kematian (D = *Death*)

3.4 Langkah Penelitian

Berikut ini adalah langkah-langkah yang diperlukan untuk membentuk kurva estimasi fungsi survival dengan Kaplan-Meier serta perbandingan dari ketiga kelompok menggunakan *Log-Rank Test* dengan program R:

1. Studi literatur
2. Mengumpulkan data pasien *Mayo Clinic* pengidap PBC dalam rentang sekitar 10 tahun dimulai dari tahun 1974.
3. Menentukan *treatment* pengobatan sebagai faktor yang dapat berpengaruh pada ketahanan hidup pasien PBC.
4. Membagi data pasien menjadi tiga kelompok, yaitu pasien yang diberi *treatment* pemberian obat *placebo*, *D-penicillamine*, dan NA (*treatment* yang tidak diketahui).
5. Menghitung ketahanan hidup $S(t)$ pasien dengan metode Kaplan-Meier dari setiap kelompok.
6. Menguji keakuratan hasil fungsi survival yang didapatkan dari *code* R di atas dengan mengestimasi persamaan yang mungkin dari fungsi survival tersebut kemudian memilih persamaan terbaik dilihat dari error terkecil yang dihasilkan dari metode RMSE. Setelah itu, peramalan dari beberapa data fungsi survival dilakukan dan diuji kecocokannya dengan *Structural Equation Model* (SEM).
7. Membuat plot kurva Kaplan-Meier hasil perhitungan ketahanan hidup $S(t)$ dari setiap kelompok.
8. Menguji apakah terdapat perbedaan ketahanan hidup $S(t)$ antar kelompok dengan metode *Log-Rank Test*. Selama tahap ini, dilakukan pencarian nilai observasi dan ekspektasi dari setiap kelompok untuk memperoleh hasil statistik uji atau nilai χ^2_{hitung} yang akan dibandingkan dengan χ^2_{tabel} . Selain itu, *p-value* akan diperoleh yang bisa dibandingkan dengan nilai *alpha* atau taraf signifikansi yang digunakan.
perhitungan *Log-Rank Test* pada data *pbcc*
9. Membuat kesimpulan dan interpretasi dari hasil perhitungan ketahanan hidup $S(t)$ pasien berdasarkan pemberian jenis obat dengan metode Kaplan-Meier dan hasil perbandingan pemberian jenis obat berdasarkan nilai statistik uji yang didapatkan dari *Log-Rank Test*.
10. Apabila hasil dari *Log-Rank Test* menyatakan bahwa tidak terdapat perbedaan ketahanan hidup antara ketiga *treatment* maka tahap dihentikan dan dapat diambil simpulan.
11. Apabila hasil dari *Log-Rank Test* menyatakan bahwa terdapat perbedaan ketahanan hidup antara ketiga *treatment* maka tahap dilanjutkan dengan pengujian *Log-Rank Test* dengan *treatment* berpasang-pasangan.
12. Melakukan pengujian Kaplan-Meier dan *Log-Rank Test* dari *placebo* dengan *D-penicillamine*. Setelah itu, pengujian Kaplan-Meier dan *Log-Rank Test* dari *placebo* dengan NA (*treatment* yang tidak diketahui) Kaplan-Meier dan *Log-Rank Test* dari *D-penicillamine* dengan NA (*treatment* yang tidak diketahui). Hal tersebut dilakukan supaya dapat diketahui pasangan kelompok mana yang menyatakan terdapat perbedaan ketahanan hidup pasien. Berdasarkan hal itu dapat diketahui apakah ketahanan hidup variabel berbeda secara keseluruhan atau sebagian variabel.

3.5 Alur Penelitian



Gambar 3.1 Diagram Alir

4. HASIL DAN PEMBAHASAN

Murtaugh, dkk. (1994) menyampaikan bahwa *Mayo Clinic* menggunakan metode *Risk Scores* dengan mempertimbangkan variabel usia pasien, bilirubin, albumin, prothrombin, dan edema. James (1985) mengemukakan bahwa hasil penelitian *Mayo Clinic* pada rentang 10 tahun tersebut menyatakan bahwa *D-penicillamine* tidak memiliki manfaat pada mortalitas atau usaha penurunan peluang kematian pada pengidap PBC. Oleh karena itu, hasil dan pembahasan ini akan membuktikan apakah hasil penelitian ini dengan yang dilakukan oleh *Mayo Clinic* adalah sama-sama menyatakan bahwa *D-penicillamine* tidak memiliki manfaat untuk meningkatkan fungsi survival atau peluang tahan hidup pasien PBC dengan cukup mempertimbangkan variabel waktu status pasien dan jenis *treatment* yang didapatkan pasien menggunakan ilmu statistika yakni metode Kaplan-Meier dan Log-Rank test.

4.1 Fungsi Survival Hasil Kaplan-Meier

Berdasarkan *output R* yang dihasilkan, fungsi survival atau peluang ketahanan hidup tiap waktu *event* (waktu pasien meninggal) dapat diketahui. Pada kelompok *treatment D-penicillamine*, waktu awal terjadinya *event* (pasien mengalami kematian) adalah 41 hari setelah pasien tersebut mulai mengikuti pengamatan, sedangkan waktu terakhir pasien mengalami *event* (kematian) adalah 4191 hari atau 11,5 tahun setelah pasien tersebut mulai mengikuti pengamatan. Pada kelompok NA, waktu awal terjadinya *event* (pasien mengalami kematian) adalah 41 hari setelah pasien tersebut mulai mengikuti pengamatan, sedangkan waktu terakhir pasien mengalami *event* (kematian) adalah 3561 hari atau 9,8 tahun setelah pasien tersebut mulai mengikuti pengamatan. Pada kelompok *treatment Placebo*, waktu awal terjadinya *event* (pasien mengalami kematian) adalah 51 hari setelah pasien tersebut mulai mengikuti pengamatan, sedangkan waktu terakhir pasien mengalami

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ferennia Putri, BAP Martadiputra , Lukman

event (kematian) adalah 3853 hari atau 10,6 tahun setelah pasien tersebut mulai mengikuti pengamatan.

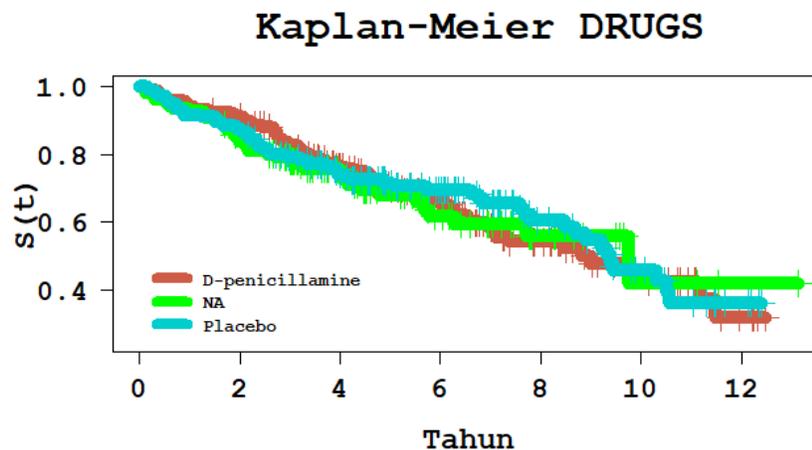
Berdasarkan *output* tersebut pula, kurva Kaplan-Meier dapat dibentuk dengan menjadikan waktu *event* dalam tahun (data yang telah dikonversi dari satuan hari) sebagai variabel sumbu-x dan peluang ketahanan hidup atau fungsi survival sebagai variabel sumbu-y. Kurva Kaplan-Meier untuk kasus ini dapat diperoleh dari *code* R sebagai berikut.

```
> plot(km_pbc, conf.int = F, mark.time = T, col = c("coral3","green","cyan3"), lty = 1, lwd = 10, cex = 2, xscale= 365, xlab = "Tahun", ylab = "S(t)", firstx = 41, ymin=0.25, main = " Kaplan-Meier DRUGS", las = 1, font.main = 11, col.main = "black",cex.main = 2, font.axis = 11, font.lab = 11, col.lab = "black", cex.lab=1.5, cex.axis=1.4)
```

```
> legend(0,0.5, legend=c("D-penicillamine ", "NA", "Placebo"),lty=1, lwd=8, col=c("coral3","green","cyan3"),text.font=11, cex=0.9 ,box.lwd=1, box.col="white")
```

Output yang dihasilkan:

Gambar 4.1 Kurva Kaplan-Meier Pasien Pengidap PBC di Mayo Clinic



Pada kurva Gambar 4.1, ketiga kelompok tersebut mengalami *event* awal terjadi dalam waktu kurang dari satu tahun sejak awal pendataan atau observasi oleh *Mayo Clinic* dengan peluang tahan hidup tertinggi mendekati 1. Sementara itu, ketiga kelompok tersebut mengalami *event* terakhir dalam waktu lebih dari 12 tahun sejak awal pendataan atau observasi oleh *Mayo Clinic* dengan peluang tahan hidup terendah mendekati nol. Hal tersebut mengartikan semakin lama peluang tahan hidup pasien PBC di *Mayo Clinic* semakin menurun.

Meskipun waktu awal dan waktu akhir *event* terjadi pada setiap kelompok berbeda yang mengakibatkan kurva Kaplan-Meier (dapat dilihat pada Gambar 4.1) setiap kelompok berbeda, pengujian perbedaan fungsi survival dalam kelompok masih perlu dilakukan. Oleh karena itu, metode *Log-Rank test* dilakukan pada tahap selanjutnya. Akan tetapi, validasi atau pengujian kebenaran dari fungsi survival atau peluang ketahanan hidup yang dihasilkan di atas perlu dilakukan terlebih dahulu.

4.2 Validasi Hasil Fungsi Survival dari Kaplan-Meier

Hasil fungsi survival yang dihasilkan dari *code* R dapat diuji kebenarannya dengan mengestimasi persamaan dari fungsi survival tersebut kemudian memilih persamaan yang mungkin dengan metode RMSE. Setelah memilih persamaan dengan error paling rendah, beberapa data terakhir fungsi survival diramal kemudian dibandingkan dengan fungsi survival hasil *code* R. Setelah itu dilanjutkan dengan *Structural Equation Model* (SEM) untuk memastikan persamaan model sesuai sehingga dapat disimpulkan fungsi survival hasil *code* R cocok dengan fungsi survival hasil persamaan model.

Berikut beberapa persamaan yang telah diestimasi dari fungsi survival *D-penicillamine* beserta error yang dihasilkan dari metode RMSE.

a. Linear

$$Y = -0.000164283933153526 X + 1.00706924341146$$

$$\varepsilon = 0.012276200750$$

b. Logaritma

$$Y = -0.127560000675321 \log X + 1.66349440165332$$

$$\varepsilon = 0.5204965066518010$$

c. Kuadrat

$$Y = (-2.09284897219916E - 10) X^2 - 0.000163646975888378 X$$

$$+ 1.00674233297752$$

$$\varepsilon = 0.012275125669$$

d. Kubik

$$Y = (9.9242968128021E - 12) X^3 + (-4.73819238628802E - 08) X^2$$

$$- 0.0001047129866848 X + 0.99142656284051$$

$$\varepsilon = 0.01006572706524$$

e. Eksponensial

$$Y = -0.000222550943867983^X + 1.04223887879983$$

$$\varepsilon = 0.29481860913567$$

Keterangan:

Y : peluang ketahanan hidup pasien / fungsi survival

X : waktu (hari)

ε : error yang dihasilkan metode RMSE

Berdasarkan error dari estimasi persamaan-persamaan di atas, persamaan kubik memiliki error yang paling kecil sehingga persamaan yang paling mendekati atau cocok untuk fungsi survival *D-penicillamine* adalah persamaan kubik yakni

$$Y = (9.9242968128021E - 12) X^3 + (-4.73819238628802E - 08) X^2$$

$$- 0.0001047129866848 X + 0.99142656284051$$

Berikut adalah data ramalan fungsi survival dari data ke-60 hingga data ke-63.

Tabel 4.1 Ramalan Fungsi Survival D-penicillamine

Data ke-	Waktu (Hari)	Fungsi Survival Hasil Code R	Fungsi Survival Hasil Ramalan Persamaan Kubik
60	3574	0.451	0.465

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ferennia Putri, BAP Martadiputra , Lukman

61	3584	0.425	0.464
62	4079	0.372	0.449
63	4191	0.319	0.451

Ramalan yang dihasilkan oleh persamaan kubik seperti yang tertera pada Tabel 4.1 belum sepenuhnya menunjukkan bahwa fungsi survival hasil *code* R dan fungsi survival hasil ramalan persamaan bersifat sesuai atau cocok meskipun dilihat dari selisih yang tidak besar atau kedua fungsi survival yang bernilai tidak berbeda jauh. Maka dari itu, setelah mendapatkan error dan ramalan fungsi survival, kecocokan antara fungsi survival dari hasil *code* R dan fungsi survival hasil ramalan persamaan yang telah ditentukan harus diuji dengan menggunakan *Structural Equation Model* (SEM). Berikut adalah hasil dari uji kecocokan SEM pada kedua sampel tersebut.

Tabel 4.2 Indeks Hasil SEM Fungsi Survival D-penicillamine Hasil Ramalan

Indeks	Nilai
CFI	1.000
TLI	1.000
RMSEA	0.000
SRMR	0.000

Tabel 4.2 menunjukkan bahwa indeks $CFI = 1.000 \geq 0.95$, $TLI = 1.000 \geq 0.95$, $RMSEA = 0.000 \leq 0.05$, dan $SRMR = 0.000 \leq 0.08$ yang artinya semua indeks tersebut memenuhi kriteria model sesuai atau bisa dikatakan fungsi survival *D-penicillamine* hasil ramalan dari persamaan yang telah ditentukan cocok dengan fungsi survival *D-penicillamine* hasil *code* R.

Berikut beberapa persamaan yang telah diestimasi dari fungsi survival kelompok pasien yang tidak diketahui *treatment*-nya (NA) beserta error yang dihasilkan dari metode RMSE.

a. Linear

$$Y = -0.000173431769294945 X + 0.979675530675612$$

$$\varepsilon = 0.013345399920111800$$

b. Logaritma

$$Y = -0.091063589201495 \log X + 1.41173940401465$$

$$\varepsilon = 0.336754435216654000$$

c. Kuadrat

$$Y = (2.30786671704398E - 08) X^2 - 0.000221107131941861 X$$

$$+ 0.995416500085636$$

$$\varepsilon = 0.0103793431$$

d. Kubik

$$Y = (2.54963081716911E - 12) X^3 + (1.51816612306367E - 08) X^2$$

$$- 0.00021487891879494 X + 0.994439868343943$$

$$\varepsilon = 0.01036709370$$

e. Eksponensial

$$Y = -0.000216892491058093^X + 0.992618855048565$$

$$\varepsilon = 0.193689476563113$$

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ferennia Putri, BAP Martadiputra , Lukman

Keterangan:

Y : peluang ketahanan hidup pasien / fungsi survival

X : waktu (hari)

ε : error yang dihasilkan metode RMSE

Berdasarkan error dari estimasi persamaan-persamaan di atas, persamaan kubik memiliki error yang paling kecil sehingga persamaan yang paling mendekati atau cocok untuk fungsi survival NA adalah persamaan kubik yakni

$$Y = (2.54963081716911E - 12) X^3 + (1.51816612306367E - 08) X^2 - 0.00021487891879494 X + 0.994439868343943$$

Berikut adalah data ramalan fungsi survival dari data ke-34 hingga data ke-36.

Tabel 4.3 Ramalan Fungsi Survival NA

Data ke-	Waktu (Hari)	Fungsi Survival Hasil Code R	Fungsi Survival Hasil Ramalan Persamaan Kubik
34	2286	0.594	0.613
35	2812	0.559	0.567
36	3561	0.419	0.537

Sama seperti tahap yang dilakukan pada kelompok *D-penicillamine*, uji kecocokan perlu dilakukan dengan metode SEM dengan hasil sebagai berikut.

Tabel 4.4 Indeks Hasil SEM Fungsi Survival Treatment NA Hasil Ramalan

Indeks	Nilai
CFI	1.000
TLI	1.000
RMSEA	0.000
SRMR	0.000

Tabel 4.4 menunjukkan bahwa indeks $CFI = 1.000 \geq 0.95$, $TLI = 1.000 \geq 0.95$, $RMSEA = 0.000 \leq 0.05$, dan $SRMR = 0.000 \leq 0.08$ yang artinya semua indeks tersebut memenuhi kriteria model sesuai atau bisa dikatakan fungsi survival *treatment* NA hasil ramalan dari persamaan yang telah ditentukan cocok dengan fungsi survival *treatment* NA hasil *code R*.

Berikut beberapa persamaan yang telah diestimasi dari fungsi survival *treatment Placebo* beserta error yang dihasilkan dari metode RMSE.

a. Linear

$$Y = -0.000131074532852683X + 0.961311080985628$$

$$\varepsilon = 0.0233907106815649$$

b. Logaritma

$$Y = -0.121449184027789 \log X + 1.61285768867344$$

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ferennia Putri, BAP Martadiputra , Lukman

$$\varepsilon = 0.47179527$$

c. Kuadrat

$$Y = (1.46480583287291E - 08) X^2 - 0.00018062843749032 X + 0.985820566707826$$

$$\varepsilon = 0.01929954262687$$

d. Kubik

$$Y = (-1.9360152260644E - 11) X^3 + (1.10097126567786E - 07) X^2 - 0.000298925675054536 X + 1.01613884025536$$

$$\varepsilon = 0.01198112931$$

e. Eksponensial

$$Y = -0.000178818875280345^X + 0.981965212535263$$

$$\varepsilon = 0.23186039109$$

Keterangan:

Y : peluang ketahanan hidup pasien / fungsi survival

X : waktu (hari)

ε : error yang dihasilkan metode RMSE

Berdasarkan error dari estimasi persamaan-persamaan di atas, persamaan kubik memiliki error yang paling kecil sehingga persamaan yang paling mendekati atau cocok untuk fungsi survival *treatment Placebo* adalah persamaan kubik yakni

$$Y = (-1.9360152260644E - 11) X^3 + (1.10097126567786E - 07) X^2 - 0.000298925675054536 X + 1.01613884025536$$

Berikut adalah data ramalan fungsi survival dari data ke-56 hingga data ke-59.

Tabel 4.5 Ramalan Fungsi Survival Placebo

Data ke-	Waktu (Hari)	Fungsi Survival Hasil Code R	Fungsi Survival Hasil Ramalan Persamaan Kubik
56	3445	0.457	0.501
57	3762	0.427	0.419
58	3839	0.394	0.396
59	3853	0.361	0.391

Sama seperti tahap yang dilakukan pada kelompok *D-penicillamine* dan NA, uji kecocokan perlu dilakukan dengan metode SEM dengan hasil sebagai berikut.

Tabel 4.6 Indeks Hasil SEM Fungsi Survival Placebo Hasil Ramalan

Indeks	Nilai
CFI	1.000
TLI	1.000
RMSEA	0.000
SRMR	0.000

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ferennia Putri, BAP Martadiputra , Lukman

Tabel 4.6 menunjukkan bahwa indeks $CFI = 1.000 \geq 0.95$, $TLI = 1.000 \geq 0.95$, $RMSEA = 0.000 \leq 0.05$, dan $SRMR = 0.000 \leq 0.08$ yang artinya semua indeks tersebut memenuhi kriteria model sesuai atau bisa dikatakan fungsi survival *Placebo* hasil ramalan dari persamaan yang telah ditentukan cocok dengan fungsi survival *Placebo* hasil *code R*.

Berdasarkan hasil SEM yang diperoleh dari ketiga kelompok, semua indeks tersebut memiliki nilai yang menunjukkan bahwa fungsi survival hasil *code R* dan fungsi survival hasil ramalan dari persamaan yang telah ditentukan bersifat sesuai atau cocok sehingga dapat dipastikan bahwa fungsi survival hasil *code R* terbukti benar.

4.3 Log-Rank Test

Setelah mendapatkan fungsi survival masing-masing kelompok yang diperoleh dari metode Kaplan-Meier, *Log-Rank Test* dapat dilakukan sebagai lanjutan dari Kaplan-Meier untuk mengetahui adanya perbedaan atau bahkan persamaan di antara kelompok tersebut meskipun secara numerik dan kurva Kaplan-Meier fungsi survival dari setiap kelompok terlihat berbeda.

Log-Rank Test membutuhkan jumlah observasi dan ekspektasi jumlah kegagalan dalam setiap kelompok. Hal tersebut dapat diperoleh dengan *code R* sebagai berikut.

```
> survdiff(Surv(N_Days, Status)~Drug, data=pbcc)
```

Output yang dihasilkan:

Table 4.7 Log-Rank Test Pasien Pengidap PBC di Mayo Clinic

	N	Observed	Expected	$\frac{(O-E)^2}{E}$
Drug=D-penicillamine	158	65	64.0	0.0156
Drug=NA	106	36	34.6	0.0567
Drug=Placebo	154	60	62.4	0.0923
Chisq= 0.2 on 2 degrees of freedom, p= 0.9				

Pada Tabel 4.7, jumlah observasi dan ekspektasi kegagalan pada setiap kelompok telah diperoleh sehingga statistik uji *Log-Rank Test* dapat langsung diperhitungkan.

$$\begin{aligned} \chi^2 &= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} \\ \Rightarrow \chi^2 &= \frac{(65 - 64)^2}{64} + \frac{(36 - 34.6)^2}{34.6} + \frac{(60 - 62.4)^2}{62.4} \\ \Rightarrow \chi^2 &= 0.0156 + 0.0567 + 0.0923 \\ \Rightarrow \chi^2 &= 0.1646 \cong 0.2 \end{aligned}$$

Selain memperhitungkan nilai χ^2 (*Chi-Square*) dengan cara di atas, nilai χ^2 juga sudah tertera pada *output* yang dapat dilihat pada Tabel 4.7 yang sama-sama bernilai 0.2.

Setelah mendapatkan nilai χ^2_{hitung} , nilai χ^2_{tabel} harus ditentukan dengan mempertimbangkan derajat kebebasan pada data. Dengan demikian diperoleh:

$$\chi^2_{tabel} = 5.991, df = 2$$

dengan

$$df = n - 1 = 3 - 1 = 2, n \text{ adalah banyaknya kelompok}$$

sehingga

$$0.2 < 5.991$$

atau

$$\chi^2_{hitung} < \chi^2_{tabel}$$

Hal di atas menyatakan bahwa H_0 diterima pada taraf signifikansi $\alpha = 0.05$ karena $\chi^2_{hitung} < \chi^2_{tabel}$ atau $S_i(t) = S_j(t), i \neq j, i, j = 1, 2, 3$ yang artinya tidak ada perbedaan nilai fungsi survival antara kelompok.

5. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa fungsi survival pasien hasil Kaplan-Meier pada setiap waktu terhadap penyakit PBC berdasarkan pemberian jenis obat yang telah diperoleh menunjukkan bahwa kurva menurun di setiap data waktu yang artinya semakin lama waktu maka pasien memiliki peluang tahan hidup semakin rendah. Selain itu, hasil dari perbandingan pemberian jenis obat terhadap pasien pengidap PBC menggunakan metode *Log-Rank* menyatakan bahwa tidak terdapat perbedaan fungsi survival antara ketiga *treatment* dengan kata lain tidak ada perbedaan peluang tahan hidup pasien pengidap PBC berdasarkan *treatment D-penicillamine, placebo*, dan pasien yang tidak diketahui *treatment* yang diberikan atau NA. Oleh karena itu, *D-penicillamine* tidak cukup baik untuk meningkatkan ketahanan hidup pasien pengidap PBC karena tidak ada dampak langsung dari *treatment D-penicillamine* terhadap peningkatan peluang tahan hidup pasien PBC sehingga penelitian ini memiliki hasil terbukti sama dengan penelitian terdahulu yang dilakukan oleh *Mayo Clinic*.

REFERENCES

- [1] Abdullah, S., 2020. Metode Kaplan-Meier. *Prodi Statistika UI*. dari www.youtube.com.
- [2] Adrian, K., 2021. Placebo, Obat Semu yang Bisa Membuat Orang Merasa Lebih Sehat. *Alodokter*. <https://www.alodokter.com/placebo-obat-semu-yang-bisa-membuat-orang-merasa-lebih-sehat>.
- [3] Budhiarta, D. M. F., 2017. Penatalaksanaan dan Edukasi Pasien Sirosis Hati dengan Varises Esofagus di RSUP Sanglah Denpasar Tahun 2014. *Intisari Sains Medis*, Vol. 8, No. 1, 19-23.
- [4] Cook, A., 2008. Kaplan Meier Estimate of S(t). *National University of Singapore*. Singapore.
- [5] Deiss, A., Lynch, R. E., Lee, G. R., & Cartwright, G. E., 1971. Long-term Therapy of Wilson's Disease. *Annals of Internal Medicine*, Vol. 75, No. 1, 57-65.
- [6] Fedesoriano, 2021. *Cirrhosis Prediction Dataset*. Retrieved [Date Retrieved] from <https://www.kaggle.com/fedesoriano/cirrhosis-prediction-dataset>.
- [7] Firsawan, I., Debatara, N. N., & Rizki, S. W., 2022. Analisis Survival pada Data Tersensor Tipe I dengan Metode Kaplan Meier. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, Vol. 11, No. 1.
- [8] Gayatri, D., 2005. Mengenal Analisis Ketahanan (Survival Analysis). *Jurnal Keperawatan Indonesia*, Vol. 9, No. 1. <https://doi.org/10.7454/jki.v9i1.158>.

- [9] Ginting, D., 2009. Structural Equation Model (SEM). *Media Informatika*, Vol. 8, No. 3, 121-123.
- [10] Gong, Y., Frederiksen, S. L., & Gluud, C., 2004. D-penicillamine for Primary Biliary Cirrhosis. *The Cochrane Database of Systematic Reviews*, 2004(4). <https://doi.org/10.1002/14651858.CD004789.pub2>.
- [11] Herrhyanto, Nar & Gantini, T., 2009. *Pengantar Statistika Matematis*. Yrama Widya, Bandung.
- [12] Khoiri, H., 2018. Analisis Data Survival dengan Survival Least Square-Svm (Studi Simulasi dan Studi Kasus Pasien Kanker Serviks). *Tesis, Institut Teknologi Sepuluh Nopember*.
- [13] James, O.F.W., 1985. D-penicillamine for Primary Biliary Cirrhosis. *Newcastle-upon-Tyne*, Vol. 26, 111.
- [14] Kleinbaum, D. G., & Klein, M., 2012. *Survival Analysis: A Self-Learning Text*, Third Edition. Springer Publisher, New York. Chapter 1, 4-13.
- [15] Kleinbaum, D. G., & Klein, M., 2012. *Survival Analysis: A Self-Learning Text*, Third Edition. Springer Publisher, New York. Chapter 2, 67-73.
- [16] Kosourova, E., 2022. *RStudio Tutorial*. Datacamp. <https://www.datacamp.com/tutorial/r-studio-tutorial>.
- [17] Latumeten, R., Lesnussa, Y.A., & Rumlawang, F.Y., 2018. Penggunaan Structural Equation Modeling (SEM) untuk Menganalisis Faktor yang Mempengaruhi Loyalitas Nasabah (Studi Kasus : PT Bank Negara Indonesia (BNI) KCU Ambon). *Sainmatika: Jurnal Ilmiah Matematika dan Ilmu Pengetahuan Alam*, Vol. 15, No. 2, 80.
- [18] Lawless, J. F., 2003. *The Statistical Analysis of Recurrent Event*. Springer Science+Business Media Inc, New York.
- [19] Maruddan, dkk., 2021. *Survival Analysis*. UNDIP Press Semarang, Semarang.
- [20] Maulid, R., 2022. *Kriteria Jenis Teknik Analisis Data dalam Forecasting*. <https://dqlab.id/kriteria-jenis-teknik-analisis-data-dalam-forecasting>.
- [21] Murtaugh P.A., Dickson, E.R., Van Dam G.M., dkk., 1994. Primary Biliary Cirrhosis: Prediction of Short-term Survival Based on Repeated Patient Visits. *Hepatology*, Vol. 20, No. 1, 126-128.
- [22] Muhajir, M., & Palupi, Y. D., 2018. Analisis Survival terhadap Pasien Diare Anak Menggunakan Metode Kaplan Meier dan Uji Log Rank. *Jurnal Ilmu-Ilmu MIPA*, Vol. 18, No. 1, 74-84.
- [23] Purba, B.T. & Sinurat, L. R., 2018. Peningkatan Status Gizi Pasien Sirosis Hepatis Melalui Regimen Nutrisi di RS Sari Mutiara Medan. *Idea Nursing Jurnal*, Vol. 9, No. 2, 1-6.
- [24] Rachmawati, R., 2020. *Model Kaplan-Meier Menggunakan R Studio*. Ramya Rachmawati, www.youtube.com.
- [25] RevoU., 2023. *9 Library Python Terbaik untuk Data Analytics*. RevoU. <https://revou.co/panduan-teknis/library>.
- [26] Robi, M., Kusnandar, D., & Sulistianingsih, E., 2017. Penerapan Structural Equation Modeling (SEM) untuk Analisis Kompetensi Alumni. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, Vol. 6, No. 2, 113-120.
- [27] Rosadi, D., 2016. *Analisis Statistika dengan R*. Gadjah Mada University Press, Yogyakarta
- [28] Ruslianah, I., 2023. Analisis Ketahanan Hidup dengan Model Cox Proportional Hazard pada Pasien Covid-19 di Rumah Sakit Paru Dr. H. A. Rotinsulu Bandung. *Skripsi Sarjana, Universitas Pendidikan Indonesia*.
- [29] Schutte, D., 2019. *Survival Analysis in R For Beginners*. Datacamp. <https://www.datacamp.com/tutorial/survival-analysis-R>.

- [30] Sternlieb, I., & Scheinberg, I. H., 1964. Penicillamine Therapy for Hepatolenticular Degeneration. *JAMA*, Vol. 189, No. 10, 748-754.
- [31] Suhartini, A., Rahmawati, R., & Suparti, S., 2018. Analisis Kurva Survival Kaplan Meier Menggunakan Uji Log Rank (Studi Kasus :Pasien Penyakit Jantung Koroner di RSUD Undata Palu). *Jurnal Gaussian*, Vol.7, No. 1, 33-42. <https://doi.org/10.14710/j.gauss.7.1.33-42>.
- [32] Sutadi, S. M., 2003. *Sirosis Hepatitis*. Internal Medicine of Medical Faculty Sumatera Utara University.