

Survival Analysis of Ovarian Carcinoma Patients in Clinical Trials

```
data(ovarian)
ovarian %>%
  mutate(fustat = as.integer(fustat)) ->
  ovarian
head(ovarian)
```

	futime <dbl>	fustat <int>	age <dbl>	resid.ds <dbl>	rx <dbl>	ecog.ps <dbl>
1	59	1	72.3315	2	1	1
2	115	1	74.4932	2	1	1
3	156	1	66.4658	2	1	2
4	421	0	53.3644	2	2	1
5	431	1	50.3397	2	1	1
6	448	0	56.4301	1	1	2

6 rows

```
table(ovarian$rx)
```

```
##
##  1  2
## 13 13
```

```
table(ovarian$fustat)
```

```
##
##  0  1
## 14 12
```

```
table(ovarian$resid.ds)
```

```
##
##  1  2
## 11 15
```

fustat: censoring status

0 = censored

1 = uncensored

rx: treatment type

1 = cyclophosphamide

2 = combined (cyclophosphamide + adriamycin)

resid.ds

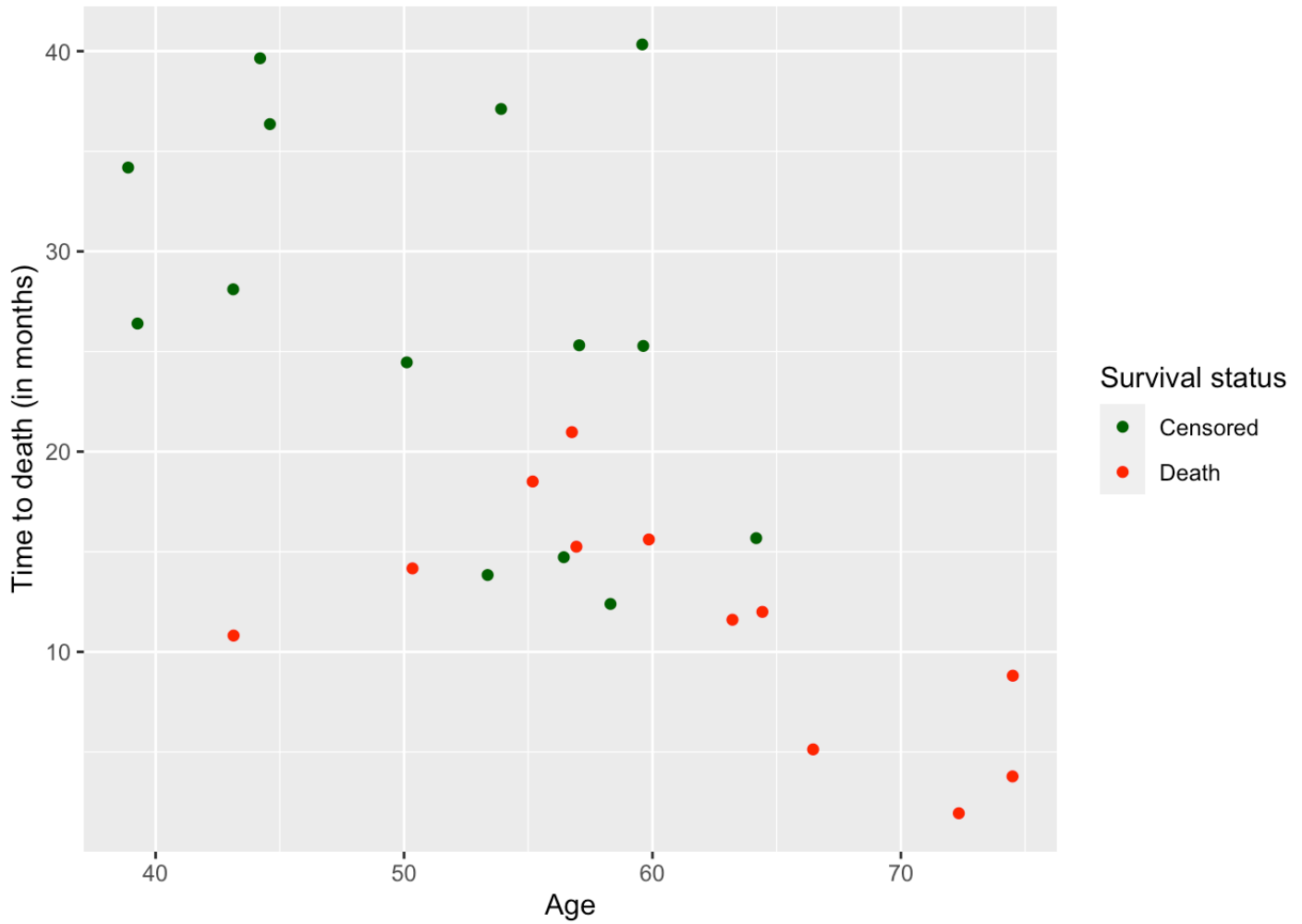
1 = microscopic residual disease

2 = macroscopic residual disease

```
summary(ovarian)
```

```
##      futime      fustat      age      resid.ds
## Min.   : 59.0   Min.   :0.0000   Min.   :38.89   Min.   :1.000
## 1st Qu.: 368.0   1st Qu.:0.0000   1st Qu.:50.17   1st Qu.:1.000
## Median : 476.0   Median :0.0000   Median :56.85   Median :2.000
## Mean   : 599.5   Mean   :0.4615   Mean   :56.17   Mean   :1.577
## 3rd Qu.: 794.8   3rd Qu.:1.0000   3rd Qu.:62.38   3rd Qu.:2.000
## Max.   :1227.0   Max.   :1.0000   Max.   :74.50   Max.   :2.000
##      rx      ecog.ps
## Min.   :1.0   Min.   :1.000
## 1st Qu.:1.0   1st Qu.:1.000
## Median :1.5   Median :1.000
## Mean   :1.5   Mean   :1.462
## 3rd Qu.:2.0   3rd Qu.:2.000
## Max.   :2.0   Max.   :2.000
```

```
ggplot(ovarian, aes(x = age, y = futime/30.42, color = factor(fustat))) +
  geom_point() +
  scale_color_manual(values = c("darkgreen", "red"), labels = c("Censored", "Death
")) +
  labs(x = "Age", y = "Time to death (in months)", color = "Survival status")
```



```
# Compute Kaplan-Meier survival curve
```

```
km.fit <- survfit(Surv(futime/30.42, fustat) ~ 1, data = ovarian)
summary(km.fit)
```

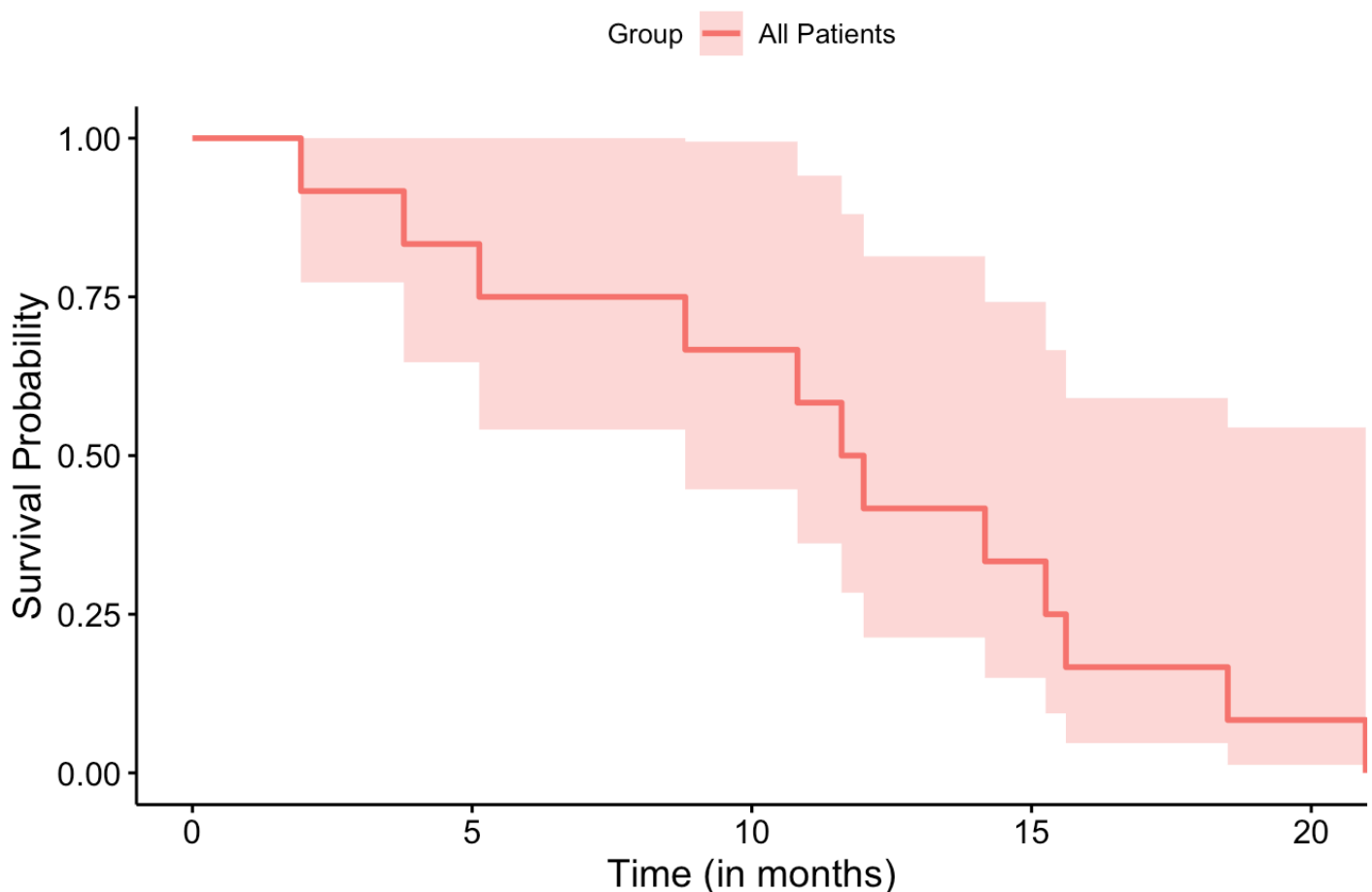
```
## Call: survfit(formula = Surv(futime/30.42, fustat) ~ 1, data = ovarian)
```

```
##
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	1.94	26	1	0.962	0.0377	0.890	1.000
##	3.78	25	1	0.923	0.0523	0.826	1.000
##	5.13	24	1	0.885	0.0627	0.770	1.000
##	8.81	23	1	0.846	0.0708	0.718	0.997
##	10.82	22	1	0.808	0.0773	0.670	0.974
##	11.60	21	1	0.769	0.0826	0.623	0.949
##	12.00	20	1	0.731	0.0870	0.579	0.923
##	14.17	17	1	0.688	0.0919	0.529	0.894
##	15.25	15	1	0.642	0.0965	0.478	0.862
##	15.61	14	1	0.596	0.0999	0.429	0.828
##	18.51	12	1	0.546	0.1032	0.377	0.791
##	20.97	11	1	0.497	0.1051	0.328	0.752

```
# Kaplan Meier on uncensored patients
uncensored_patients <- ovarian[ovarian$fustat == 1, ]
km.uncensored <- survfit(Surv(futime/30.42, fustat) ~ 1, data = uncensored_patients)
ggsurvplot(km.uncensored,
            xlab = "Time (in months)",
            ylab = "Survival Probability",
            legend.title = "Group",
            legend.labs = c("All Patients")) +
ggtitle("Kaplan-Meier on Uncensored Patients")
```

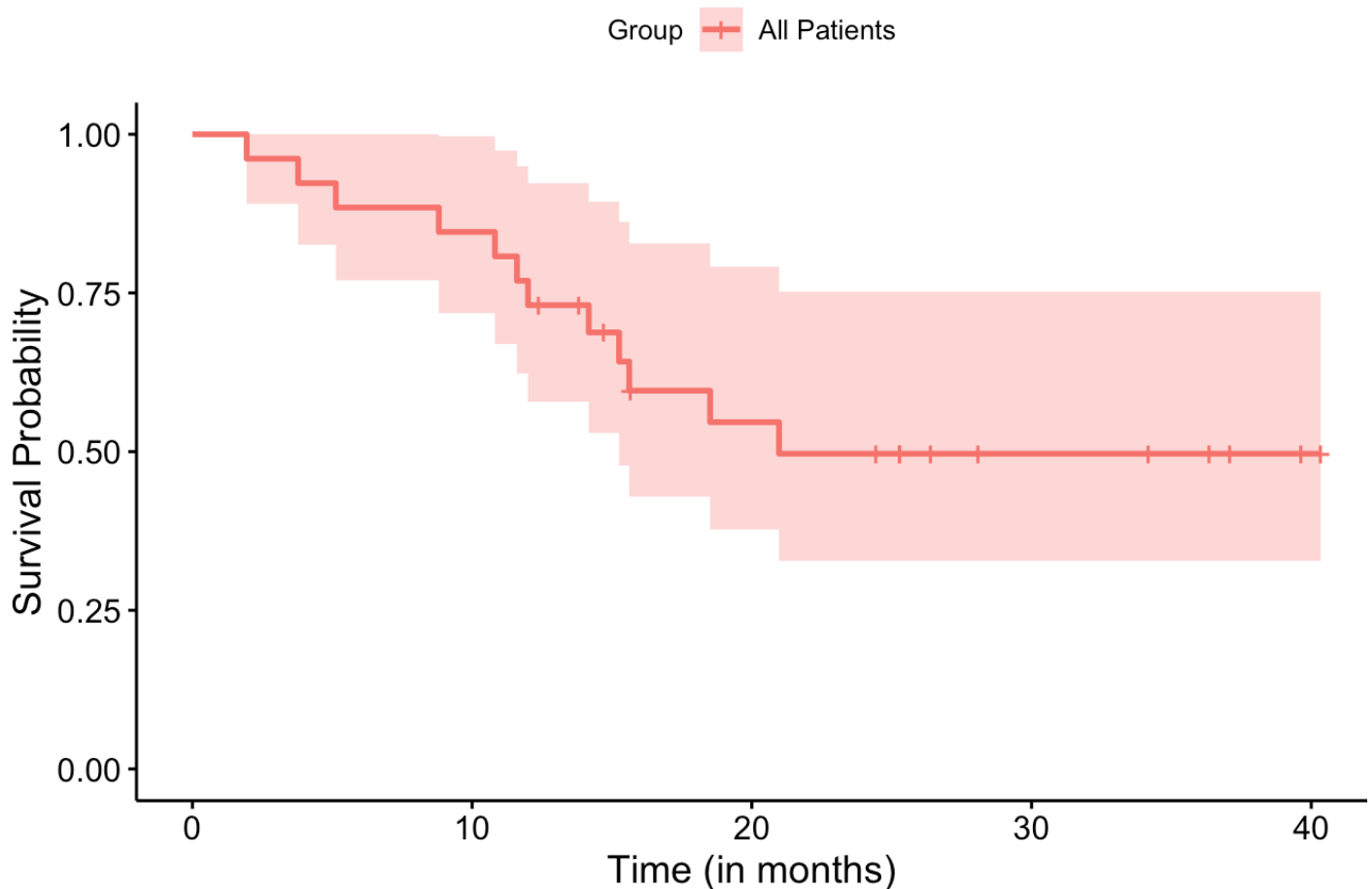
Kaplan-Meier on Uncensored Patients



Here it is easy to determine the survival rate/survival probability because we have legitimate data of each patient where he/she is alive or dead at a particular instance of time. But in theory this graph is wrong as we are not considering the uncensored patients so the graph gives us a wrong estimation of survival rate at a particular time.

```
ggsurvplot(km.fit,
            xlab = "Time (in months)",
            ylab = "Survival Probability",
            legend.title = "Group",
            legend.labs = c("All Patients")) +
ggtitle("Kaplan-Meier Survival Curve")
```

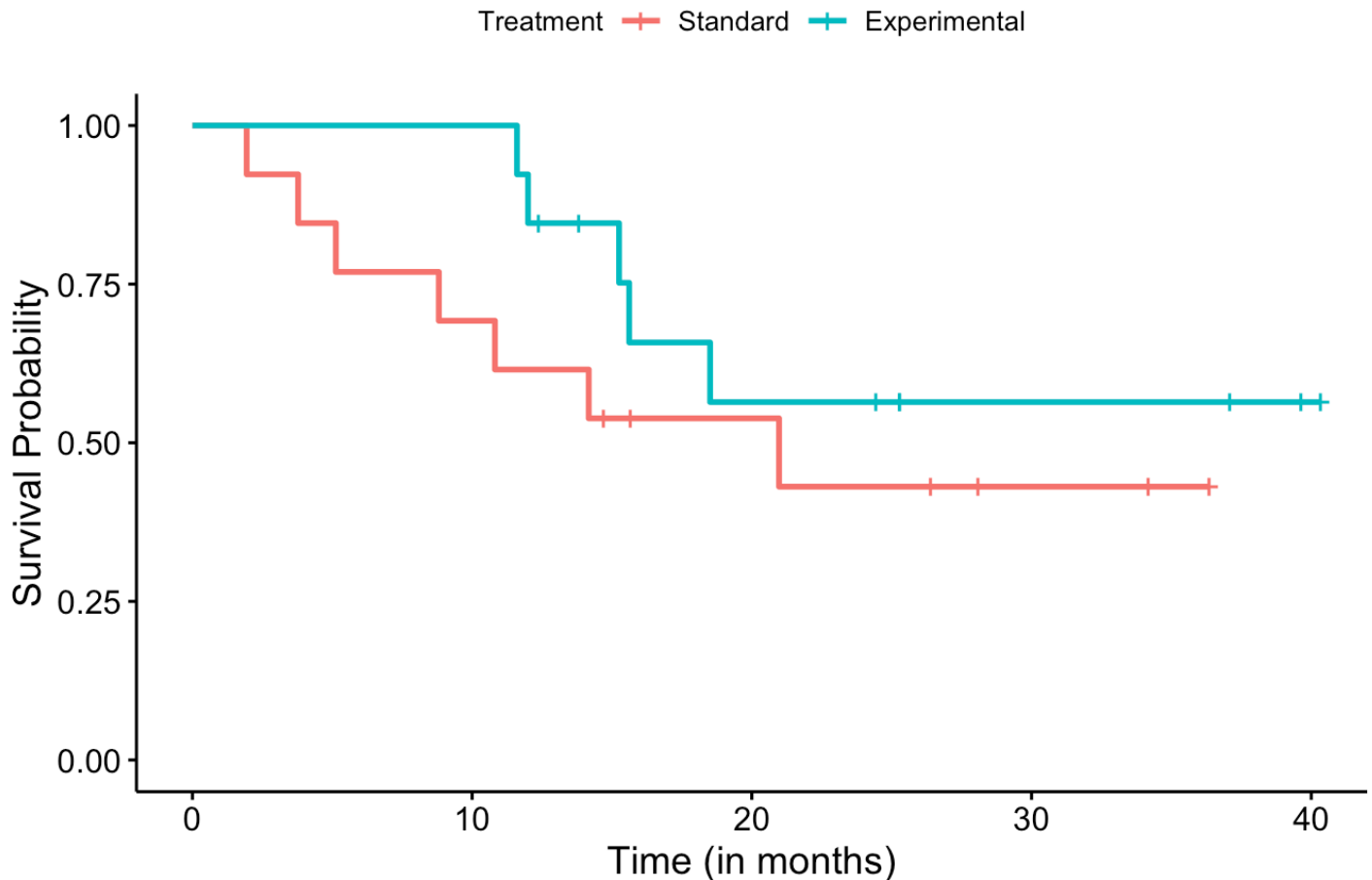
Kaplan-Meier Survival Curve



Here the vertical lines indicate an event where the person gets censored hence in order to plot the survival rate we use the law of total probability through which we can estimate the survival rate/survival probability. And if we don't consider censored patients the Kaplan-Meier curve wrongly gives the survival rate. For eg. After 5 months from the first graph the survival rate is 0.75 but after considering censored patients it outputs 0.95. Hence this shows the importance of not neglecting censored subjects.

```
# Computing Kaplan-Meier survival curves by treatment regimen
ovarian_km_rx <- survfit(Surv(futime/30.42, fustat) ~ rx, data = ovarian)
ggsurvplot(ovarian_km_rx, color = "rx",
            xlab = "Time (in months)",
            ylab = "Survival Probability",
            legend.title = "Treatment",
            legend.labs = c("Standard", "Experimental")) +
ggtitle("Kaplan-Meier Survival Curve by Treatment")
```

Kaplan-Meier Survival Curve by Treatment



From the plot, it appears that the combined approach seems to have the potential to increase the survival rate.

```
# Estimated survival probability after 2 years (24 months)
summary(km.fit, times = 24)
```

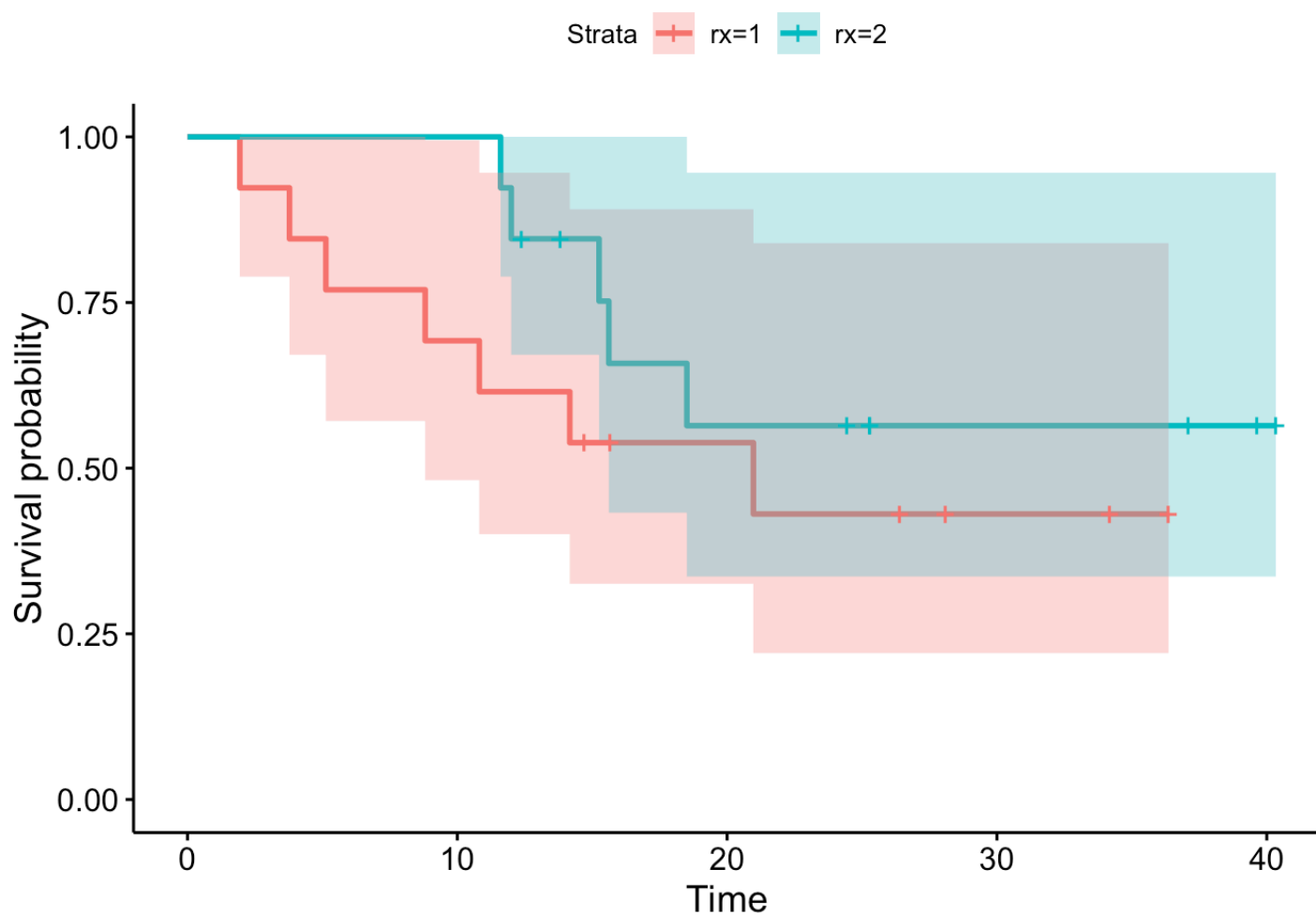
```
## Call: survfit(formula = Surv(futime/30.42, fustat) ~ 1, data = ovarian)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    24     10     12   0.497   0.105    0.328    0.752
```

```
# naive estimate for survival after 2 years (24 months)
num_deaths <- filter(ovarian, fustat==1 & futime <= 24) %>%
  nrow()
num_alive <- filter(ovarian, futime > 24) %>%
  nrow()

binom.test(c(num_alive, num_deaths))
```

```
##
## Exact binomial test
##
## data: c(num_alive, num_deaths)
## number of successes = 26, number of trials = 26, p-value = 2.98e-08
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.8677254 1.0000000
## sample estimates:
## probability of success
##                                1
```

```
ggsurvplot(ovarian_km_rx, conf.int = 0.95)
```



```
summary(ovarian_km_rx)
```

```
## Call: survfit(formula = Surv(futime/30.42, fustat) ~ rx, data = ovarian)
##
##               rx=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   1.94    13      1   0.923  0.0739    0.789    1.000
##   3.78    12      1   0.846  0.1001    0.671    1.000
##   5.13    11      1   0.769  0.1169    0.571    1.000
##   8.81    10      1   0.692  0.1280    0.482    0.995
##  10.82     9      1   0.615  0.1349    0.400    0.946
##  14.17     8      1   0.538  0.1383    0.326    0.891
##  20.97     5      1   0.431  0.1467    0.221    0.840
##
##               rx=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   11.6    13      1   0.923  0.0739    0.789    1.000
##   12.0    12      1   0.846  0.1001    0.671    1.000
##   15.3     9      1   0.752  0.1256    0.542    1.000
##   15.6     8      1   0.658  0.1407    0.433    1.000
##   18.5     7      1   0.564  0.1488    0.336    0.946
```

```
data.logrank <- survdiff(Surv(futime, fustat) ~ factor(rx), data = ovarian)
data.logrank
```

```
## Call:
## survdiff(formula = Surv(futime, fustat) ~ factor(rx), data = ovarian)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## factor(rx)=1 13         7      5.23    0.596    1.06
## factor(rx)=2 13         5      6.77    0.461    1.06
##
##   Chisq= 1.1  on 1 degrees of freedom, p= 0.3
```

The results from the log-rank test did not reveal a significant difference in survival between the treatment types (rx=1 and rx=2) for ovarian carcinoma patients in the clinical trials. However, more thorough research and larger sample sizes might be required to offer more conclusive evidence.