

statistical method

yimin chen

2023-12-06

```
library(biostat3)
## Creating times relativ to spouse death (year=0)
brv2 <- mutate(brv,
               id=NULL,
               y_before_sp_dth = as.numeric(doe -dosp) / 365.24,
               y_after_sp_dth = as.numeric(dox - dosp) / 365.24)

## Splitting at spouse death (year=0)
brvSplit <- survSplit(brv2, cut = 0, end="y_after_sp_dth", start="y_before_sp_dth", id="id", event="fail")

## Calculating risk times
brvSplit <- mutate(brvSplit,
                  t_sp_at_risk = y_after_sp_dth - y_before_sp_dth,
                  brv = ifelse(y_after_sp_dth > 0, 1, 0))
```

The mutate function from the dplyr package (assumed as it's not explicitly loaded but commonly used for such operations) is used to modify the brv data frame. Two new columns are created: y_before_sp_dth and y_after_sp_dth. These represent the number of years before and after the death of a spouse (dosp), calculated by subtracting the date of the event of interest (doe or dox) from the date of the spouse's death and converting the difference into years (dividing by 365.24, the average number of days in a year accounting for leap years).

The survSplit function from the survival package is used to split the data into periods before and after the spouse's death. This is done by specifying a cut point at year 0 (the year of the spouse's death). The function creates new observations in the dataset, splitting any observation that spans the time point 0 into two, one before and one after the spouse's death. Further Data Transformation:

Another mutate function is used to calculate two new variables: t_sp_at_risk (the time at risk after the spouse's death, calculated as the difference between y_after_sp_dth and y_before_sp_dth) and brv (a binary indicator set to 1 if the event occurred after the spouse's death, otherwise 0).

```
summary(brvSplit)
```

##	couple	dob	doe	dox
##	Min. : 1.0	Min. :1888-02-22	Min. :1981-01-15	Min. :1981-03-13
##	1st Qu.: 65.5	1st Qu.:1900-11-23	1st Qu.:1981-03-10	1st Qu.:1985-02-27
##	Median :131.0	Median :1903-02-24	Median :1981-04-08	Median :1988-09-04
##	Mean :132.0	Mean :1902-05-28	Mean :1981-04-10	Mean :1987-11-08
##	3rd Qu.:196.0	3rd Qu.:1904-10-28	3rd Qu.:1981-05-11	3rd Qu.:1991-01-01
##	Max. :266.0	Max. :1906-03-12	Max. :1981-10-23	Max. :1991-01-01
##	dosp	group	disab	health
##	Min. :1981-05-22	Min. :1.000	Min. :0.0000	Min. :0.000
##	1st Qu.:1983-10-16	1st Qu.:1.000	1st Qu.:0.0000	1st Qu.:1.000
##	Median :1986-12-14	Median :1.000	Median :0.0000	Median :2.000

```
## Mean :1989-07-20 Mean :1.544 Mean :0.5568 Mean :1.532
## 3rd Qu.:2000-01-01 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:2.000
## Max. :2000-01-01 Max. :3.000 Max. :3.0000 Max. :2.000
## sex id y_before_sp_dth y_after_sp_dth
## Min. :1.000 Min. : 1.0 Min. : -18.960 Min. : -18.804
## 1st Qu.:1.000 1st Qu.:111.5 1st Qu.: -18.618 1st Qu.: -9.000
## Median :1.000 Median :221.0 Median : -4.288 Median : 0.000
## Mean :1.468 Mean :210.8 Mean : -7.259 Mean : -2.871
## 3rd Qu.:2.000 3rd Qu.:309.5 3rd Qu.: 0.000 3rd Qu.: 0.690
## Max. :2.000 Max. :399.0 Max. : 0.000 Max. : 9.583
## fail t_sp_at_risk brv
## Min. :0.0000 Min. :0.008214 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:1.794710 1st Qu.:0.0000
## Median :1.0000 Median :3.926186 Median :0.0000
## Mean :0.5009 Mean :4.388663 Mean :0.2811
## 3rd Qu.:1.0000 3rd Qu.:6.654529 3rd Qu.:1.0000
## Max. :1.0000 Max. :9.889388 Max. :1.0000
```

```
library(skimr)
skimr::skim(brvSplit)
```

Table 1: Data summary

Name	brvSplit
Number of rows	555
Number of columns	15
Column type frequency:	
Date	4
numeric	11
Group variables	None

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
dob	0	1	1888-02-22	1906-03-12	1903-02-24	376
doe	0	1	1981-01-15	1981-10-23	1981-04-08	93
dox	0	1	1981-03-13	1991-01-01	1988-09-04	264
dosp	0	1	1981-05-22	2000-01-01	1986-12-14	235

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
couple	0	1	131.99	76.60	1.00	65.50	131.00	196.00	266.00	
group	0	1	1.54	0.72	1.00	1.00	1.00	2.00	3.00	
disab	0	1	0.56	0.97	0.00	0.00	0.00	1.00	3.00	
health	0	1	1.53	0.61	0.00	1.00	2.00	2.00	2.00	
sex	0	1	1.47	0.50	1.00	1.00	1.00	2.00	2.00	
id	0	1	210.77	115.77	1.00	111.50	221.00	309.50	399.00	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
y_before_sp_dth	0	1	-7.26	7.67	-	-	-4.29	0.00	0.00	
y_after_sp_dth	0	1	-2.87	7.04	-	-9.00	0.00	0.69	9.58	
fail	0	1	0.50	0.50	0.00	0.00	1.00	1.00	1.00	
t_sp_at_risk	0	1	4.39	2.99	0.01	1.79	3.93	6.65	9.89	
brv	0	1	0.28	0.45	0.00	0.00	0.00	1.00	1.00	

```
brvSplit2 <- mutate(brvSplit,
  sex = as.factor(sex),
  brv = as.factor(brv))
## Translate time scale from years from spouse death to ages
brvSplit3 <- brvSplit2 %>%
  mutate(age_sp_dth = as.numeric(dosp - dob) / 365.24, # Age at spouse death
    age_start = age_sp_dth + y_before_sp_dth, # Age at start of timeband
    age_end = age_sp_dth + y_after_sp_dth) # Age at end of timeband

age_cat <- seq(70,100,5) # Split at these ages
brvSplit4 <- survSplit(brvSplit3, cut=age_cat, start="age_start", end="age_end", event="fail", zero = 0)

brvSplit4 <- mutate(brvSplit4,
  t_at_risk = age_end - age_start, # Creating new time at risk
  age = cut(age_end, age_cat)) # Creating age band category

## Calculate crude rates
survRate(Surv(t_at_risk, fail) ~ age, data=brvSplit4)

##          age      tstop event      rate      lower      upper
## age=(75,80] (75,80] 703.612419    45 0.06395566 0.04664970 0.08557771
## age=(80,85] (80,85] 1184.684043   123 0.10382515 0.08628885 0.12387811
## age=(85,90] (85,90] 490.021356    95 0.19386910 0.15685168 0.23699492
## age=(90,95] (90,95] 55.090352    12 0.21782399 0.11255283 0.38049467
## age=(95,100] (95,100] 2.299858     3 1.30442857 0.26900453 3.81209383

summary(coxph(Surv(age_start, age_end, fail) ~ brv,
  data = brvSplit4))
```

```
## Call:
## coxph(formula = Surv(age_start, age_end, fail) ~ brv, data = brvSplit4)
##
##      n= 1036, number of events= 278
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## brv1 -0.2070      0.8131   0.1390 -1.488   0.137
##
##      exp(coef) exp(-coef) lower .95 upper .95
## brv1    0.8131      1.23    0.6191    1.068
##
## Concordance= 0.511 (se = 0.014 )
## Likelihood ratio test= 2.26 on 1 df,  p=0.1
## Wald test = 2.22 on 1 df,  p=0.1
## Score (logrank) test = 2.22 on 1 df,  p=0.1
```

```
summary(coxph(Surv(age_start, age_end, fail) ~ brv + sex,
              data = brvSplit4))
```

```
## Call:
## coxph(formula = Surv(age_start, age_end, fail) ~ brv + sex, data = brvSplit4)
##
##      n= 1036, number of events= 278
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## brv1 -0.07842    0.92458  0.14245 -0.551 0.581971
## sex2 -0.47291    0.62318  0.13075 -3.617 0.000298 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## brv1  0.9246      1.082    0.6993    1.2224
## sex2  0.6232      1.605    0.4823    0.8052
##
## Concordance= 0.56 (se = 0.018 )
## Likelihood ratio test= 15.85 on 2 df,  p=4e-04
## Wald test               = 15.21 on 2 df,  p=5e-04
## Score (logrank) test = 15.51 on 2 df,  p=4e-04
```

```
summary(coxph(Surv(age_start, age_end, fail) ~ brv,
              data = brvSplit4))
```

```
## Call:
## coxph(formula = Surv(age_start, age_end, fail) ~ brv, data = brvSplit4)
##
##      n= 1036, number of events= 278
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## brv1 -0.2070    0.8131  0.1390 -1.488  0.137
##
##      exp(coef) exp(-coef) lower .95 upper .95
## brv1  0.8131      1.23    0.6191    1.068
##
## Concordance= 0.511 (se = 0.014 )
## Likelihood ratio test= 2.26 on 1 df,  p=0.1
## Wald test               = 2.22 on 1 df,  p=0.1
## Score (logrank) test = 2.22 on 1 df,  p=0.1
```

```
summary(coxph(Surv(age_start, age_end, fail) ~ brv + sex,
              data = brvSplit4))
```

```
## Call:
## coxph(formula = Surv(age_start, age_end, fail) ~ brv + sex, data = brvSplit4)
##
##      n= 1036, number of events= 278
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## brv1 -0.07842    0.92458  0.14245 -0.551 0.581971
## sex2 -0.47291    0.62318  0.13075 -3.617 0.000298 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##      exp(coef) exp(-coef) lower .95 upper .95
## brv1    0.9246    1.082    0.6993    1.2224
## sex2    0.6232    1.605    0.4823    0.8052
##
## Concordance= 0.56 (se = 0.018 )
## Likelihood ratio test= 15.85 on 2 df,  p=4e-04
## Wald test              = 15.21 on 2 df,  p=5e-04
## Score (logrank) test = 15.51 on 2 df,  p=4e-04
write.csv(brvSplit4, "brvSplit4.csv")

surv_object <- Surv(time = brvSplit$dox-brvSplit$doe, event = brvSplit$fail)

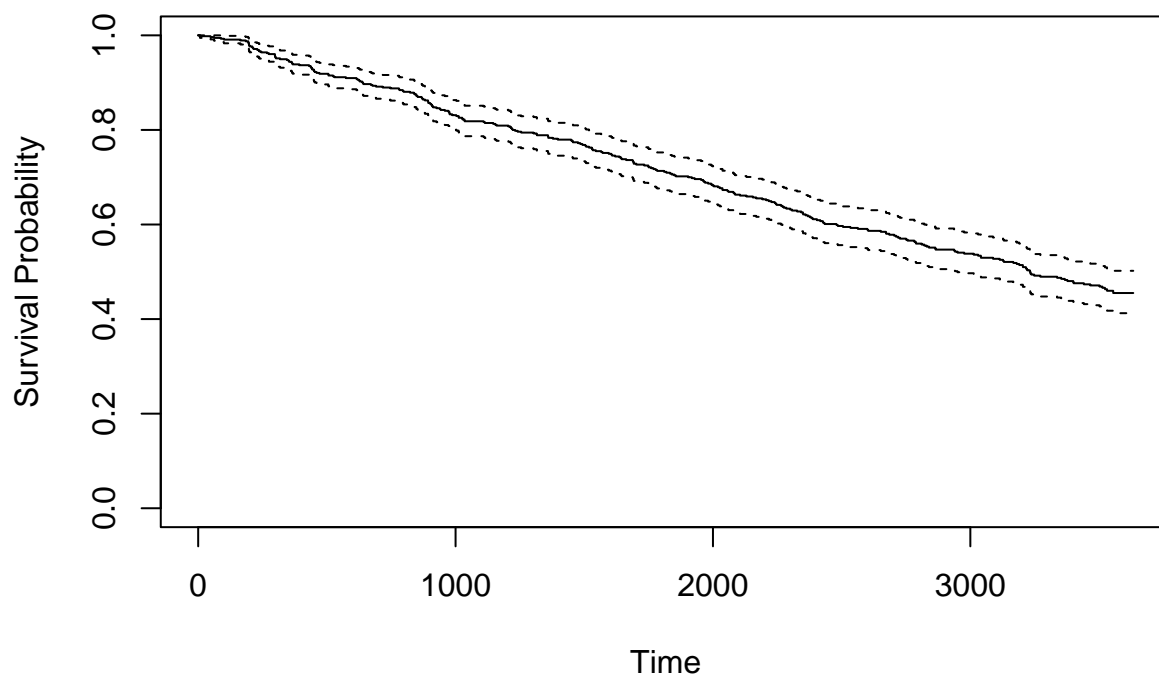
# Generate the life table using Kaplan-Meier estimate
life_table <- survfit(surv_object ~ 1)

# Print the life table
print(life_table)

## Call: survfit(formula = surv_object ~ 1)
##
##      n events median 0.95LCL 0.95UCL
## [1,] 555    278   3233   2950      NA
life_table

## Call: survfit(formula = surv_object ~ 1)
##
##      n events median 0.95LCL 0.95UCL
## [1,] 555    278   3233   2950      NA
plot(life_table, main = "Survival Curve", xlab = "Time", ylab = "Survival Probability")
```

Survival Curve



male

```
lifetable1=lifetab2(Surv(time = brvSplit$dox-brvSplit$doe, brvSplit$fail==1) ~ 1, brvSplit[brvSplit$sex
print(lifetable1)
```

##	tstart	tstop	nsubs	nlost	nrisk	nevent	surv	pdf
## 0-300	0	300	295	1	294.5	25	1.00000000	0.0002829655
## 300-600	300	600	269	3	267.5	25	0.91511036	0.0002850811
## 600-900	600	900	241	4	239.0	29	0.82958602	0.0003355369
## 900-1200	900	1200	208	3	206.5	26	0.72892496	0.0003059249
## 1200-1500	1200	1500	179	7	175.5	22	0.63714748	0.0002662345
## 1500-1800	1500	1800	150	8	146.0	29	0.55727714	0.0003689735
## 1800-2100	1800	2100	113	9	108.5	26	0.44658511	0.0003567193
## 2100-2400	2100	2400	78	14	71.0	26	0.33956932	0.0004144978
## 2400-2700	2400	2700	38	1	37.5	16	0.21521999	0.0003060907
## 2700-3000	2700	3000	21	13	14.5	19	0.12339280	0.0005389570
## 3000-Inf	3000	Inf	-11	214	-118.0	35	-0.03829432	NA
##	hazard	se.surv	se.pdf	se.hazard				
## 0-300	0.0002955083	0.00000000	5.413775e-05	5.904356e-05				
## 300-600	0.0003267974	0.01624132	5.452184e-05	6.528090e-05				
## 600-900	0.0004305865	0.02195479	5.907645e-05	7.979095e-05				
## 900-1200	0.0004478898	0.02606034	5.714913e-05	8.763995e-05				
## 1200-1500	0.0004457953	0.02832075	5.438765e-05	9.483115e-05				
## 1500-1800	0.0007351077	0.02944824	6.436005e-05	1.356737e-04				
## 1800-2100	0.0009075044	0.02992482	6.551898e-05	1.763196e-04				
## 2100-2400	0.0014942529	0.02920041	7.388269e-05	2.855912e-04				
## 2400-2700	0.0018079096	0.02682268	6.937243e-05	4.350404e-04				
## 2700-3000	0.0126666667	0.02320880	7.437499e-05	NaN				
## 3000-Inf	NA	NaN	NA	NA				

female

```
lifetable2=lifetab2(Surv(time = brvSplit$dox-brvSplit$doe, brvSplit$fail==1) ~ 1, brvSplit[brvSplit$sex==1,])
print(lifetable2)
```

##	tstart	tstop	nsubs	nlost	nrisk	nevent	surv	pdf
## 0-300	0	300	260	1	259.5	25	1.00000000	0.0003211304
## 300-600	300	600	234	3	232.5	25	0.90366089	0.0003238928
## 600-900	600	900	206	4	204.0	29	0.80649305	0.0003821617
## 900-1200	900	1200	173	3	171.5	26	0.69184453	0.0003496202
## 1200-1500	1200	1500	144	7	140.5	22	0.58695848	0.0003063603
## 1500-1800	1500	1800	115	8	111.0	29	0.49505039	0.0004311250
## 1800-2100	1800	2100	78	9	73.5	26	0.36571290	0.0004312261
## 2100-2400	2100	2400	43	14	36.0	26	0.23634507	0.0005689789
## 2400-2700	2400	2700	3	1	2.5	16	0.06565141	0.0014005634
## 2700-3000	2700	3000	-14	13	-20.5	19	-0.35451761	0.0010952576
## 3000-Inf	3000	Inf	-46	214	-153.0	35	-0.68309490	NA
##	hazard		se.surv		se.pdf		se.hazard	
## 0-300	0.0003373819		0.00000000		6.105400e-05		6.738992e-05	
## 300-600	0.0003787879		0.01831620		6.154794e-05		7.563519e-05	
## 600-900	0.0005101143		0.02458190		6.675243e-05		9.444814e-05	
## 900-1200	0.0005467928		0.02887035		6.481848e-05		1.068736e-04	
## 1200-1500	0.0005662806		0.03096618		6.212426e-05		1.202951e-04	
## 1500-1800	0.0010017271		0.03171679		7.414652e-05		1.839042e-04	
## 1800-2100	0.0014325069		0.03122682		7.731692e-05		2.743754e-04	
## 2100-2400	0.0037681159		0.02869232		9.071912e-05		6.096225e-04	
## 2400-2700	-0.0096969697		0.01935995		NaN		NaN	
## 2700-3000	-0.0021111111		NaN		NaN		4.593974e-04	
## 3000-Inf		NA	NaN		NA		NA	

KM and FH

```
fit <- brvSplit%>%
  survfit(Surv(brvSplit$dox-brvSplit$doe, fail==1) ~ brv, data = .)

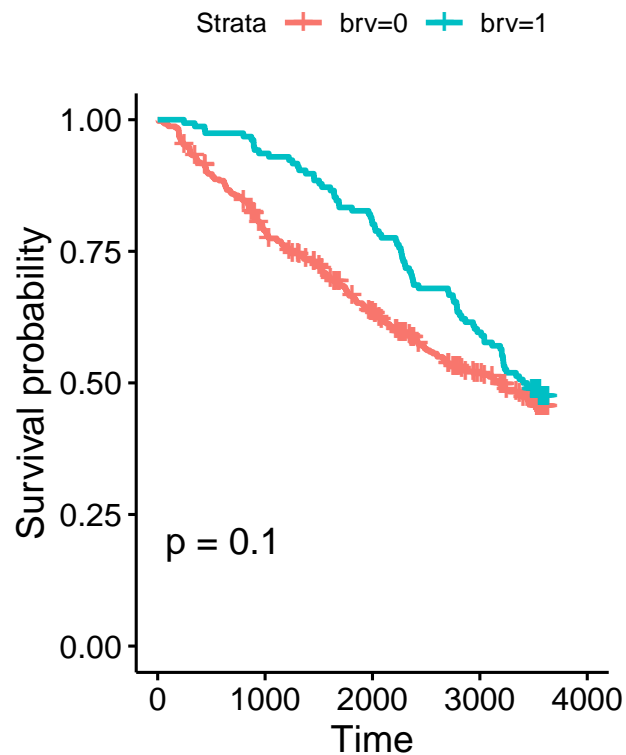
fit2 <- brvSplit %>%
  survfit(Surv(brvSplit$dox-brvSplit$doe, fail==0) ~ brv, data = ., type = "fleming")

plots <- list()

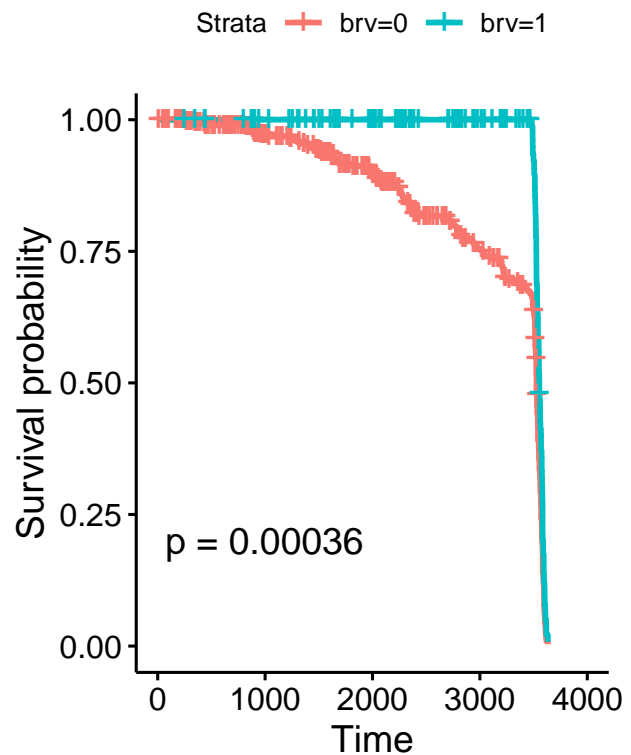
plots[[1]] <- ggsurvplot(fit, data = brvSplit, pval = TRUE, title = "Kaplan-Meier")
plots[[2]] <- ggsurvplot(fit2, data = brvSplit, pval = TRUE, title = "Fleming-Harrington")

arrange_ggsurvplots(plots, print = TRUE,
  ncol = 2, nrow = 1)
```

Kaplan–Meier



Fleming–Harrington



```
fit3 <- brvSplit%>%
  survfit(Surv(brvSplit$dox-brvSplit$doe, fail==1) ~ sex, data = .)

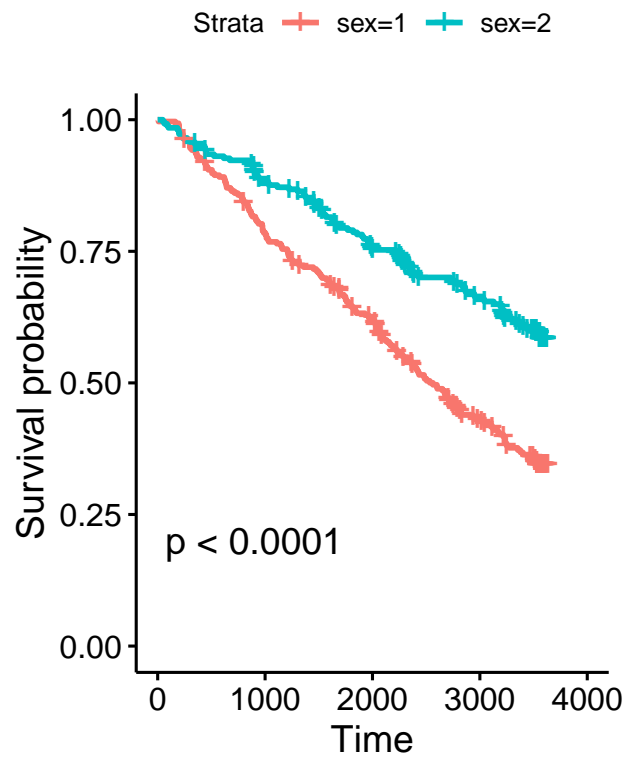
fit4 <- brvSplit %>%
  survfit(Surv(brvSplit$dox-brvSplit$doe, fail==0) ~ sex, data = ., type = "fleming")

plots <- list()

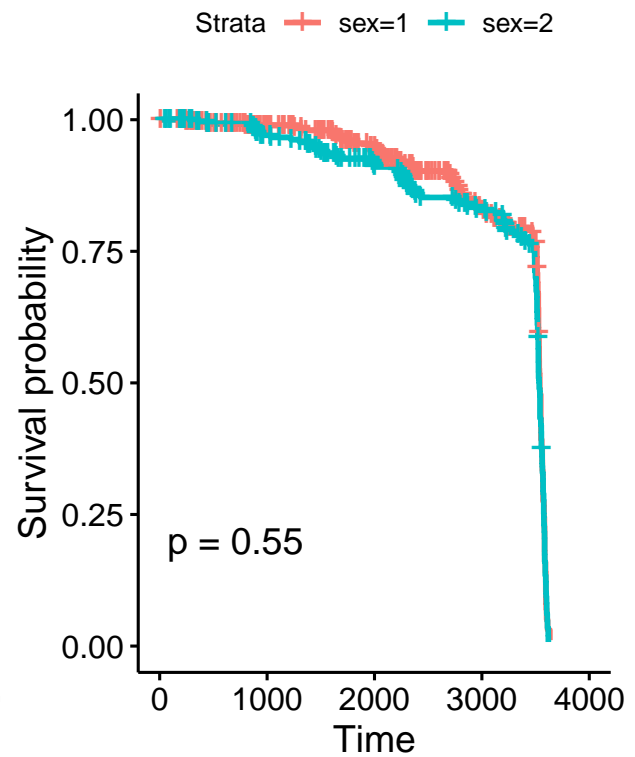
plots[[1]] <- ggsurvplot(fit3, data = brvSplit, pval = TRUE, title = "Kaplan-Meier")
plots[[2]] <- ggsurvplot(fit4, data = brvSplit, pval = TRUE, title = "Fleming-Harrington")

arrange_ggsurvplots(plots, print = TRUE,
  ncol = 2, nrow = 1)
```


Kaplan–Meier

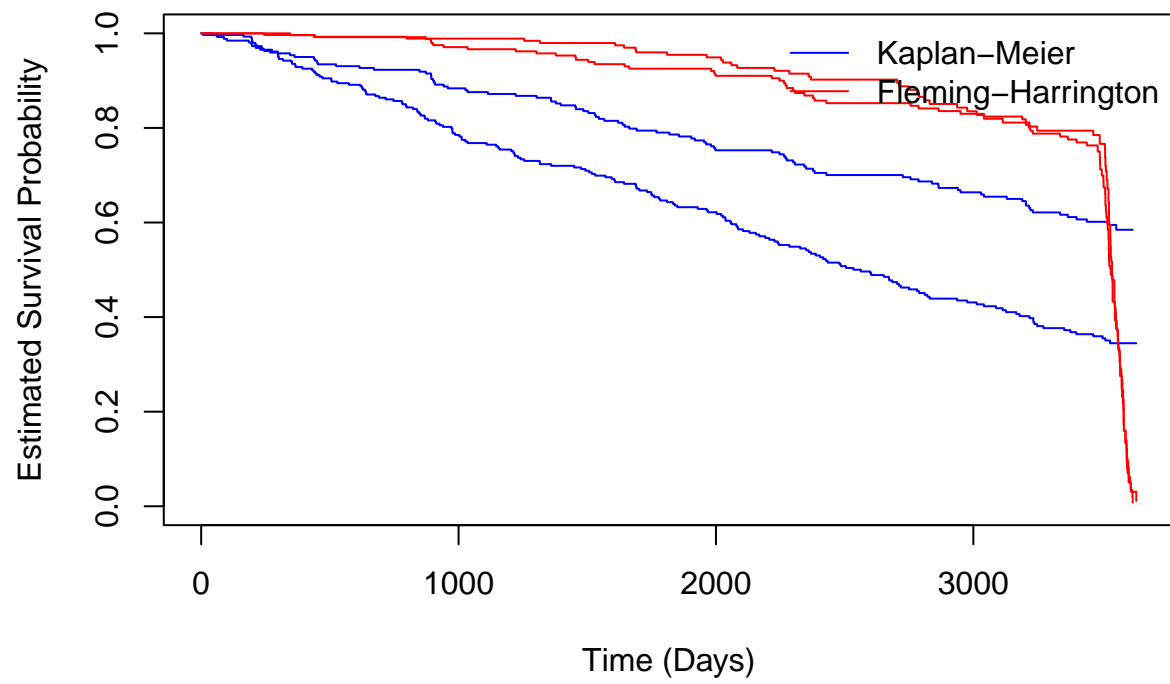


Fleming–Harrington



```
plot(fit3, conf.int = FALSE, col = "blue",
     xlab = "Time (Days)", ylab = "Estimated Survival Probability",
     main = "Comparison of S(t) between K-M and F-H methods")
lines(fit4, conf.int = FALSE, col = "red")
legend("topright", c("Kaplan-Meier", "Fleming-Harrington"),
     col = c("blue", "red"), lty = 1, bty = "n")
```

Comparison of $S(t)$ between K-M and F-H methods

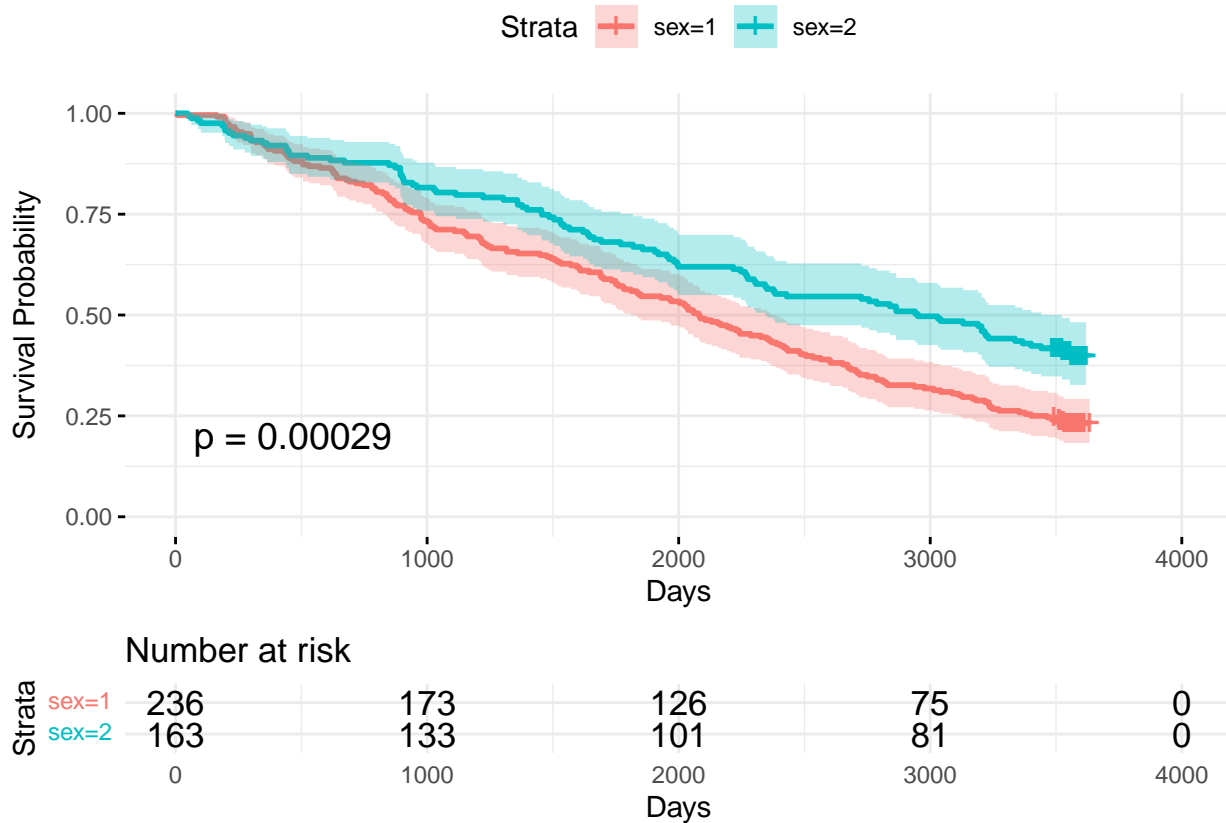


```

surv_obj <- Surv(time = brv$dox-brv$doe, event = brv$fail)

km_fit1 <- survfit(surv_obj ~ brv$sex)
ggsurvplot(km_fit1, data= brv,
  pval = TRUE,
  conf.int = TRUE,
  risk.table = TRUE,
  ggtheme = theme_minimal(),
  palette = "Dark",
  main = "Kaplan-Meier Survival Curve",
  xlab = "Days",
  ylab = "Survival Probability")

```

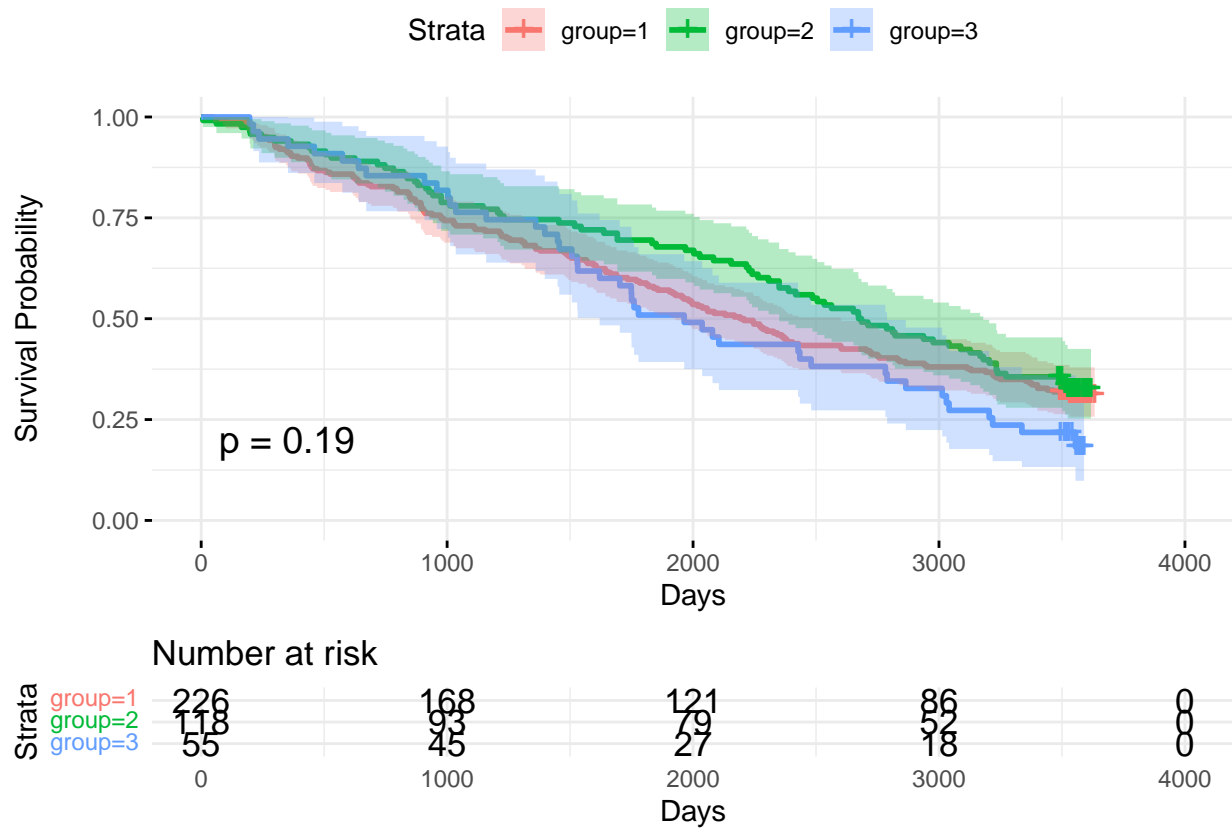


```

surv_obj <- Surv(time = brv$dox-brv$doe, event = brv$fail)

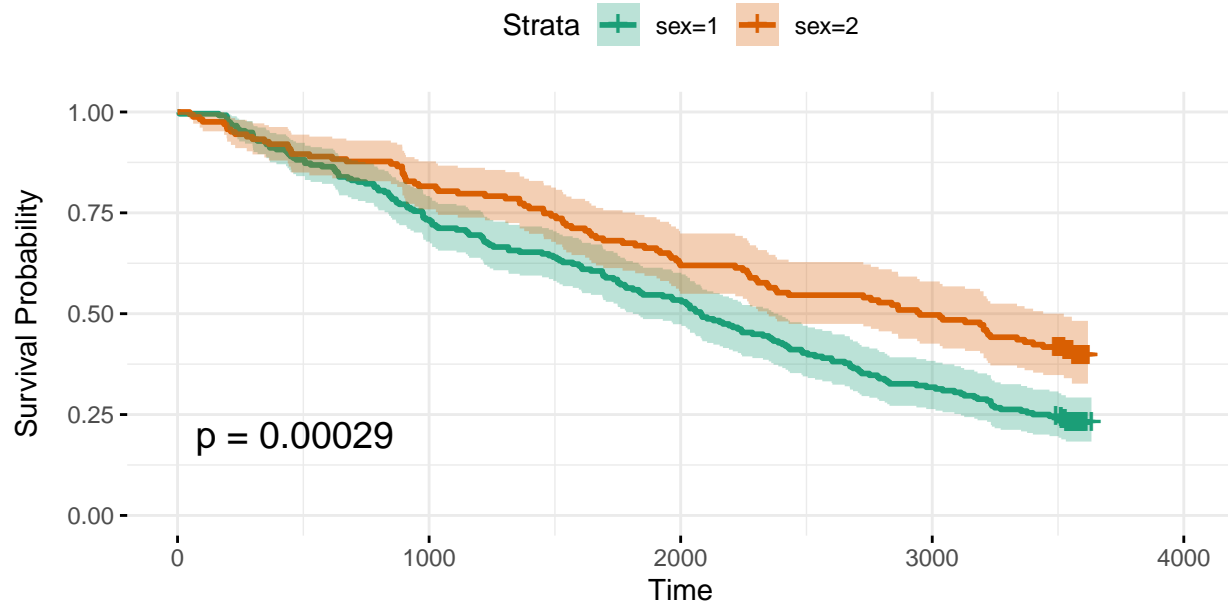
km_fit1 <- survfit(surv_obj ~ brv$group)
ggsurvplot(km_fit1, data= brv,
  pval = TRUE,
  conf.int = TRUE,
  risk.table = TRUE,
  ggtheme = theme_minimal(),
  palette = "Dark",
  main = "Kaplan-Meier Survival Curve",
  xlab = "Days",
  ylab = "Survival Probability")

```



```
km_fit2 <- survfit(surv_obj ~ brv$sex)

ggsurvplot(km_fit2, data=brv,
  pval = TRUE,
  conf.int = TRUE,
  risk.table = TRUE,
  ggtheme = theme_minimal(),
  palette = "Dark2",
  main = "Kaplan-Meier Survival Curve",
  xlab = "Time",
  ylab = "Survival Probability")
```



Number at risk

Strata					
+ sex=1	236	173	126	75	0
+ sex=2	163	133	101	81	0
	0	1000	2000	3000	4000
	Time				

```
## Call:
## survdiff(formula = surv_obj ~ group, data = brv)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## group=1 226      155   153.2    0.0212    0.0472
## group=2 118       79    89.5    1.2217    1.8051
## group=3  55       44    35.3    2.1183    2.4336
##
##  Chisq= 3.4  on 2 degrees of freedom, p= 0.2
```

Log-rank test (death as event) comparing group

```
surv_obj1 <- Surv(time = brv$dox-brv$doe, event = brv$fail)

log_rank_test <- survdiff(surv_obj1 ~ group, data = brv)

print(log_rank_test)
```

```
## Call:
## survdiff(formula = surv_obj1 ~ group, data = brv)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## group=1 226      155   153.2    0.0212    0.0472
## group=2 118       79    89.5    1.2217    1.8051
## group=3  55       44    35.3    2.1183    2.4336
##
##  Chisq= 3.4  on 2 degrees of freedom, p= 0.2
```

```

surv_obj1 <- Surv(time = brv$dox-brv$doe, event = brv$fail)

log_rank_test2 <- survdiff(surv_obj1 ~ sex, data = brv)

print(log_rank_test2)

```

```

## Call:
## survdiff(formula = surv_obj1 ~ sex, data = brv)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 236      181      151      5.95      13.1
## sex=2 163       97      127      7.08      13.1
##
##  Chisq= 13.1 on 1 degrees of freedom, p= 3e-04

```

```

# Creating the survival object
surv_obj <- Surv(time = brv$dox - brv$doe, event = brv$fail)

# Fit Cox model (specify variables or use '.' for all variables)
cox_model <- coxph(surv_obj ~ ., data = brv)

```

```

## Warning in coxph.fit(X, Y, istrat, offset, init, control, weights = weights, :
## Ran out of iterations and did not converge

## Warning in coxph.fit(X, Y, istrat, offset, init, control, weights = weights, :
## one or more coefficients may be infinite

```

```

summary(cox_model)

```

```

## Call:
## coxph(formula = surv_obj ~ ., data = brv)
##
##   n= 399, number of events= 278
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## id          2.802e-06  1.000e+00  1.184e-05   0.237   0.813
## couple    -2.653e-03  9.974e-01  1.645e-03  -1.613   0.107
## dob        4.051e-05  1.000e+00  8.990e-05   0.451   0.652
## doe        3.118e-01  1.366e+00  2.607e-03  119.608 <2e-16 ***
## dox       -3.110e-01  7.327e-01  2.602e-03 -119.551 <2e-16 ***
## dosp       5.955e-06  1.000e+00  4.813e-05   0.124   0.902
## fail       6.643e+00  7.671e+02  5.758e+00   1.154   0.249
## group     -8.489e-02  9.186e-01  1.577e-01  -0.538   0.590
## disab      2.197e-02  1.022e+00  1.162e-01   0.189   0.850
## health    -8.354e-03  9.917e-01  2.094e-01  -0.040   0.968
## sex        7.317e-02  1.076e+00  2.562e-01   0.286   0.775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## id              1.0000   0.999997  0.999980  1.000e+00
## couple          0.9974   1.002656  0.994140  1.001e+00
## dob              1.0000   0.999959  0.999864  1.000e+00
## doe             1.3658   0.732151  1.358878  1.373e+00
## dox              0.7327   1.364817  0.728972  7.364e-01
## dosp             1.0000   0.999994  0.999912  1.000e+00

```

```
## fail      767.0831    0.001304    0.009628 6.111e+07
## group      0.9186    1.088598    0.674428 1.251e+00
## disab      1.0222    0.978270    0.813937 1.284e+00
## health     0.9917    1.008389    0.657892 1.495e+00
## sex        1.0759    0.929441    0.651162 1.778e+00
##
## Concordance= 1 (se = 0 )
## Likelihood ratio test= 2947 on 11 df, p=<2e-16
## Wald test              = 28603 on 11 df, p=<2e-16
## Score (logrank) test = 906.5 on 11 df, p=<2e-16
```

```
# Check proportional hazards assumption
cox.zph(cox_model)
```

```
##          chisq df    p
## id      NaN  1 NaN
## couple  NaN  1 NaN
## dob     NaN  1 NaN
## doe     NaN  1 NaN
## dox     NaN  1 NaN
## dosp    NaN  1 NaN
## fail    NaN  1 NaN
## group   NaN  1 NaN
## disab   NaN  1 NaN
## health  NaN  1 NaN
## sex     NaN  1 NaN
## GLOBAL  NaN 11 NaN
```

Kaplan-Meier and Fleming-Harrington model

For nonparametric estimator, Kaplan-Meier(KM) model and Fleming-Harrington(FH) model were used to measure the fraction of subjects living for a certain amount of time after treatment with the stratify of sex.[3]

The Kaplan-Meier estimator

$$\hat{S}_K(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t} [1 - \frac{d_i}{n_i}] & \text{if } t \geq t_1 \end{cases}$$

note: d_i = # of failure at time t_i , n_i = # at risk at t_i^- , c_i = # censored during the interval $[t_i, t_{i+1}]$

The Fleming-Harrington estimator

$$\hat{S}_F(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t} \exp[-\frac{d_i}{n_i}] & \text{if } t \geq t_1 \end{cases}$$