statistical method

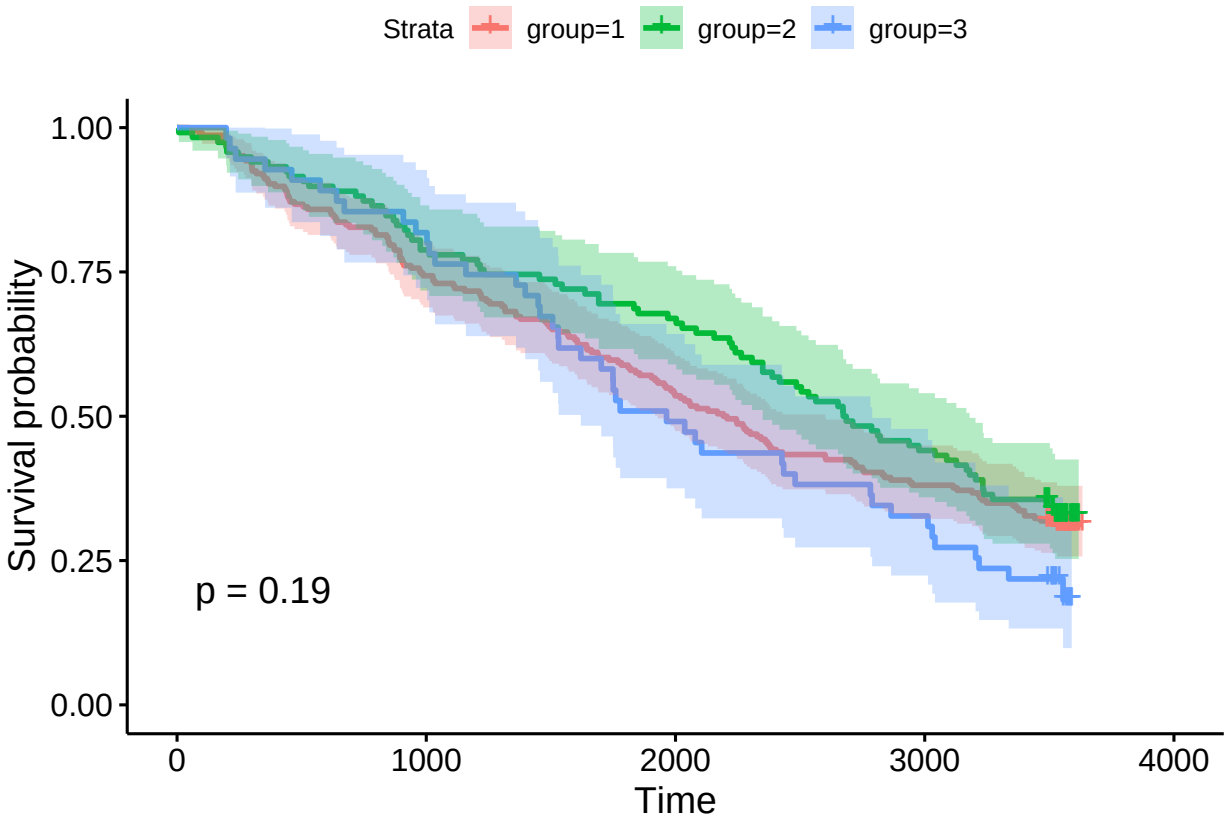
yimin chen

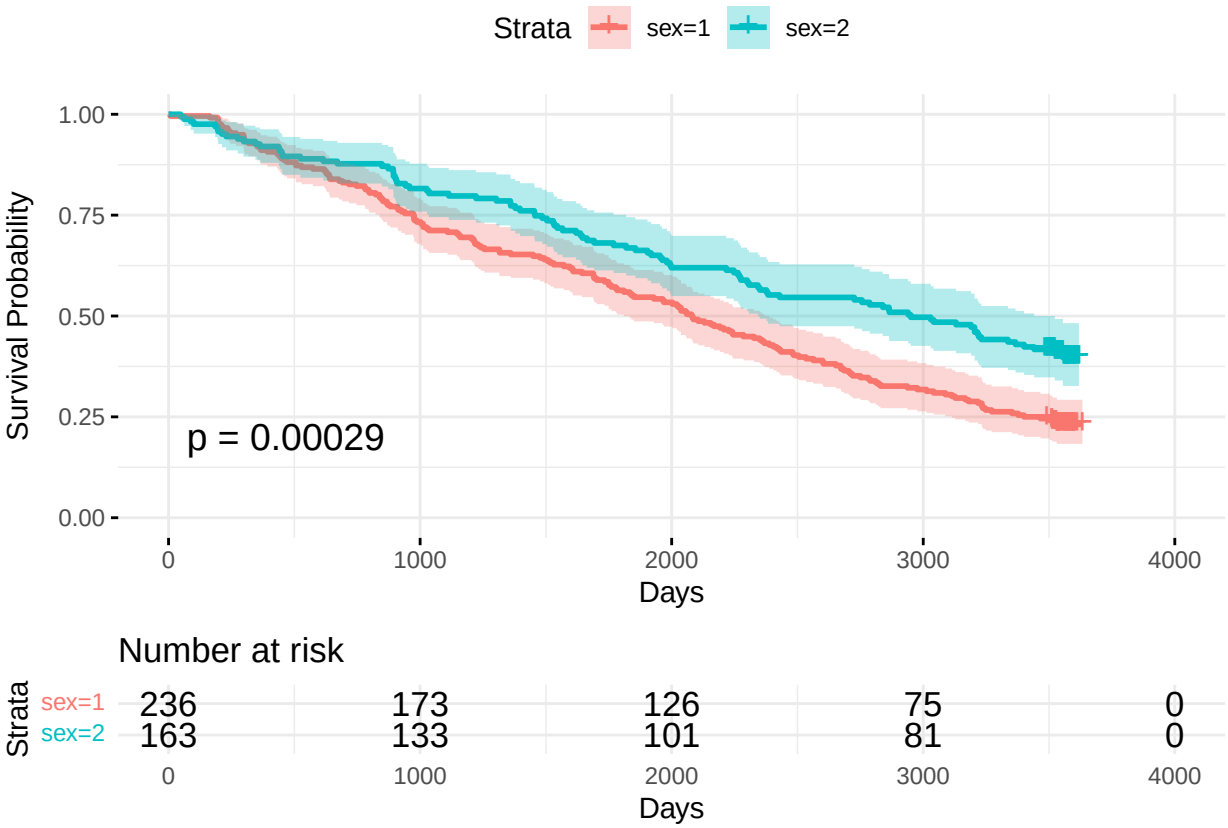
2023-12-06

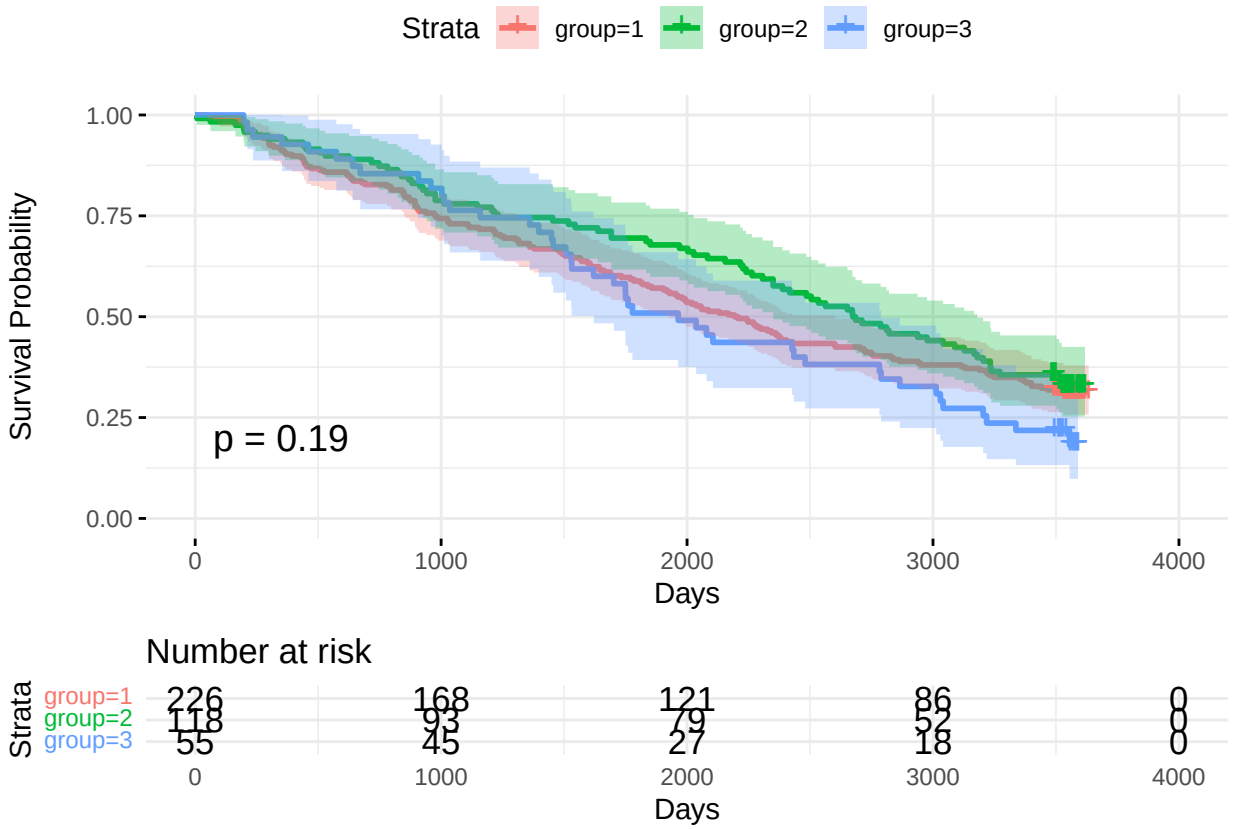
```
## Call:
## coxph(formula = surv_obj ~ ., data = brv)
##
##   n= 399, number of events= 278
##
##              coef  exp(coef)   se(coef)        z Pr(>|z|)
## id      2.802e-06  1.000e+00  1.184e-05    0.237    0.813
## couple -2.653e-03  9.974e-01  1.645e-03   -1.613    0.107
## dob     4.051e-05  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
```

```
##        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
```

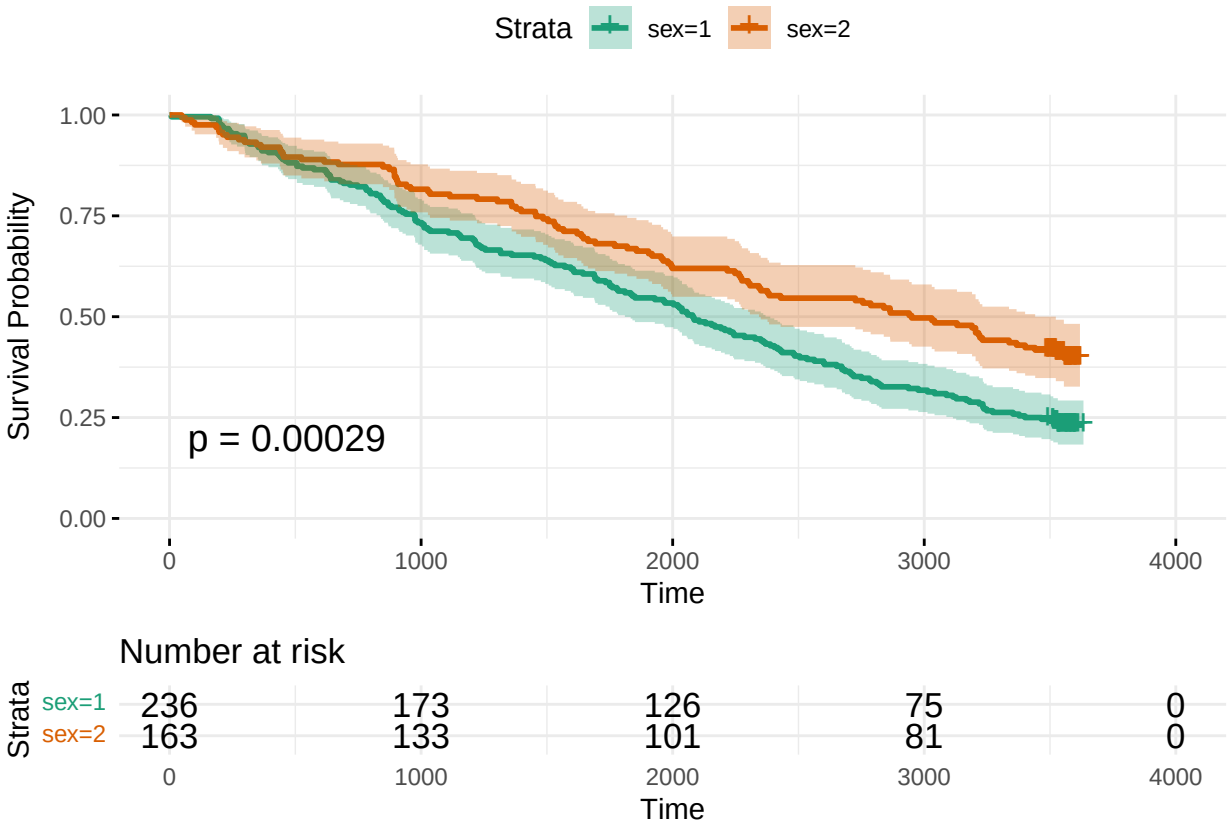

Strata
group=1
group=2
group=3
1.00
0.75
0.50
0.25
0.00
Survival probability
p = 0.19
0
1000
2000
3000
4000
Time


Strata
sex=1
sex=2
1.00
0.75
0.50
0.25
0.00
Survival Probability
p = 0.00029
0
1000
2000
3000
4000
Days
Number at risk
Strata
sex=1
sex=2
236
173
126
75
0
163
133
101
81
0
0
1000
2000
3000
4000
Days




```
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")
```




```
## 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
```

```r
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
```

```r
# 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
```

```r
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
```

```r
# 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 Estimation:

$S(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$

Where $S(t)$ is the survival probability at time t_i, d_i is the number of events at time t_i, and n_i is the number of subjects at risk at time t_i.

Cox Proportional Hazards Model:

$h(t) = h_0(t) \exp(\beta_1 \cdot \text{pspline(age)} + \beta_2 \cdot \text{size} + \beta_3 \cdot \text{grade} + \beta_4 \cdot \text{nodes} + \beta_5 \cdot \text{pgr} + \beta_6 \cdot \text{er} + \beta_7 \cdot \text{hormon} + \beta_8 \cdot \text{chemo})$

Where $h(t)$ is the hazard at time t_i, $h_0(t)$ is the baseline hazard, $\beta_1, \beta_2, ..., \beta_8$ are the coefficients for each covariate, which include age modeled with a penalized spline, tumor size, grade, number of positive lymph nodes, progesterone receptor levels, estrogen receptor levels, hormonal treatment, and chemotherapy, respectively.