01-26-26-coxph_Rscript.R

sgrambow

2026-01-22

```
# ============================================================
# Time-to-Event Analysis of CGD Clinical Trial Data
# Documentation enhanced by Claude 3.5 Sonnet on 2025-01-30
# ============================================================
```

STUDY BACKGROUND This analysis examines data from a clinical trial investigating the efficacy of gamma interferon in preventing serious infections among patients with chronic granulomatous disease (CGD). The primary endpoint is time to first serious infection from study entry.

QUESTIONS TO BE ANSWERED: Q1: What is the hazard ratio and 95% CI comparing placebo vs. gamma interferon? Q2: Is there an association between sex and infection risk? Q3: Create and interpret Kaplan-Meier plots stratified by sex Q4: Is there an association between age and infection risk? Q5: What is the hazard ratio for a 5-year increase in age?

```
# Load required packages
if (!require(survival)) install.packages("survival")
```

```
## Loading required package: survival
```

```
if (!require(survminer)) install.packages("survminer")
```

```
## Loading required package: survminer
```

```
## Loading required package: ggplot2
```

```
## Loading required package: ggpubr
```

```
##
## Attaching package: 'survminer'
```

```
## The following object is masked from 'package:survival':
##
##     myeloma
```

```r
library(survival)   # For survival analysis functions
library(survminer)  # For creating Kaplan-Meier plots

# Load data
load(url("https://www.duke.edu/~sgrambow/crp241data/cgd.RData"))

# ============================================================
# INITIAL DATA EXPLORATION
# ============================================================

# Examine data structure and descriptive statistics
str(cgd)
```

```
## 'data.frame':    128 obs. of  10 variables:
##  $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ treat   : num  1 0 1 1 0 1 0 1 0 1 ...
##  $ sex     : num  0 1 1 1 1 0 1 1 1 1 ...
##  $ age     : int  12 15 19 12 17 44 22 7 27 5 ...
##  $ height  : num  147 159 171 142 162 ...
##  $ weight  : num  62 47.5 72.7 34 52.7 45 59.7 17.4 82.8 19.5 ...
##  $ inherit : num  0 0 1 1 1 0 1 1 0 1 ...
##  $ propylac: int  0 1 1 1 1 0 1 1 1 1 ...
##  $ tstop   : int  219 8 382 388 246 364 292 363 294 371 ...
##  $ status  : int  1 1 0 0 1 0 1 0 1 0 ...
```

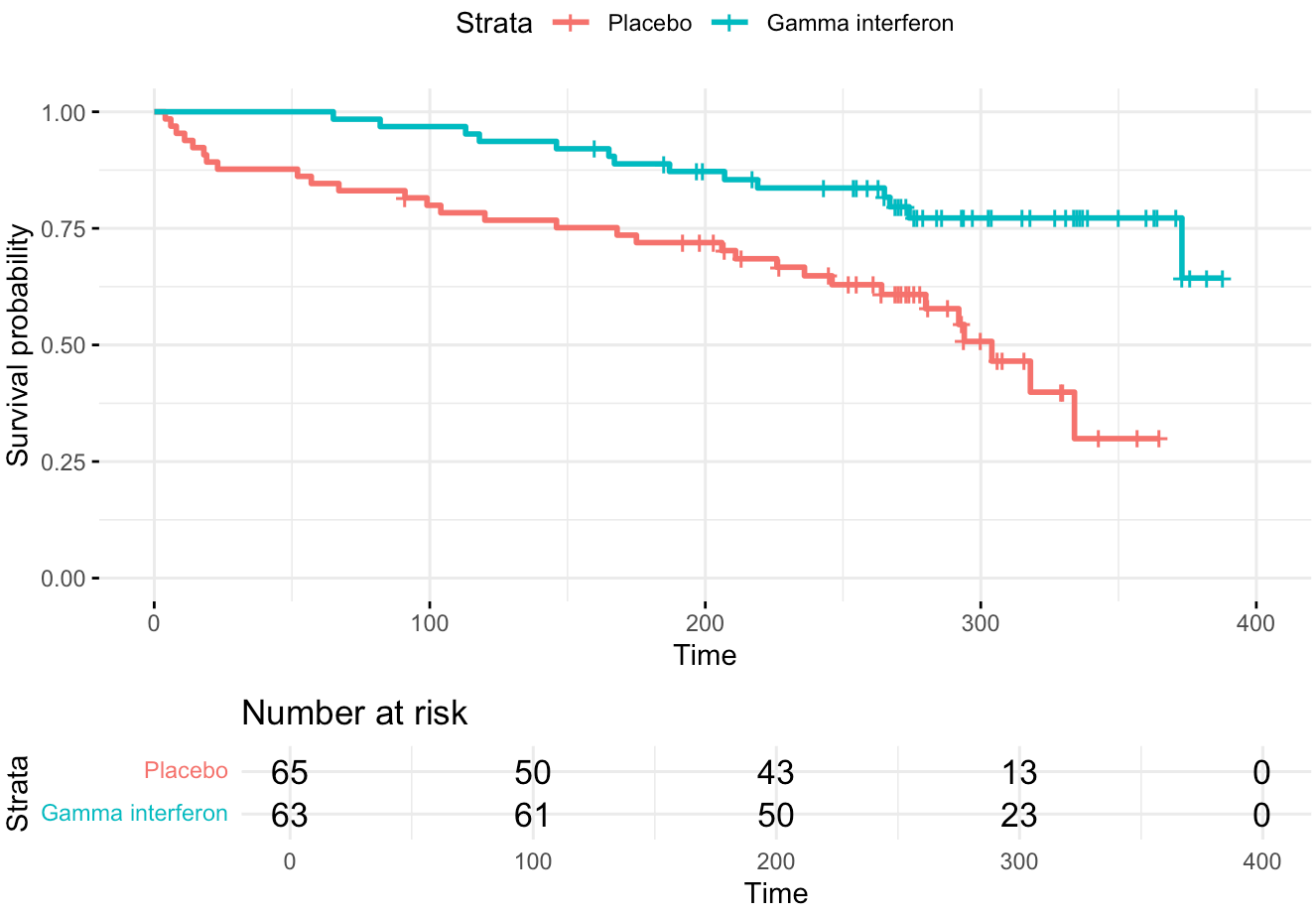
```r
summary(cgd)
```

```
##        id             treat             sex              age       
##  Min.   :  1.00   Min.   :0.0000   Min.   :0.0000   Min.   : 1.00  
##  1st Qu.: 32.75   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.: 7.00  
##  Median : 64.50   Median :0.0000   Median :1.0000   Median :12.00  
##  Mean   : 64.81   Mean   :0.4922   Mean   :0.8125   Mean   :14.64  
##  3rd Qu.: 96.25   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:22.00  
##  Max.   :135.00   Max.   :1.0000   Max.   :1.0000   Max.   :44.00  
##      height          weight          inherit          propylac     
##  Min.   : 76.3   Min.   : 10.40   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:116.5   1st Qu.: 20.68   1st Qu.:0.0000   1st Qu.:1.0000  
##  Median :140.8   Median : 34.85   Median :1.0000   Median :1.0000  
##  Mean   :140.1   Mean   : 40.56   Mean   :0.6719   Mean   :0.8672  
##  3rd Qu.:169.7   3rd Qu.: 59.17   3rd Qu.:1.0000   3rd Qu.:1.0000  
##  Max.   :189.0   Max.   :101.50   Max.   :1.0000   Max.   :1.0000  
##      tstop           status      
##  Min.   :  4.0   Min.   :0.0000  
##  1st Qu.:197.0   1st Qu.:0.0000  
##  Median :269.0   Median :0.0000  
##  Mean   :241.1   Mean   :0.3438  
##  3rd Qu.:304.5   3rd Qu.:1.0000  
##  Max.   :388.0   Max.   :1.0000
```

Key Findings from Initial Data Review: - Sample size: 128 patients - Treatment groups: 49.2% gamma interferon, 50.8% placebo - Sex distribution: 81.3% male, 18.7% female - Age: median 12 years (range: 1-44 years) - Events: 44 infections observed (34.4% event rate) - Follow-up time: median 269 days (range: 4-388 days)

```
# Create initial Kaplan-Meier curves stratified by treatment
fit.km <- survfit(Surv(tstop, status) ~ treat, data=cgd)

# Create visualization with risk table
ggsurvplot(fit.km, data=cgd,
           ggtheme = theme_minimal(),
           legend.labs = c("Placebo","Gamma interferon"),
           risk.table = TRUE)
```



Initial Visual Assessment: - Clear separation between treatment arms - Gamma interferon group shows better survival - Effect appears early and maintains throughout follow-up - Good retention in both arms based on risk table

```
# ============================================================
# QUESTION 1: Treatment Effect Analysis
# What is the hazard ratio and 95% CI comparing placebo vs. gamma interferon?
# ============================================================

# Fit Cox model comparing treatments
# Note: Model estimates gamma (treat=1) vs placebo (treat=0)
mfit <- coxph(Surv(tstop, status) ~ treat, data=cgd)
summary(mfit)
```

```
## Call:
## coxph(formula = Surv(tstop, status) ~ treat, data = cgd)
##
##   n= 128, number of events= 44
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## treat -1.0940    0.3349   0.3348 -3.268  0.00108 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##       exp(coef) exp(-coef) lower .95 upper .95
## treat    0.3349      2.986    0.1737    0.6454
##
## Concordance= 0.621  (se = 0.036 )
## Likelihood ratio test= 11.8  on 1 df,   p=6e-04
## Wald test            = 10.68  on 1 df,   p=0.001
## Score (logrank) test = 11.74  on 1 df,   p=6e-04
```

```
# Calculate hazard ratio for placebo vs gamma
1/.3349  # HR = 2.99
```

```
## [1] 2.985966
```

```
# Calculate confidence intervals
1/.6454  # Lower CI = 1.55
```

```
## [1] 1.549427
```

```
1/.1737  # Upper CI = 5.76
```

```
## [1] 5.757052
```

INTERPRETATION OF TREATMENT EFFECT: Model Results: - Log hazard coefficient = -1.0940 (SE: 0.3348) - Hazard Ratio (Placebo vs Gamma): 2.99 [95% CI: 1.55, 5.76] - P-value = 0.001 (statistically significant) - Concordance = 0.621 (SE: 0.036)

Clinical Interpretation: - Patients on placebo had approximately 3 times the risk of infection compared to those on gamma interferon - We are 95% confident the true hazard ratio is between 1.55 and 5.76 - The effect is both statistically and clinically significant

```
# ============================================================
# QUESTION 2: Sex Effect Analysis
# Is there an association between sex and infection risk?
# ============================================================

# Fit Cox model examining sex effect
mfit2.cgd <- coxph(Surv(tstop, status) ~ sex, data=cgd)
summary(mfit2.cgd)
```

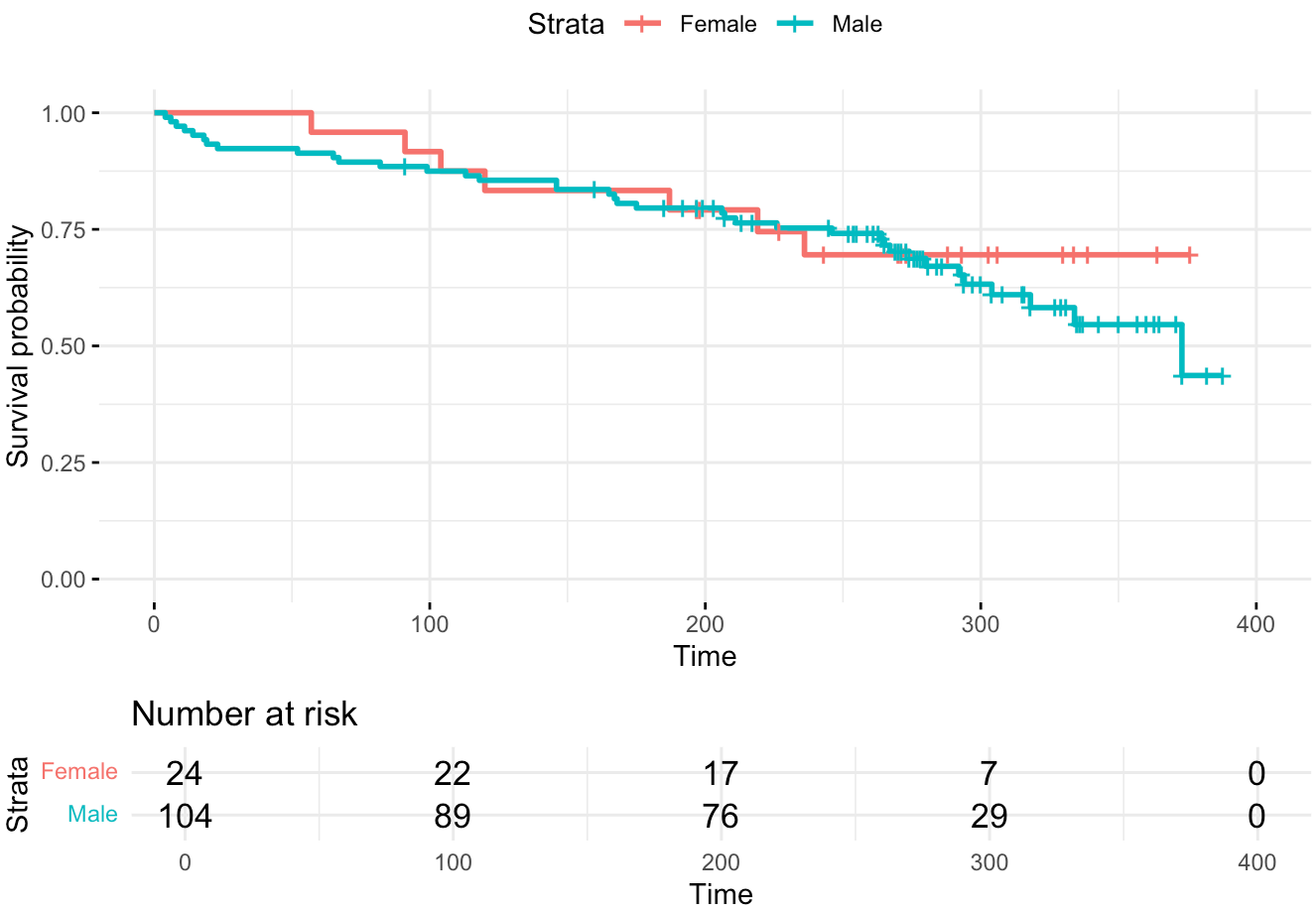
```
## Call:
## coxph(formula = Surv(tstop, status) ~ sex, data = cgd)
##
##   n= 128, number of events= 44
##
##       coef exp(coef) se(coef)     z Pr(>|z|)
## sex 0.2127    1.2370   0.4123 0.516    0.606
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex     1.237     0.8084    0.5514     2.775
##
## Concordance= 0.507  (se = 0.031 )
## Likelihood ratio test= 0.28  on 1 df,   p=0.6
## Wald test            = 0.27  on 1 df,   p=0.6
## Score (logrank) test = 0.27  on 1 df,   p=0.6
```

INTERPRETATION OF SEX EFFECT: Model Results: - Log hazard coefficient = 0.2127 (SE: 0.4123) - Hazard Ratio (Male vs Female): 1.24 [95% CI: 0.55, 2.78] - P-value = 0.606 (not statistically significant) - Concordance = 0.507 (SE: 0.031)

Clinical Interpretation: - Males had a 24% higher risk of infection compared to females - However, this difference was not statistically significant - The wide confidence interval suggests considerable uncertainty - Cannot conclude there is a true difference in risk between sexes

```
# ============================================================
# QUESTION 3: Visualizing Sex Effect
# Create and interpret Kaplan-Meier plots stratified by sex
# ============================================================

# Create Kaplan-Meier plot stratified by sex
mfit.sex <- survfit(Surv(tstop, status) ~ sex, data=cgd)
ggsurvplot(mfit.sex, data=cgd,
           ggtheme=theme_minimal(),
           legend.labs = c("Female","Male"),
           risk.table = TRUE)
```



INTERPRETATION OF KM CURVES: Visual Assessment: - The survival curves for males and females overlap considerably - No clear pattern of sex difference in infection risk - Risk table shows adequate initial numbers but female group small - Results align with Cox model showing no significant sex effect

```
# ============================================================
# QUESTION 4: Age Effect Analysis
# Is there an association between age and infection risk?
# ============================================================

# Fit Cox model examining age effect
mfit.age <- coxph(Surv(tstop, status) ~ age, data=cgd)
summary(mfit.age)
```

```
## Call:
## coxph(formula = Surv(tstop, status) ~ age, data = cgd)
##
##   n= 128, number of events= 44
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## age -0.02121   0.97901  0.01682 -1.261    0.207
##
##     exp(coef) exp(-coef) lower .95 upper .95
## age     0.979      1.021    0.9473     1.012
##
## Concordance= 0.57  (se = 0.048 )
## Likelihood ratio test= 1.69  on 1 df,   p=0.2
## Wald test            = 1.59  on 1 df,   p=0.2
## Score (logrank) test = 1.6  on 1 df,   p=0.2
```

INTERPRETATION OF AGE EFFECT: Model Results: - Log hazard coefficient = -0.02121 (SE: 0.01682) - Hazard Ratio (per 1-year increase): 0.979 [95% CI: 0.947, 1.012] - P-value = 0.207 (not statistically significant) - Concordance = 0.57 (SE: 0.048)

Clinical Interpretation: - For each 1-year increase in age, the hazard of infection decreases by 2.1% - However, this trend is not statistically significant - Cannot conclude there is a true association between age and infection risk

```
# ============================================================
# QUESTION 5: Age Effect Over 5 Years
# What is the hazard ratio for a 5-year increase in age?
# ============================================================

# Calculate 5-year age effect
exp(5*-0.02121)  # Multiply coefficient by 5 years then exp()
```

```
## [1] 0.8993797
```

```
# Calculate CI by multiplying original confidence limits by 5
exp(5*confint(mfit.age))  # Calculate CI for 5-year effect
```

```
##         2.5 % 97.5 %
## age 0.7627288 1.0605
```

INTERPRETATION OF 5-YEAR AGE EFFECT: Calculations: - Log hazard coefficient for 5 years = 5 * (-0.02121) = -0.10605

Results: - Hazard Ratio (per 5-year increase): 0.90 [95% CI: 0.76, 1.06]

Clinical Interpretation: - For every 5-year increase in age, the hazard of infection decreases by approximately 10% - However, this remains statistically non-significant - The confidence interval crosses 1, suggesting we cannot rule out no effect or even a small harmful effect of increasing age

```
# End of analysis
```