# BS852 Final Project

**Kaden Bailey**

**2025-12-07**

# Introduction

Longevity research hopes to identify biological and behavioral factors that influence survival in aging populations. Inflammation is a well-known risk factor for age-related disease and death. Elevated levels of inflammatory biomarkers, such as interleukin-6 (IL-6) and high-sensitivity C-reactive protein (hsCRP), have been associated with cardiovascular disease, and premature death.

The aim of this project is to evaluate the association between systemic inflammation and mortality in a cohort of adults followed for approximately eight years. Specifically, we test whether higher concentrations of log-transformed IL-6 are associated with increased risk of death, while accounting for demographic, physiological, and cognitive covariates. IL-6 was selected as the primary inflammatory biomarker based on its role as a proximal pro-inflammatory cytokine with established mechanistic links to age related mortality, including effects on cardiovascular dysfunction, and immune dysregulation. While hsCRP was considered as a secondary marker, IL-6 was prioritized given its upstream position in the inflammatory response and its more direct involvement in aging.

# Methods

Participants were drawn from a longitudinal study of aging and longevity. Eligible subjects were adults older than 25 years at enrollment who underwent detailed phenotyping. Follow-up extended for approximately eight years, with vital status collected at last contact in November 2021. Age at enrollment and age at last contact were recorded for each participant. The primary outcome was all cause mortality. Survival time was calculated as the difference between age at last contact and age at enrollment. Vital status was coded as an event indicator, with death defined as event = 1 and alive at last contact defined as event = 0.

The primary exposure of interest was systemic inflammation, measured by the log-transformed concentration of interleukin-6 (IL-6). High sensitivity C-reactive protein (hsCRP), also log-transformed, was considered as a secondary biomarker. Potential confounders included demographic, physiological, and cognitive measures: sex, age at enrollment, years of education, body mass index (z-score), systolic blood pressure (z-score), and cognitive performance assessed by the Digit Symbol Substitution Test (DSST). Additional functional measures (grip strength, gait speed, lung function) were summarized descriptively but not included in the primary adjusted model.

Baseline characteristics were summarized using means and standard deviations for continuous variables and proportions for categorical variables, stratified by survival status. Survival analyses were conducted using Cox proportional hazards regression. Crude models included only the primary predictor (log-IL6). Adjusted models incorporated covariates to account for potential confounding. Hazard ratios (HRs) with 95% confidence intervals (CIs) were reported. Complete case analysis was performed; participants with missing values in any model covariate were excluded from the adjusted analysis (n=381, 8.3% of the sample). Kaplan-Meier survival curves were generated to visualize differences in survival between participants with high versus low IL-6 concentrations, defined by a median split. The log-rank test was used to compare survival distributions. Forest plots were produced to display hazard ratios from adjusted Cox models. All analyses were performed using R (version 4.5.1) with the survival, survminer, and tableone packages. Statistical significance was defined as a two-sided p-value $< 0.05$.

# Results

## Baseline Characteristics

A total of 4,581 participants were included in the analysis. At last contact, 1,260 individuals had died and 3,315 were alive. Table 1 summarizes baseline characteristics stratified by survival status. Compared with survivors, those who died were older at enrollment (mean age 89.1 vs. 64.2 years), had fewer years of education (mean 9.9 vs. 12.2), and lower cognitive performance (mean DSST score 27.1 vs. 48.9). Mean log-IL6 concentrations were higher among deceased participants (0.85 vs. −0.15), as were hsCRP levels (0.85 vs. 0.34).

## Association Between IL-6 and Mortality

In crude Cox regression, higher log-IL6 was strongly associated with increased risk of death (HR = 2.02, 95% CI: 1.94–2.10, $p < 0.001$). After adjustment for sex, age at enrollment, education, BMI, systolic blood pressure, and DSST score, the association remained significant though substantially weaker (HR = 1.30, 95% CI: 1.22–1.39, $p < 0.001$).

Several covariates were independently associated with mortality. Older age at enrollment was the strongest predictor (HR = 1.09 per year, 95% CI: 1.09–1.10, $p < 0.001$), indicating a 9% increase in mortality risk per additional year of age. Lower DSST scores strongly predicted higher risk of death (HR = 0.97 per unit, 95% CI: 0.97–0.98, $p < 0.001$). Male sex was associated with increased mortality (HR for female = 0.70, 95% CI: 0.62–0.79, $p < 0.001$). Higher BMI was protective (HR = 0.89 per z-score unit, 95% CI: 0.84–0.95, $p < 0.001$). Education and systolic blood pressure were not statistically significant predictors in the adjusted model.

## Survival Curves and Forest Plot

Kaplan-Meier curves (Figure 1) demonstrated significantly lower survival among participants with high IL-6 concentrations compared to those with low IL-6 (log-rank $p < 0.001$). The forest plot of adjusted Cox regression results (Figure 2) visually confirmed the independent contribution of IL-6 alongside demographic and clinical covariates.

# Discussion and Conclusion

Higher concentrations of log-transformed IL-6 were significantly associated with increased mortality risk, even after adjustment for demographic, physiological, and cognitive covariates. The substantial reduction from crude (HR = 2.02) to adjusted models (HR = 1.30) suggests that much of IL-6's effect, operates through or is confounded by age, sex, cognitive function, and body composition, however the association remained robust. Kaplan-Meier curves demonstrated poorer survival among individuals with elevated IL-6, highlighting its potential as a prognostic biomarker. These findings highlight inflammation as a critical determinant of survival and suggest that interventions targeting inflammatory pathways may improve health outcomes in older adults.

# Important plots and Figures

**Table 1: Baseline Characteristics Stratified by Survival Status**

| Variable | Deceased (n=1,260) | Alive (n=3,315) |
|---|---|---|
| Sex (% Male) | 51% | 43% |
| Age at enrollment (years) | 89.1 (10.1) | 64.2 (11.6) |
| Education (years) | 9.9 (4.1) | 12.2 (3.2) |
| DSST score | 27.1 (13.6) | 48.9 (13.3) |
| BMI (z-score) | -0.01 (1.03) | 0.02 (0.98) |
| Systolic BP (z-score) | 0.01 (1.02) | 0.03 (0.99) |
| Log-IL6 | 0.85 (0.95) | -0.15 (0.85) |
| Log-hsCRP | 0.85 (1.19) | 0.34 (1.03) |

*Values are mean (SD) or %.*

**Figure 1**: Kaplan-Meier survival curves by IL-6 level (log-rank p < 0.001).

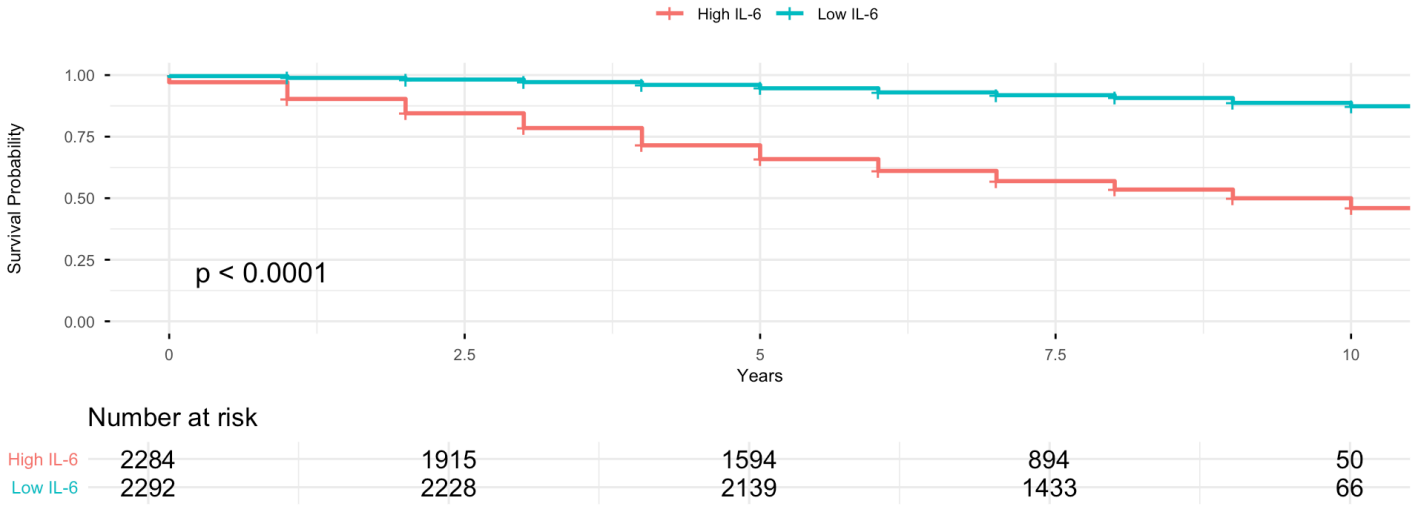
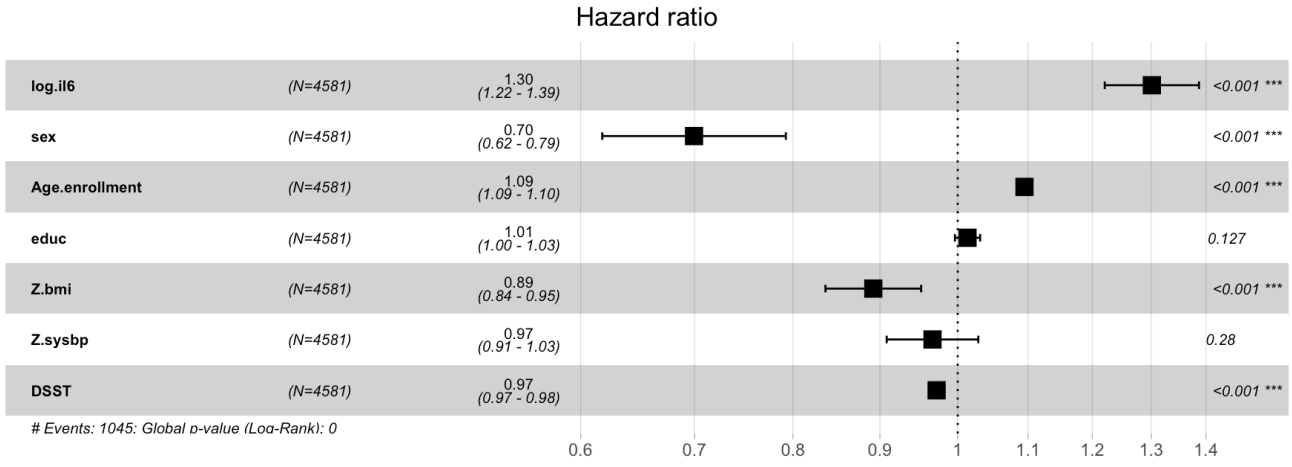



**Figure 2**: Forest plot of adjusted hazard ratios.



# Appendix

```
data <- read.csv("project.2025.csv")

str(data)
```

```
## 'data.frame':    4581 obs. of  13 variables:
##  $ educ            : int  14 14 12 15 14 10 13 14 14 14 ...
##  $ DSST            : int  68 67 60 61 54 35 74 73 65 64 ...
##  $ Alive           : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Age.last.contact: int  27 40 44 44 45 45 38 44 47 46 ...
##  $ Age.enrollment  : int  25 32 36 36 37 37 37 37 38 38 ...
##  $ sex             : int  2 1 2 2 1 2 2 1 2 2 ...
##  $ log.il6         : num  0.14 -1.273 -1.05 -0.329 -1.609 ...
##  $ log.new.hscrp   : num  1.5412 -0.0726 0.793 1.4207 2.4397 ...
##  $ Z.grip.strength : num  -1.1 -0.287 -1.424 -0.857 -1.303 ...
##  $ Z.bmi           : num  -0.6252 -0.5189 -1.0691 -0.0703 1.1555 ...
##  $ Z.gait.speed    : num  -1.164 -0.592 -0.335 1.13 -1.175 ...
##  $ Z.fev1.7        : num  NA 0.963 -0.147 -0.396 -0.125 ...
##  $ Z.sysbp         : num  -0.00271 0.29306 -0.73983 1.76637 -0.59087 ...
```

```
summary(data)
```

```
##       educ           DSST          Alive           Age.last.contact
##  Min.   : 0.0   Min.   : 0.00   Length:4581        Min.   : 27.00
##  1st Qu.: 9.0   1st Qu.:32.00   Class :character   1st Qu.: 66.00
##  Median :13.0   Median :44.00   Mode  :character   Median : 74.50
##  Mean   :11.6   Mean   :43.45                      Mean   : 77.75
##  3rd Qu.:14.0   3rd Qu.:55.00                      3rd Qu.: 92.00
##  Max.   :17.0   Max.   :93.00                      Max.   :111.00
##  NA's   :8      NA's   :230                        NA's   :5
##  Age.enrollment        sex           log.il6        log.new.hscrp
##  Min.   : 25.00   Min.   :1.000   Min.   :-2.5257   Min.   :-3.1869
##  1st Qu.: 59.00   1st Qu.:1.000   1st Qu.:-0.5621   1st Qu.:-0.2614
##  Median : 67.00   Median :2.000   Median : 0.0000   Median : 0.3850
##  Mean   : 71.04   Mean   :1.547   Mean   : 0.1282   Mean   : 0.4818
##  3rd Qu.: 87.00   3rd Qu.:2.000   3rd Qu.: 0.6831   3rd Qu.: 1.1442
##  Max.   :110.00   Max.   :2.000   Max.   : 4.8828   Max.   : 5.3124
##
##  Z.grip.strength         Z.bmi           Z.gait.speed          Z.fev1.7
##  Min.   :-3.736542   Min.   :-3.03003   Min.   :-4.444448   Min.   :-4.67987
##  1st Qu.:-0.669820   1st Qu.:-0.70679   1st Qu.:-0.672063   1st Qu.:-0.62165
##  Median : 0.002763   Median :-0.10577   Median :-0.002695   Median : 0.00491
##  Mean   : 0.007183   Mean   : 0.01405   Mean   : 0.006469   Mean   :-0.00036
##  3rd Qu.: 0.656129   3rd Qu.: 0.65137   3rd Qu.: 0.645025   3rd Qu.: 0.66550
##  Max.   : 6.709879   Max.   : 3.95051   Max.   : 4.579952   Max.   : 6.91806
##  NA's   :40          NA's   :173        NA's   :117         NA's   :511
##     Z.sysbp
##  Min.   :-2.95117
##  1st Qu.:-0.69356
##  Median :-0.09511
##  Mean   : 0.02250
##  3rd Qu.: 0.60338
##  Max.   : 5.36219
##
```

```
data$event <- ifelse(data$Alive == "Yes", 0, 1)
data$time <- data$Age.last.contact - data$Age.enrollment
surv_obj <- Surv(time = data$time, event = data$event)

main_predictor <- "log.il6"

covariates <- c("sex", "Age.enrollment", "educ", "Z.bmi", "Z.sysbp", "DSST")

#Crude
cox_crude <- coxph(as.formula(paste("surv_obj ~", main_predictor)), data = data)
summary(cox_crude)
```

```
## Call:
## coxph(formula = as.formula(paste("surv_obj ~", main_predictor)),
##     data = data)
##
##   n= 4576, number of events= 1261
##    (5 observations deleted due to missingness)
##
##            coef exp(coef) se(coef)     z Pr(>|z|)
## log.il6 0.70205   2.01789  0.02092 33.56   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## log.il6     2.018     0.4956     1.937     2.102
##
## Concordance= 0.774  (se = 0.006 )
## Likelihood ratio test= 864.1  on 1 df,   p=<2e-16
## Wald test            = 1126  on 1 df,   p=<2e-16
## Score (logrank) test = 1137  on 1 df,   p=<2e-16
```

```
# Adjusted
cox_adj <- coxph(as.formula(
  paste("surv_obj ~", main_predictor, "+", paste(covariates, collapse = "+"))
), data = data)
summary(cox_adj)
```

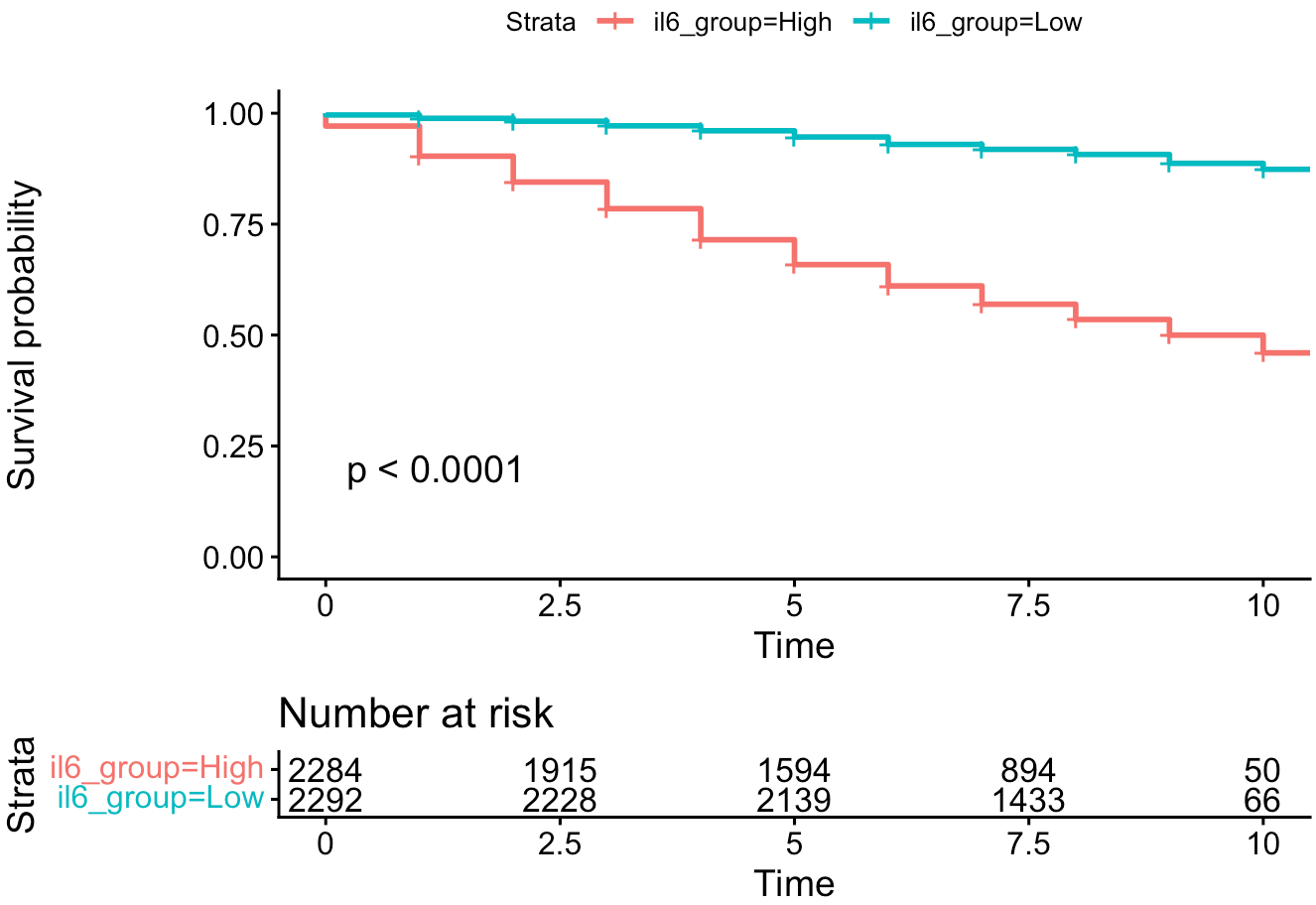
```
## Call:
## coxph(formula = as.formula(paste("surv_obj ~", main_predictor,
##     "+", paste(covariates, collapse = "+"))), data = data)
##
##   n= 4200, number of events= 1045
##    (381 observations deleted due to missingness)
##
##                     coef exp(coef)  se(coef)      z Pr(>|z|)
## log.il6         0.263075  1.300925  0.032586  8.073 6.85e-16 ***
## sex            -0.357242  0.699603  0.063469 -5.629 1.82e-08 ***
## Age.enrollment  0.090491  1.094711  0.003785 23.910  < 2e-16 ***
## educ            0.013294  1.013383  0.008706  1.527 0.126773
## Z.bmi          -0.114347  0.891948  0.033088 -3.456 0.000549 ***
## Z.sysbp        -0.034169  0.966409  0.031653 -1.079 0.280377
## DSST           -0.028572  0.971832  0.002868 -9.963  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## log.il6           1.3009     0.7687    1.2204    1.3867
## sex               0.6996     1.4294    0.6178    0.7923
## Age.enrollment    1.0947     0.9135    1.0866    1.1029
## educ              1.0134     0.9868    0.9962    1.0308
## Z.bmi             0.8919     1.1211    0.8359    0.9517
## Z.sysbp           0.9664     1.0348    0.9083    1.0283
## DSST              0.9718     1.0290    0.9664    0.9773
##
## Concordance= 0.891  (se = 0.005 )
## Likelihood ratio test= 2506  on 7 df,   p=<2e-16
## Wald test            = 1678  on 7 df,   p=<2e-16
## Score (logrank) test = 2642  on 7 df,   p=<2e-16
```

```
vars <- c("sex", "Age.enrollment", "educ", "DSST", "Z.bmi", "Z.sysbp", "log.il6", "log.n
ew.hscrp")
table1 <- CreateTableOne(vars = vars, strata = "Alive", data = data, test = FALSE)
print(table1, showAllLevels = TRUE)
```

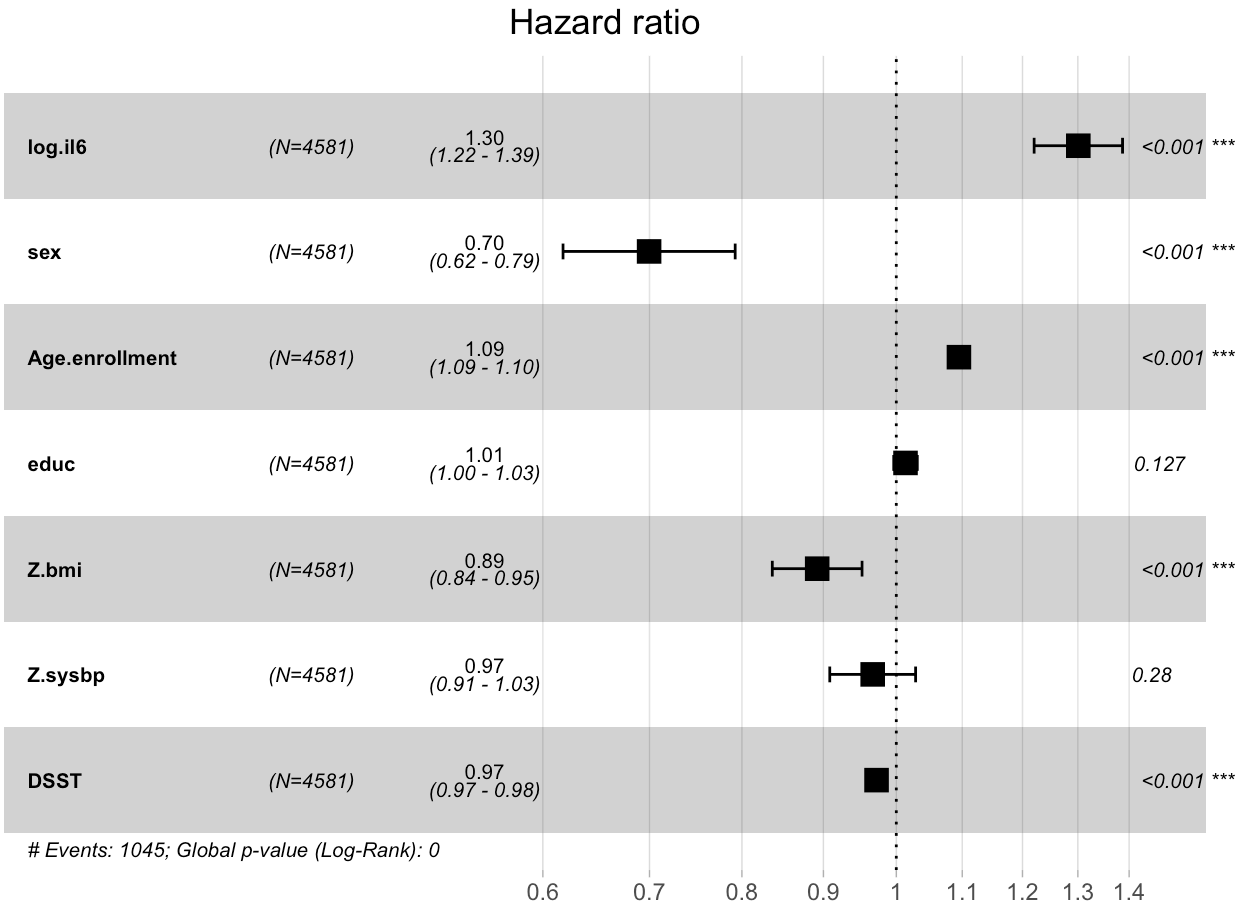
```
##                             Stratified by Alive
##                              level               No            Yes
##   n                                     6          1260          3315
##   sex (mean (SD))                    1.67 (0.52)   1.49 (0.50)   1.57 (0.50)
##   Age.enrollment (mean (SD))        63.33 (10.65) 89.08 (10.14) 64.19 (11.63)
##   educ (mean (SD))                  13.17 (2.04)   9.88 (4.07)  12.24 (3.19)
##   DSST (mean (SD))                  49.00 (16.54) 27.11 (13.59) 48.92 (13.31)
##   Z.bmi (mean (SD))                  0.65 (0.70)  -0.01 (1.03)   0.02 (0.98)
##   Z.sysbp (mean (SD))                0.79 (1.09)   0.01 (1.02)   0.03 (0.99)
##   log.il6 (mean (SD))               -0.29 (0.56)   0.85 (0.95)  -0.15 (0.85)
##   log.new.hscrp (mean (SD))          0.52 (0.66)   0.85 (1.19)   0.34 (1.03)
```

```
# Kaplan-Meier
data$il6_group <- ifelse(data$log.il6 > median(data$log.il6, na.rm = TRUE), "High", "Lo
w")
fit <- survfit(surv_obj ~ il6_group, data = data)
ggsurvplot(fit, data = data, pval = TRUE, risk.table = TRUE)
```

```
## Ignoring unknown labels:
## • colour : "Strata"
```




```
# Forest plot
ggforest(cox_adj, data = data)
```


Hazard ratio
log.il6 (N=4581) 1.30 (1.22 - 1.39) <0.001 ***
sex (N=4581) 0.70 (0.62 - 0.79) <0.001 ***
Age.enrollment (N=4581) 1.09 (1.09 - 1.10) <0.001 ***
educ (N=4581) 1.01 (1.00 - 1.03) 0.127
Z.bmi (N=4581) 0.89 (0.84 - 0.95) <0.001 ***
Z.sysbp (N=4581) 0.97 (0.91 - 1.03) 0.28
DSST (N=4581) 0.97 (0.97 - 0.98) <0.001 ***
# Events: 1045; Global p-value (Log-Rank): 0
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4