Build Survival Model: Cox Proportional Hazards Model

Mingcheng Hu

Table of contents


```r
library(tidyverse)
library(survival)
library(forestplot)
library(glmnet)
library(ggfortify)
library(kableExtra) # include knitr automatically

source("/work/users/y/u/yuukias/BIOS-Material/BIOS992/utils/csv_utils.r")
# * Don't use setwd() for Quarto documents!
# setwd("/work/users/y/u/yuukias/BIOS-Material/BIOS992/data")

adjust_type <- ifelse(exists("params"), params$adjust_type, "partial") #
↪  options: "minimal", "partial", "full"
```

```r
impute_type <- ifelse(exists("params"), params$impute_type, "unimputed") #
↪  options: "unimputed", "imputed"
include_statin <- ifelse(exists("params"), params$include_statin, "no") #
↪  options: "yes", "no"
```

```r
# string of parameters
adjust_type_str <- switch(adjust_type,
    minimal = "minimal",
    partial = "partial",
    full = "full"
)
print(paste0("Model Adjustment Type: ", adjust_type_str))
```

```
[1] "Model Adjustment Type: partial"
```

```r
impute_type_str <- switch(impute_type,
    unimputed = "unimputed",
    imputed = "imputed"
)
print(paste0("Data Imputation Type: ", impute_type_str))
```

```
[1] "Data Imputation Type: unimputed"
```

Load Data

```r
if (include_statin == "yes") {
    data_train <-
↪  read.csv(paste0("/work/users/y/u/yuukias/BIOS-Material/BIOS992/data/train_data_",
↪  impute_type_str, "_statin.csv"),
        header = TRUE
    )
} else {
    data_train <-
↪  read.csv(paste0("/work/users/y/u/yuukias/BIOS-Material/BIOS992/data/train_data_",
↪  impute_type_str, ".csv"),
        header = TRUE
    )
```

```r
}

data_train <- data_train[, -1] # the first column is the index generated by
↪ sklearn
(dim(data_train))
```

```
[1] 28127   100
```

```r
data <- select_subset(data_train, type = adjust_type)
(dim(data))
```

```
[1] 28127    75
```

```r
colnames(data)
```

```
 [1] "event"                          "time"
 [3] "HRV_MeanNN"                     "HRV_SDNN"
 [5] "HRV_RMSSD"                      "HRV_SDSD"
 [7] "HRV_CVNN"                       "HRV_CVSD"
 [9] "HRV_MedianNN"                   "HRV_MadNN"
[11] "HRV_MCVNN"                      "HRV_IQRNN"
[13] "HRV_SDRMSSD"                    "HRV_Prc20NN"
[15] "HRV_Prc80NN"                    "HRV_pNN50"
[17] "HRV_pNN20"                      "HRV_MinNN"
[19] "HRV_MaxNN"                      "HRV_HTI"
[21] "HRV_TINN"                       "HRV_LF"
[23] "HRV_HF"                         "HRV_VHF"
[25] "HRV_TP"                         "HRV_LFHF"
[27] "HRV_LFn"                        "HRV_HFn"
[29] "HRV_LnHF"                       "HRV_SD1"
[31] "HRV_SD2"                        "HRV_SD1SD2"
[33] "HRV_S"                          "HRV_CSI"
[35] "HRV_CVI"                        "HRV_CSI_Modified"
[37] "HRV_PIP"                        "HRV_IALS"
[39] "HRV_PSS"                        "HRV_PAS"
[41] "HRV_GI"                         "HRV_SI"
[43] "HRV_AI"                         "HRV_PI"
[45] "HRV_C1d"                        "HRV_C1a"
[47] "HRV_SD1d"                       "HRV_SD1a"
```

```
[49] "HRV_C2d"                      "HRV_C2a"
[51] "HRV_SD2d"                     "HRV_SD2a"
[53] "HRV_Cd"                       "HRV_Ca"
[55] "HRV_SDNNd"                    "HRV_SDNNa"
[57] "HRV_ApEn"                     "HRV_ShanEn"
[59] "HRV_FuzzyEn"                  "HRV_MSEn"
[61] "HRV_CMSEn"                    "HRV_RCMSEn"
[63] "HRV_CD"                       "HRV_HFD"
[65] "HRV_KFD"                      "HRV_LZC"
[67] "HRV_DFA_alpha1"               "HRV_MFDFA_alpha1_Width"
[69] "HRV_MFDFA_alpha1_Peak"        "HRV_MFDFA_alpha1_Mean"
[71] "HRV_MFDFA_alpha1_Max"         "HRV_MFDFA_alpha1_Delta"
[73] "HRV_MFDFA_alpha1_Asymmetry"   "HRV_MFDFA_alpha1_Fluctuation"
[75] "HRV_MFDFA_alpha1_Increment"
```

```
data <- tibble::as_tibble(data)
```

```
# * It is very hard to compare the HR as different predictors are on
↪  different magnitudes, so we need to normalize them.
time_col <- data$time
event_col <- data$event
data <- data %>%
    select(-c(time, event)) %>%
    mutate(across(where(is.numeric), scale)) %>%
    mutate(
        time = time_col,
        event = event_col
    )
```

Note now the interpretation of HR is different! For example, if HR=1.16 for the predictor in the univariate model fitted using scaled data, it means that each standard deviation increase is associated with 16% higher risk of event.

```
data_complete <- na.omit(data)
```

Univariate Cox Proportional Hazards Model

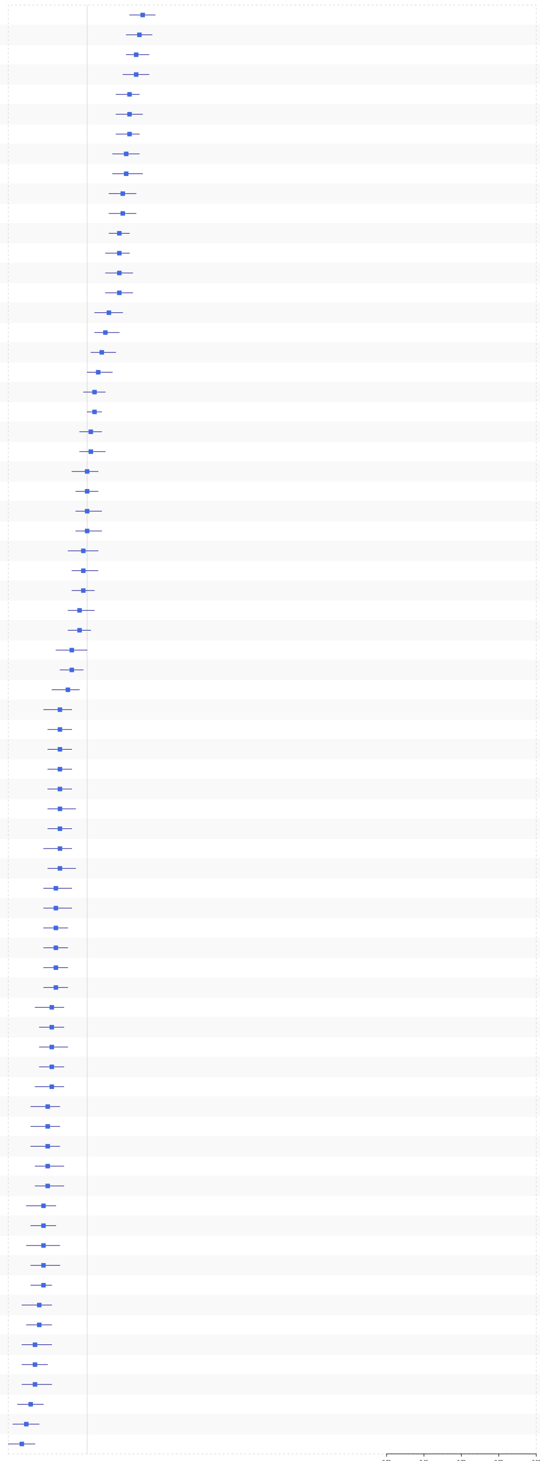
```r
if (!("time" %in% colnames(data) && "event" %in% colnames(data))) {
    stop("time and event columns are required")
}
predictors <- colnames(data)[!colnames(data) %in% c("time", "event")]

results_univariate <- map_dfr(predictors, function(predictor) {
    formula <- as.formula(paste("Surv(time, event) ~", predictor))
    # cox_model_single <- coxph(Surv(time, event) ~ get(predictor), data =
     ↪  data)  # equivalent way
    cox_model_single <- coxph(formula, data = data)

    coef <- coef(cox_model_single) # log hazard ratio
    se <- sqrt(diag(vcov(cox_model_single)))

    hr <- exp(coef)
    lower_ci <- exp(coef - 1.96 * se)
    upper_ci <- exp(coef + 1.96 * se)
    p_value <- summary(cox_model_single)$coefficients[5]

    return(
        data.frame(
            predictor = predictor,
            hr = hr,
            lower_ci = lower_ci,
            upper_ci = upper_ci,
            p_value = p_value
        )
    )
})
results_univariate$hr <- round(results_univariate$hr, 2)
results_univariate$lower_ci <- round(results_univariate$lower_ci, 2)
results_univariate$upper_ci <- round(results_univariate$upper_ci, 2)
results_univariate$ci <- paste0("(", results_univariate$lower_ci, ",",
 ↪  results_univariate$upper_ci, ")")
results_univariate$p_value <- round(results_univariate$p_value, 3)
results_univariate <- results_univariate %>% arrange(desc(hr)) # sort
 ↪  descendingly by HR
```

```r
# Create forest plot
results_univariate %>%
    forestplot(
```

```r
    labeltext = c(predictor, hr, ci, p_value),
    mean = hr,
    lower = lower_ci,
    upper = upper_ci,
    xlab = "Hazard Ratio",
    title = "Univariate Cox Models",
    xlog = TRUE, # * Make sure the CI are not symmetric and need to be
     ↪  transformed
    boxsize = 0.2,
    xticks = c(0.8, 0.9, 1.0, 1.1, 1.2),
    clip = c(0.8, 1.2),
    zero = 1
) %>%
fp_set_style(
    box = "royalblue",
    line = "darkblue",
    summary = "royalblue"
) %>%
fp_add_header(
    predictor = c("Predictor", ""),
    hr = c("Hazard Ratio", "per SD increase"),
    ci = c("95% CI", ""),
    p_value = c("p-value", "")
) %>%
fp_decorate_graph(
    box = gpar(lty = 2, col = "lightgray"),
    graph.pos = 4
) %>% # change the position of forest plot
fp_set_zebra_style("#f9f9f9")
```

Univariate Cox Models

Predictor	Hazard Ratio per SD increase	95% CI	p−value
HRV_PIP	1.16	(1.12,1.2)	0
HRV_IALS	1.15	(1.11,1.19)	0
HRV_SD1SD2	1.14	(1.11,1.18)	0
HRV_HFD	1.14	(1.1,1.18)	0
HRV_PAS	1.12	(1.08,1.15)	0
HRV_GI	1.12	(1.08,1.16)	0
HRV_SI	1.12	(1.08,1.15)	0
HRV_PSS	1.11	(1.07,1.15)	0
HRV_AI	1.11	(1.07,1.16)	0
HRV_ApEn	1.1	(1.06,1.14)	0
HRV_CMSEn	1.1	(1.06,1.14)	0
HRV_VHF	1.09	(1.06,1.12)	0
HRV_C2d	1.09	(1.05,1.12)	0
HRV_Cd	1.09	(1.05,1.13)	0
HRV_RCMSEn	1.09	(1.05,1.13)	0
HRV_C1d	1.06	(1.02,1.1)	0.001
HRV_pNN20	1.05	(1.02,1.09)	0.003
HRV_MFDFA_alpha1_Asymmetry	1.04	(1.01,1.08)	0.012
HRV_MSEn	1.03	(1,1.07)	0.075
HRV_pNN50	1.02	(0.99,1.05)	0.281
HRV_KFD	1.02	(1,1.04)	0.05
HRV_ShanEn	1.01	(0.98,1.04)	0.621
HRV_MFDFA_alpha1_Width	1.01	(0.98,1.05)	0.429
HRV_HF	1	(0.96,1.03)	0.845
HRV_TP	1	(0.97,1.03)	0.984
HRV_MFDFA_alpha1_Fluctuation	1	(0.97,1.04)	0.914
HRV_MFDFA_alpha1_Increment	1	(0.97,1.04)	0.788
HRV_LFHF	0.99	(0.95,1.03)	0.611
HRV_HFn	0.99	(0.96,1.03)	0.719
HRV_LnHF	0.99	(0.96,1.02)	0.598
HRV_S	0.98	(0.95,1.02)	0.301
HRV_MFDFA_alpha1_Mean	0.98	(0.95,1.01)	0.181
HRV_TINN	0.96	(0.92,1)	0.028
HRV_FuzzyEn	0.96	(0.93,0.99)	0.011
HRV_C1a	0.95	(0.91,0.98)	0.001
HRV_SDSD	0.93	(0.89,0.96)	0
HRV_CVSD	0.93	(0.9,0.96)	0
HRV_MadNN	0.93	(0.9,0.96)	0
HRV_MCVNN	0.93	(0.9,0.96)	0
HRV_IQRNN	0.93	(0.9,0.96)	0
HRV_HTI	0.93	(0.9,0.97)	0
HRV_LF	0.93	(0.9,0.96)	0
HRV_SD1	0.93	(0.89,0.96)	0
HRV_SD1d	0.93	(0.9,0.97)	0
HRV_RMSSD	0.92	(0.89,0.96)	0
HRV_SD1a	0.92	(0.89,0.96)	0
HRV_C2a	0.92	(0.89,0.95)	0
HRV_Ca	0.92	(0.89,0.95)	0
HRV_MFDFA_alpha1_Max	0.92	(0.89,0.95)	0
HRV_MFDFA_alpha1_Delta	0.92	(0.89,0.95)	0
HRV_SDNN	0.91	(0.87,0.94)	0
HRV_PI	0.91	(0.88,0.94)	0
HRV_SDNNd	0.91	(0.88,0.95)	0
HRV_CD	0.91	(0.88,0.94)	0
HRV_MFDFA_alpha1_Peak	0.91	(0.87,0.94)	0
HRV_MedianNN	0.9	(0.86,0.93)	0
HRV_Prc80NN	0.9	(0.86,0.93)	0
HRV_SD2	0.9	(0.86,0.93)	0
HRV_SD2d	0.9	(0.87,0.94)	0
HRV_SDNNa	0.9	(0.87,0.94)	0
HRV_MeanNN	0.89	(0.85,0.92)	0
HRV_LFn	0.89	(0.86,0.92)	0
HRV_CSI_Modified	0.89	(0.85,0.93)	0
HRV_SD2a	0.89	(0.86,0.93)	0
HRV_LZC	0.89	(0.86,0.91)	0
HRV_CVNN	0.88	(0.84,0.91)	0
HRV_DFA_alpha1	0.88	(0.85,0.91)	0
HRV_Prc20NN	0.87	(0.84,0.91)	0
HRV_MinNN	0.87	(0.84,0.9)	0
HRV_MaxNN	0.87	(0.84,0.91)	0
HRV_CVI	0.86	(0.83,0.89)	0
HRV_CSI	0.85	(0.82,0.88)	0
HRV_SDRMSSD	0.84	(0.81,0.87)	0



Multivariate Cox Proportional Hazards Model

```r
cox_model_full <- coxph(Surv(time, event) ~ ., data = data)
summary(cox_model_full)
```

```r
cox_model_full_complete <- coxph(Surv(time, event) ~ ., data = data_complete)
summary(cox_model_full_complete)
```

PH Assumption Assessment

```r
cox.zph(cox_model_full)
```

```
                                chisq df    p
HRV_MeanNN                   1.10e+00  1 0.29
HRV_SDNN                     7.94e-01  1 0.37
HRV_RMSSD                    8.20e-01  1 0.37
HRV_SDSD                     8.32e-01  1 0.36
HRV_CVNN                     3.02e-02  1 0.86
HRV_CVSD                     1.03e-01  1 0.75
HRV_MedianNN                 1.79e+00  1 0.18
HRV_MadNN                    3.93e-01  1 0.53
HRV_MCVNN                    6.12e-01  1 0.43
HRV_IQRNN                    7.95e-01  1 0.37
HRV_SDRMSSD                  2.28e-01  1 0.63
HRV_Prc20NN                  2.64e-01  1 0.61
HRV_Prc80NN                  1.55e+00  1 0.21
HRV_pNN50                    8.50e-01  1 0.36
HRV_pNN20                    1.17e+00  1 0.28
HRV_MinNN                    1.49e-02  1 0.90
HRV_MaxNN                    4.55e-01  1 0.50
HRV_HTI                      9.39e-02  1 0.76
HRV_TINN                     5.92e-01  1 0.44
HRV_LF                       5.05e-01  1 0.48
HRV_HF                       2.94e-01  1 0.59
HRV_VHF                      1.67e+00  1 0.20
HRV_LFHF                     6.86e-01  1 0.41
HRV_LFn                      2.09e-02  1 0.88
```

HRV_HFn	1.73e-01	1	0.68
HRV_LnHF	9.73e-02	1	0.76
HRV_SD2	6.66e-01	1	0.41
HRV_SD1SD2	7.02e-01	1	0.40
HRV_S	1.50e+00	1	0.22
HRV_CSI	5.09e-01	1	0.48
HRV_CVI	3.17e-02	1	0.86
HRV_CSI_Modified	5.10e-01	1	0.48
HRV_PIP	5.48e-02	1	0.81
HRV_IALS	1.16e-01	1	0.73
HRV_PSS	8.10e-02	1	0.78
HRV_PAS	2.00e+00	1	0.16
HRV_GI	5.72e-01	1	0.45
HRV_SI	7.15e-02	1	0.79
HRV_AI	6.37e-01	1	0.42
HRV_PI	7.35e-01	1	0.39
HRV_C1d	3.06e-01	1	0.58
HRV_SD1d	7.57e-01	1	0.38
HRV_SD1a	8.99e-01	1	0.34
HRV_C2d	2.00e-01	1	0.65
HRV_SD2d	4.67e-01	1	0.49
HRV_SD2a	7.42e-01	1	0.39
HRV_Cd	6.00e-01	1	0.44
HRV_SDNNd	6.66e-01	1	0.41
HRV_SDNNa	8.69e-01	1	0.35
HRV_ApEn	9.19e-01	1	0.34
HRV_ShanEn	3.02e-01	1	0.58
HRV_FuzzyEn	2.47e-01	1	0.62
HRV_MSEn	7.32e-02	1	0.79
HRV_CMSEn	9.87e-01	1	0.32
HRV_RCMSEn	3.56e-01	1	0.55
HRV_CD	3.82e-02	1	0.85
HRV_HFD	1.19e-01	1	0.73
HRV_KFD	9.54e-01	1	0.33
HRV_LZC	3.03e-02	1	0.86
HRV_DFA_alpha1	5.27e-01	1	0.47
HRV_MFDFA_alpha1_Width	3.60e-02	1	0.85
HRV_MFDFA_alpha1_Peak	8.66e-02	1	0.77
HRV_MFDFA_alpha1_Mean	4.83e-02	1	0.83
HRV_MFDFA_alpha1_Max	4.88e-01	1	0.48
HRV_MFDFA_alpha1_Delta	1.16e-01	1	0.73
HRV_MFDFA_alpha1_Asymmetry	1.03e-05	1	1.00
HRV_MFDFA_alpha1_Fluctuation	4.80e-01	1	0.49

```
HRV_MFDFA_alpha1_Increment   2.00e-01  1 0.66
GLOBAL                       5.46e+01 68 0.88
```

```r
cox.zph(cox_model_full_complete)
```

```
                                chisq df    p
HRV_MeanNN                   1.10e+00  1 0.29
HRV_SDNN                     7.94e-01  1 0.37
HRV_RMSSD                    8.20e-01  1 0.37
HRV_SDSD                     8.32e-01  1 0.36
HRV_CVNN                     3.02e-02  1 0.86
HRV_CVSD                     1.03e-01  1 0.75
HRV_MedianNN                 1.79e+00  1 0.18
HRV_MadNN                    3.93e-01  1 0.53
HRV_MCVNN                    6.12e-01  1 0.43
HRV_IQRNN                    7.95e-01  1 0.37
HRV_SDRMSSD                  2.28e-01  1 0.63
HRV_Prc20NN                  2.64e-01  1 0.61
HRV_Prc80NN                  1.55e+00  1 0.21
HRV_pNN50                    8.50e-01  1 0.36
HRV_pNN20                    1.17e+00  1 0.28
HRV_MinNN                    1.49e-02  1 0.90
HRV_MaxNN                    4.55e-01  1 0.50
HRV_HTI                      9.39e-02  1 0.76
HRV_TINN                     5.92e-01  1 0.44
HRV_LF                       5.05e-01  1 0.48
HRV_HF                       2.94e-01  1 0.59
HRV_VHF                      1.67e+00  1 0.20
HRV_LFHF                     6.86e-01  1 0.41
HRV_LFn                      2.09e-02  1 0.88
HRV_HFn                      1.73e-01  1 0.68
HRV_LnHF                     9.73e-02  1 0.76
HRV_SD2                      6.66e-01  1 0.41
HRV_SD1SD2                   7.02e-01  1 0.40
HRV_S                        1.50e+00  1 0.22
HRV_CSI                      5.09e-01  1 0.48
HRV_CVI                      3.17e-02  1 0.86
HRV_CSI_Modified             5.10e-01  1 0.48
HRV_PIP                      5.48e-02  1 0.81
HRV_IALS                     1.16e-01  1 0.73
HRV_PSS                      8.10e-02  1 0.78
HRV_PAS                      2.00e+00  1 0.16
```

HRV_GI	5.72e-01	1	0.45
HRV_SI	7.15e-02	1	0.79
HRV_AI	6.37e-01	1	0.42
HRV_PI	7.35e-01	1	0.39
HRV_C1d	3.06e-01	1	0.58
HRV_SD1d	7.57e-01	1	0.38
HRV_SD1a	8.99e-01	1	0.34
HRV_C2d	2.00e-01	1	0.65
HRV_SD2d	4.67e-01	1	0.49
HRV_SD2a	7.42e-01	1	0.39
HRV_Cd	6.00e-01	1	0.44
HRV_SDNNd	6.66e-01	1	0.41
HRV_SDNNa	8.69e-01	1	0.35
HRV_ApEn	9.19e-01	1	0.34
HRV_ShanEn	3.02e-01	1	0.58
HRV_FuzzyEn	2.47e-01	1	0.62
HRV_MSEn	7.32e-02	1	0.79
HRV_CMSEn	9.87e-01	1	0.32
HRV_RCMSEn	3.56e-01	1	0.55
HRV_CD	3.82e-02	1	0.85
HRV_HFD	1.19e-01	1	0.73
HRV_KFD	9.54e-01	1	0.33
HRV_LZC	3.03e-02	1	0.86
HRV_DFA_alpha1	5.27e-01	1	0.47
HRV_MFDFA_alpha1_Width	3.60e-02	1	0.85
HRV_MFDFA_alpha1_Peak	8.66e-02	1	0.77
HRV_MFDFA_alpha1_Mean	4.83e-02	1	0.83
HRV_MFDFA_alpha1_Max	4.88e-01	1	0.48
HRV_MFDFA_alpha1_Delta	1.16e-01	1	0.73
HRV_MFDFA_alpha1_Asymmetry	1.03e-05	1	1.00
HRV_MFDFA_alpha1_Fluctuation	4.80e-01	1	0.49
HRV_MFDFA_alpha1_Increment	2.00e-01	1	0.66
GLOBAL	5.46e+01	68	0.88

The proportional hazards assumption was tested using Schoenfeld residuals. None of the variables violated the PH assumption (all $p > 0.05$), indicating that the Cox proportional hazards model was appropriate for our analysis.

Variable Selection

LASSO

```
# * LASSO doesn't allow missing values
set.seed(1234)
x <- as.matrix(data_complete %>% select(-c(time, event)))
y <- Surv(data_complete$time, data_complete$event)

# cox_model_lasso.cv <- cv.glmnet(
#     x,
#     y,
#     family = "cox",
#     alpha = 1, # 1 for LASSO, 0 for Ridge
#     nfolds = 10
# )
# plot(cox_model_lasso.cv)  # Plot partial likelihood deviance vs log(lambda)

# * We choose the range based on plot(cox_model_lasso.cv) for previous run
↪  when not providing lambda_seq
lambda_seq <- exp(seq(-8, -6, length.out = 100))
cox_model_lasso.cv <- cv.glmnet(
    x,
    y,
    family = "cox",
    alpha = 1, # 1 for LASSO, 0 for Ridge
    nfolds = 10,
    lambda = lambda_seq
)
print(cox_model_lasso.cv$lambda.min)
```

```
[1] 0.0007083833
```

```
print(cox_model_lasso.cv$lambda.1se)
```

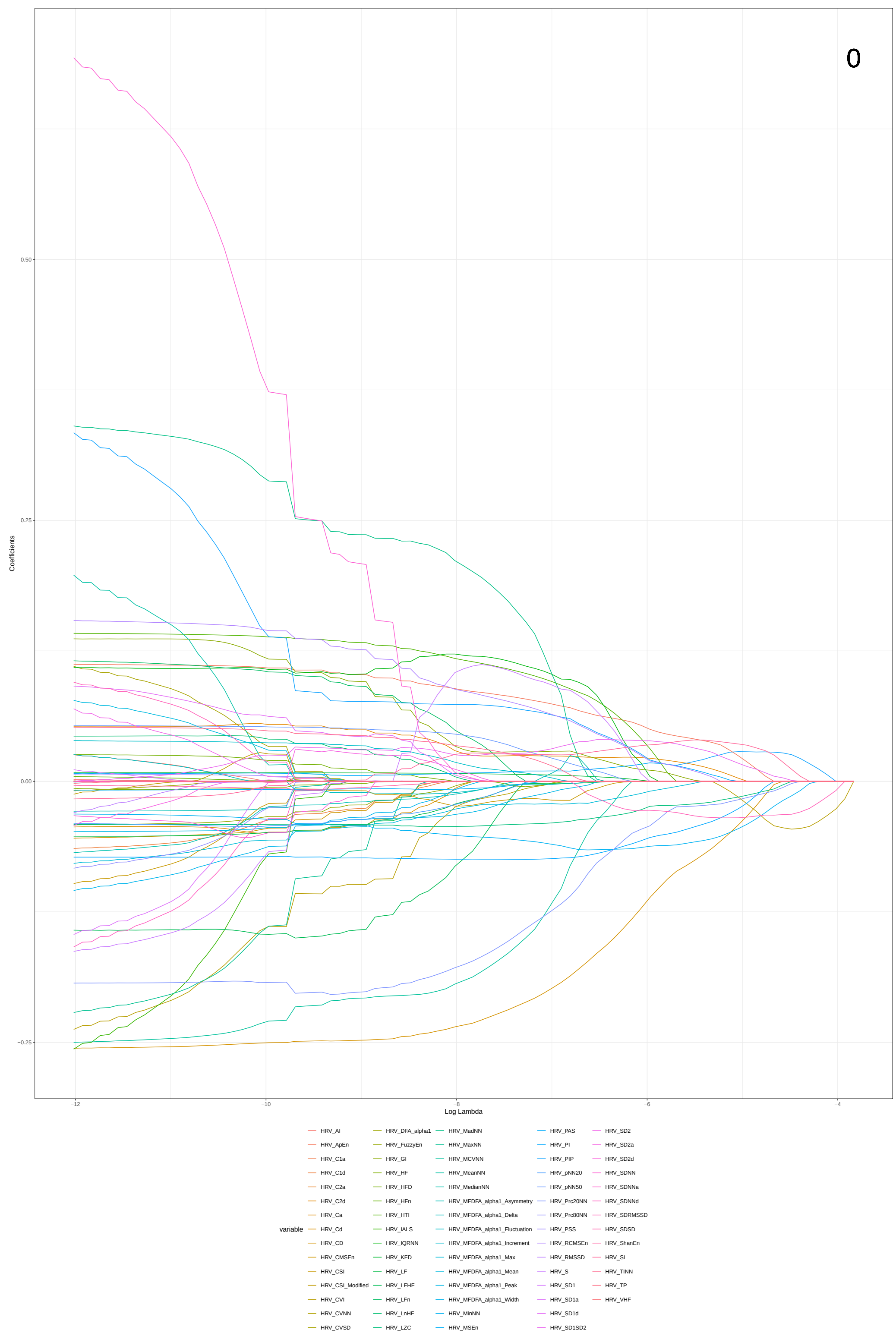
```
[1] 0.002478752
```

As mentioned in the paper, we will use the value of hyperparameter `lambda.1se` that gave the most shrunk model but still was within one standard error from the value that gave the lowest error. This is shown to produce consistently better performance than `lambda.min`.

```r
cox_model_lasso <- glmnet(
    x,
    y,
    family = "cox",
    alpha = 1,
    lambda = cox_model_lasso.cv$lambda.1se
)
cox_model_lasso.coef <- coef(cox_model_lasso)
print(cox_model_lasso.coef)
selected_vars <- rownames(cox_model_lasso.coef)[which(cox_model_lasso.coef !=
 ↪  0)]
print(selected_vars)
```

```r
# * To visualize the LASSO path, we should not supply lambda
cox_model_lasso_fullpath <- glmnet(
    x,
    y,
    family = "cox",
    alpha = 1
)
```

```r
# plot(cox_model_lasso_fullpath, xvar = "lambda", label = TRUE)
autoplot(cox_model_lasso_fullpath, xvar = "lambda", label = TRUE, label.size
 ↪  = 15) +
    theme_bw() +
    theme(legend.position = "bottom") # better way of visualizing the LASSO
     ↪  path
```


0
Coefficients
0.50
0.25
0.00
−0.25
−12
−10
−8
−6
−4
Log Lambda
variable
HRV_AI
HRV_ApEn
HRV_C1a
HRV_C1d
HRV_C2a
HRV_C2d
HRV_Ca
HRV_Cd
HRV_CD
HRV_CMSEn
HRV_CSI
HRV_CSI_Modified
HRV_CVI
HRV_CVNN
HRV_CVSD
HRV_DFA_alpha1
HRV_FuzzyEn
HRV_GI
HRV_HF
HRV_HFD
HRV_HFn
HRV_HTI
HRV_IALS
HRV_IQRNN
HRV_KFD
HRV_LF
HRV_LFHF
HRV_LFn
HRV_LnHF
HRV_LZC
HRV_MadNN
HRV_MaxNN
HRV_MCVNN
HRV_MeanNN
HRV_MedianNN
HRV_MFDFA_alpha1_Asymmetry
HRV_MFDFA_alpha1_Delta
HRV_MFDFA_alpha1_Fluctuation
HRV_MFDFA_alpha1_Increment
HRV_MFDFA_alpha1_Max
HRV_MFDFA_alpha1_Mean
HRV_MFDFA_alpha1_Peak
HRV_MFDFA_alpha1_Width
HRV_MinNN
HRV_MSEn
HRV_PAS
HRV_PI
HRV_PIP
HRV_pNN20
HRV_pNN50
HRV_Prc20NN
HRV_Prc80NN
HRV_PSS
HRV_RCMSEn
HRV_RMSSD
HRV_S
HRV_SD1
HRV_SD1a
HRV_SD1d
HRV_SD1SD2
HRV_SD2
HRV_SD2a
HRV_SD2d
HRV_SDNN
HRV_SDNNa
HRV_SDNNd
HRV_SDRMSSD
HRV_SDSD
HRV_ShanEn
HRV_SI
HRV_TINN
HRV_TP
HRV_VHF

Stepwise Selection based on BIC

```r
# * Stepwise selection doesn't allow missing values
cox_model_step <- MASS::stepAIC(cox_model_full_complete,
    direction = "both",
    k = log(nrow(data)), # Use BIC instead of AIC
    trace = FALSE
)
```

```r
summary(cox_model_step)
```

```
Call:
coxph(formula = Surv(time, event) ~ HRV_MeanNN + HRV_SDNN + HRV_RMSSD +
    HRV_SDSD + HRV_CVNN + HRV_CVSD + HRV_MedianNN + HRV_MadNN +
    HRV_MCVNN + HRV_IQRNN + HRV_SDRMSSD + HRV_Prc20NN + HRV_Prc80NN +
    HRV_pNN50 + HRV_pNN20 + HRV_MinNN + HRV_MaxNN + HRV_HTI +
    HRV_TINN + HRV_LF + HRV_HF + HRV_VHF + HRV_TP + HRV_LFHF +
    HRV_LFn + HRV_HFn + HRV_LnHF + HRV_SD1 + HRV_SD2 + HRV_SD1SD2 +
    HRV_S + HRV_CSI + HRV_CVI + HRV_CSI_Modified + HRV_PIP +
    HRV_IALS + HRV_PSS + HRV_PAS + HRV_GI + HRV_SI + HRV_AI +
    HRV_PI + HRV_C1d + HRV_C1a + HRV_SD1d + HRV_SD1a + HRV_C2d +
    HRV_C2a + HRV_SD2d + HRV_SD2a + HRV_Cd + HRV_RCMSEn + HRV_CD +
    HRV_HFD, data = data_complete)

  n= 26782, number of events= 3386

                    coef  exp(coef)   se(coef)      z Pr(>|z|)
HRV_MeanNN        7.564e-02  1.079e+00  5.484e-01  0.138  0.89030
HRV_SDNN          5.576e-01  1.747e+00  1.253e+00  0.445  0.65624
HRV_RMSSD         2.189e+01  3.201e+09  1.185e+01  1.847  0.06470 .
HRV_SDSD         -2.120e+01  6.180e-10  1.148e+01 -1.847  0.06472 .
HRV_CVNN         -2.283e-01  7.959e-01  3.701e-01 -0.617  0.53740
HRV_CVSD          2.681e-01  1.308e+00  2.849e-01  0.941  0.34663
HRV_MedianNN     -1.269e-01  8.808e-01  1.753e-01 -0.724  0.46908
HRV_MadNN         4.480e-01  1.565e+00  2.019e-01  2.219  0.02650 *
HRV_MCVNN        -3.035e-01  7.382e-01  9.590e-02 -3.165  0.00155 **
HRV_IQRNN         9.520e-02  1.100e+00  1.063e-01  0.895  0.37068
HRV_SDRMSSD       1.530e-01  1.165e+00  3.365e-01  0.455  0.64930
HRV_Prc20NN      -1.625e-01  8.500e-01  9.311e-02 -1.745  0.08100 .
HRV_Prc80NN      -9.657e-02  9.079e-01  1.658e-01 -0.582  0.56032
```

```
HRV_pNN50         5.276e-02  1.054e+00  3.140e-02  1.680  0.09294 .  
HRV_pNN20        -3.467e-02  9.659e-01  4.062e-02 -0.854  0.39333    
HRV_MinNN        -1.644e-02  9.837e-01  3.407e-02 -0.483  0.62935    
HRV_MaxNN        -3.078e-01  7.351e-01  1.544e-01 -1.994  0.04617 *  
HRV_HTI           1.329e-01  1.142e+00  3.081e-02  4.313 1.61e-05 ***
HRV_TINN          5.595e-02  1.058e+00  3.656e-02  1.530  0.12593    
HRV_LF           -1.127e+00  3.239e-01  5.124e+02 -0.002  0.99824    
HRV_HF           -2.680e+00  6.859e-02  1.300e+03 -0.002  0.99836    
HRV_VHF          -1.001e+00  3.674e-01  4.914e+02 -0.002  0.99837    
HRV_TP            3.706e+00  4.067e+01  1.800e+03  0.002  0.99836    
HRV_LFHF         -1.430e-01  8.667e-01  1.627e-01 -0.879  0.37943    
HRV_LFn           1.495e-01  1.161e+00  7.499e-02  1.994  0.04614 *  
HRV_HFn           1.191e-02  1.012e+00  5.815e-02  0.205  0.83770    
HRV_LnHF          3.399e-02  1.035e+00  8.430e-02  0.403  0.68681    
HRV_SD1                  NA         NA  0.000e+00     NA       NA    
HRV_SD2           9.298e-01  2.534e+00  1.402e+00  0.663  0.50715    
HRV_SD1SD2        3.317e-02  1.034e+00  9.595e-02  0.346  0.72953    
HRV_S            -2.282e-01  7.960e-01  3.775e-01 -0.605  0.54551    
HRV_CSI          -3.304e-02  9.675e-01  3.586e-01 -0.092  0.92659    
HRV_CVI          -4.225e-01  6.554e-01  2.866e-01 -1.474  0.14051    
HRV_CSI_Modified -2.870e-01  7.505e-01  3.126e-01 -0.918  0.35856    
HRV_PIP           6.060e-01  1.833e+00  3.116e-01  1.945  0.05180 .  
HRV_IALS         -4.686e-01  6.259e-01  2.917e-01 -1.607  0.10814    
HRV_PSS          -7.486e-03  9.925e-01  3.631e-02 -0.206  0.83667    
HRV_PAS          -6.844e-04  9.993e-01  2.947e-02 -0.023  0.98147    
HRV_GI            2.707e-01  1.311e+00  1.866e-01  1.451  0.14683    
HRV_SI           -5.510e-02  9.464e-01  7.586e-02 -0.726  0.46760    
HRV_AI           -1.271e-01  8.807e-01  2.029e-01 -0.626  0.53110    
HRV_PI           -8.850e-02  9.153e-01  3.162e-02 -2.799  0.00513 ** 
HRV_C1d          -5.642e-02  9.451e-01  7.689e-02 -0.734  0.46309    
HRV_C1a                  NA         NA  0.000e+00     NA       NA    
HRV_SD1d         -1.615e-01  8.509e-01  7.370e-01 -0.219  0.82659    
HRV_SD1a         -8.564e-01  4.247e-01  1.019e+00 -0.840  0.40073    
HRV_C2d           6.128e-02  1.063e+00  6.087e-02  1.007  0.31403    
HRV_C2a                  NA         NA  0.000e+00     NA       NA    
HRV_SD2d         -6.930e-02  9.331e-01  5.377e-01 -0.129  0.89746    
HRV_SD2a         -1.531e-01  8.581e-01  5.569e-01 -0.275  0.78343    
HRV_Cd           -7.895e-02  9.241e-01  8.777e-02 -0.900  0.36835    
HRV_RCMSEn        1.190e-01  1.126e+00  3.009e-02  3.954 7.68e-05 ***
HRV_CD           -2.598e-01  7.712e-01  3.173e-02 -8.186 2.69e-16 ***
HRV_HFD           4.636e-02  1.047e+00  4.188e-02  1.107  0.26834    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
HRV_MeanNN	1.079e+00	9.271e-01	3.682e-01	3.160e+00
HRV_SDNN	1.747e+00	5.726e-01	1.499e-01	2.035e+01
HRV_RMSSD	3.201e+09	3.124e-10	2.634e-01	3.891e+19
HRV_SDSD	6.180e-10	1.618e+09	1.046e-19	3.651e+00
HRV_CVNN	7.959e-01	1.256e+00	3.853e-01	1.644e+00
HRV_CVSD	1.308e+00	7.648e-01	7.481e-01	2.285e+00
HRV_MedianNN	8.808e-01	1.135e+00	6.247e-01	1.242e+00
HRV_MadNN	1.565e+00	6.389e-01	1.054e+00	2.325e+00
HRV_MCVNN	7.382e-01	1.355e+00	6.117e-01	8.908e-01
HRV_IQRNN	1.100e+00	9.092e-01	8.929e-01	1.355e+00
HRV_SDRMSSD	1.165e+00	8.581e-01	6.026e-01	2.253e+00
HRV_Prc20NN	8.500e-01	1.176e+00	7.083e-01	1.020e+00
HRV_Prc80NN	9.079e-01	1.101e+00	6.560e-01	1.257e+00
HRV_pNN50	1.054e+00	9.486e-01	9.912e-01	1.121e+00
HRV_pNN20	9.659e-01	1.035e+00	8.920e-01	1.046e+00
HRV_MinNN	9.837e-01	1.017e+00	9.202e-01	1.052e+00
HRV_MaxNN	7.351e-01	1.360e+00	5.432e-01	9.948e-01
HRV_HTI	1.142e+00	8.756e-01	1.075e+00	1.213e+00
HRV_TINN	1.058e+00	9.456e-01	9.844e-01	1.136e+00
HRV_LF	3.239e-01	3.087e+00	0.000e+00	Inf
HRV_HF	6.859e-02	1.458e+01	0.000e+00	Inf
HRV_VHF	3.674e-01	2.722e+00	0.000e+00	Inf
HRV_TP	4.067e+01	2.459e-02	0.000e+00	Inf
HRV_LFHF	8.667e-01	1.154e+00	6.301e-01	1.192e+00
HRV_LFn	1.161e+00	8.611e-01	1.003e+00	1.345e+00
HRV_HFn	1.012e+00	9.882e-01	9.030e-01	1.134e+00
HRV_LnHF	1.035e+00	9.666e-01	8.770e-01	1.220e+00
HRV_SD1	NA	NA	NA	NA
HRV_SD2	2.534e+00	3.946e-01	1.624e-01	3.954e+01
HRV_SD1SD2	1.034e+00	9.674e-01	8.565e-01	1.248e+00
HRV_S	7.960e-01	1.256e+00	3.798e-01	1.668e+00
HRV_CSI	9.675e-01	1.034e+00	4.790e-01	1.954e+00
HRV_CVI	6.554e-01	1.526e+00	3.737e-01	1.149e+00
HRV_CSI_Modified	7.505e-01	1.332e+00	4.067e-01	1.385e+00
HRV_PIP	1.833e+00	5.455e-01	9.953e-01	3.376e+00
HRV_IALS	6.259e-01	1.598e+00	3.534e-01	1.109e+00
HRV_PSS	9.925e-01	1.008e+00	9.244e-01	1.066e+00
HRV_PAS	9.993e-01	1.001e+00	9.432e-01	1.059e+00
HRV_GI	1.311e+00	7.628e-01	9.094e-01	1.890e+00
HRV_SI	9.464e-01	1.057e+00	8.156e-01	1.098e+00
HRV_AI	8.807e-01	1.135e+00	5.918e-01	1.311e+00

```
HRV_PI            9.153e-01  1.093e+00 8.603e-01 9.738e-01
HRV_C1d           9.451e-01  1.058e+00 8.129e-01 1.099e+00
HRV_C1a                  NA         NA        NA        NA
HRV_SD1d          8.509e-01  1.175e+00 2.007e-01 3.608e+00
HRV_SD1a          4.247e-01  2.355e+00 5.762e-02 3.130e+00
HRV_C2d           1.063e+00  9.406e-01 9.436e-01 1.198e+00
HRV_C2a                  NA         NA        NA        NA
HRV_SD2d          9.331e-01  1.072e+00 3.252e-01 2.677e+00
HRV_SD2a          8.581e-01  1.165e+00 2.880e-01 2.556e+00
HRV_Cd            9.241e-01  1.082e+00 7.780e-01 1.098e+00
HRV_RCMSEn        1.126e+00  8.878e-01 1.062e+00 1.195e+00
HRV_CD            7.712e-01  1.297e+00 7.247e-01 8.207e-01
HRV_HFD           1.047e+00  9.547e-01 9.649e-01 1.137e+00

Concordance= 0.587  (se = 0.005 )
Likelihood ratio test= 320.9  on 51 df,   p=<2e-16
Wald test            = 306.5  on 51 df,   p=<2e-16
Score (logrank) test = 309.9  on 51 df,   p=<2e-16
```

Summary of Variable Selection

We will compare the selection of variables from all models we have built.

```r
# Obtain the selected variables from all models
variable_names_all <- colnames(data) %>%
    setdiff(c("time", "event"))

variable_names_univariate <- results_univariate %>%
    filter(p_value < 0.05) %>%
    pull(predictor)

variable_names_multivariate <- summary(cox_model_full_complete)$coefficients
 ↪  %>%
    as.data.frame() %>%
    rownames_to_column(var = "predictor") %>% # transpose, "predictor" will
     ↪  now be the column name
    filter(`Pr(>|z|)` < 0.05) %>%
    pull(predictor)

variable_names_lasso <-
 ↪  rownames(cox_model_lasso.coef)[which(cox_model_lasso.coef != 0)]
```

```r
variable_names_step <- summary(cox_model_step)$coefficients %>%
    as.data.frame() %>%
    filter(`Pr(>|z|)` < 0.05) %>%
    rownames()
```

```r
variable_selection_matrix <- matrix(
    0,
    nrow = length(variable_names_all),
    ncol = 4 # univariate, multivariate, lasso, stepwise
)
colnames(variable_selection_matrix) <- c("univariate", "multivariate",
 ↪  "lasso", "stepwise")
rownames(variable_selection_matrix) <- variable_names_all

for (variable in variable_names_all) {
    if (variable %in% variable_names_univariate) {
        variable_selection_matrix[variable, "univariate"] <- 1
    }
    if (variable %in% variable_names_multivariate) {
        variable_selection_matrix[variable, "multivariate"] <- 1
    }
    if (variable %in% variable_names_lasso) {
        variable_selection_matrix[variable, "lasso"] <- 1
    }
    if (variable %in% variable_names_step) {
        variable_selection_matrix[variable, "stepwise"] <- 1
    }
}
```

```r
symbol_selected <- "*"

selection_table <- data.frame(
    Variable = variable_names_all,
    Univariate = ifelse(variable_selection_matrix[, "univariate"] == 1,
     ↪  symbol_selected, ""),
    Multivariate = ifelse(variable_selection_matrix[, "multivariate"] == 1,
     ↪  symbol_selected, ""),
    LASSO = ifelse(variable_selection_matrix[, "lasso"] == 1,
     ↪  symbol_selected, ""),
    Stepwise = ifelse(variable_selection_matrix[, "stepwise"] == 1,
     ↪  symbol_selected, "")
```

```r
) %>%
    mutate(Num_Selected = rowSums(variable_selection_matrix)) %>%
    arrange(desc(Num_Selected), Variable) %>%
    as.data.frame() %>%
    remove_rownames()

variable_categories <- sapply(variable_names_all, determine_category)
category_colors <- c(
    "covariate" = "#FFB6C1", #
    "time"      = "#1E90FF", #
    "frequency" = "#32CD32", #
    "poincare"  = "#FF4500", #
    "entropy"   = "#FF8C00", #
    "fractal"   = "#FFD700", #
    "unknown"   = "#000000" #
)
category_colors_names <- c(
    "covariate"  = "pink", #
    "time"       = "blue", #
    "frequency"  = "green", #
    "poincare"   = "red", #
    "entropy"    = "orange", #
    "fractal"    = "gold" #
)
category_legend <- sapply(names(category_colors_names), function(cat) {
    sprintf(
        "%s: %s",
        tools::toTitleCase(cat),
        tools::toTitleCase(category_colors_names[cat])
    )
}) %>%
    paste(collapse = "; ")


selection_table %>%
    kbl(
        caption = "Variable Selection by Different Models",
        align = c("|l", "c", "c", "c", "c", "c|"),
        col.names = c("Variable", "Univariate", "Multivariate", "LASSO",
         ↪  "Stepwise", "Selected Times"),
        longtable = TRUE
    ) %>%
```

```
    kable_styling(
        bootstrap_options = c("striped", "hover", "condensed", "responsive"),
        position = "center",
        font_size = 9,
        latex_options = c("repeat_header", "striped", "HOLD_position")
    ) %>%
    # Add color for different categories of variables
    column_spec(1,
        color =
         ↪  category_colors[variable_categories[selection_table$Variable]],
        bold = TRUE
    ) %>%
    # Add a header colname for four columns: Univariate, Multivariate, LASSO,
     ↪  Stepwise
    add_header_above(c(
        " " = 1,
        "Selection Methods" = 4,
        " " = 1
    )) %>%
    footnote(
        general = sprintf("%s", category_legend),
        general_title = "Note:"
    )
```

```
Warning: 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")

Warning: 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

Table 1: Variable Selection by Different Models

	Selection Methods				
Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
HRV_CD	*	*	*	*	4
HRV_HTI	*	*	*	*	4
HRV_PI	*	*	*	*	4
HRV_RCMSEn	*	*	*	*	4
HRV_ApEn	*	*	*		3
HRV_MCVNN	*	*		*	3

Table 1: Variable Selection by Different Models *(continued)*

Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
HRV_MadNN	*	*		*	3
HRV_MaxNN	*	*		*	3
HRV_PIP	*	*	*		3
HRV_C2a	*		*		2
HRV_C2d	*		*		2
HRV_CSI	*		*		2
HRV_GI	*		*		2
HRV_IQRNN	*		*		2
HRV_LFn	*			*	2
HRV_LZC	*		*		2
HRV_MFDFA_alpha1_Max	*		*		2
HRV_MinNN	*		*		2
HRV_PAS	*		*		2
HRV_Prc20NN	*		*		2
HRV_SD1SD2	*		*		2
HRV_SDRMSSD	*		*		2
HRV_SI	*		*		2
HRV_AI	*				1
HRV_C1a	*				1
HRV_C1d	*				1
HRV_CMSEn	*				1
HRV_CSI_Modified	*				1
HRV_CVI	*				1
HRV_CVNN	*				1
HRV_CVSD	*				1
HRV_Ca	*				1
HRV_Cd	*				1
HRV_DFA_alpha1	*				1
HRV_FuzzyEn	*				1
HRV_HFD	*				1
HRV_IALS	*				1
HRV_KFD			*		1
HRV_LF	*				1
HRV_MFDFA_alpha1_Asymmetry	*				1
HRV_MFDFA_alpha1_Delta	*				1
HRV_MFDFA_alpha1_Peak	*				1
HRV_MeanNN	*				1
HRV_MedianNN	*				1
HRV_PSS	*				1
HRV_Prc80NN	*				1
HRV_RMSSD	*				1
HRV_SD1	*				1
HRV_SD1a	*				1
HRV_SD1d	*				1
HRV_SD2	*				1
HRV_SD2a	*				1
HRV_SD2d	*				1
HRV_SDNN	*				1

Table 1: Variable Selection by Different Models *(continued)*

Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
HRV__SDNNa	*				1
HRV__SDNNd	*				1
HRV__SDSD	*				1
HRV__TINN	*				1
HRV__VHF	*				1
HRV__pNN20	*				1
HRV__HF					0
HRV__HFn					0
HRV__LFHF					0
HRV__LnHF					0
HRV__MFDFA__alpha1__Fluctuation					0
HRV__MFDFA__alpha1__Increment					0
HRV__MFDFA__alpha1__Mean					0
HRV__MFDFA__alpha1__Width					0
HRV__MSEn					0
HRV__S					0
HRV__ShanEn					0
HRV__TP					0
HRV__pNN50					0

Note:

Covariate: Pink; Time: Blue; Frequency: Green; Poincare: Red; Entropy: Orange; Fractal: Gold