Sensitivity Analysis: Cox Proportional Hazards Model

Mingcheng Hu

Table of contents


```
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, "full") #
↪  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: full"
```

```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    89
```

```r
colnames(data)
```

```
 [1] "event"                        "time"
 [3] "age"                          "sex"
 [5] "ethnicity"                    "BMI"
 [7] "smoking"                      "diabetes"
 [9] "systolic_bp"                  "hypertension_treatment"
[11] "total_chol"                   "hdl_chol"
[13] "education"                    "activity"
[15] "max_workload"                 "max_heart_rate"
[17] "HRV_MeanNN"                   "HRV_SDNN"
[19] "HRV_RMSSD"                    "HRV_SDSD"
[21] "HRV_CVNN"                     "HRV_CVSD"
[23] "HRV_MedianNN"                 "HRV_MadNN"
[25] "HRV_MCVNN"                    "HRV_IQRNN"
[27] "HRV_SDRMSSD"                  "HRV_Prc20NN"
[29] "HRV_Prc80NN"                  "HRV_pNN50"
[31] "HRV_pNN20"                    "HRV_MinNN"
[33] "HRV_MaxNN"                    "HRV_HTI"
[35] "HRV_TINN"                     "HRV_LF"
[37] "HRV_HF"                       "HRV_VHF"
[39] "HRV_TP"                       "HRV_LFHF"
[41] "HRV_LFn"                      "HRV_HFn"
[43] "HRV_LnHF"                     "HRV_SD1"
[45] "HRV_SD2"                      "HRV_SD1SD2"
[47] "HRV_S"                        "HRV_CSI"
```

```
[49] "HRV_CVI"                      "HRV_CSI_Modified"
[51] "HRV_PIP"                      "HRV_IALS"
[53] "HRV_PSS"                      "HRV_PAS"
[55] "HRV_GI"                       "HRV_SI"
[57] "HRV_AI"                       "HRV_PI"
[59] "HRV_C1d"                      "HRV_C1a"
[61] "HRV_SD1d"                     "HRV_SD1a"
[63] "HRV_C2d"                      "HRV_C2a"
[65] "HRV_SD2d"                     "HRV_SD2a"
[67] "HRV_Cd"                       "HRV_Ca"
[69] "HRV_SDNNd"                    "HRV_SDNNa"
[71] "HRV_ApEn"                     "HRV_ShanEn"
[73] "HRV_FuzzyEn"                  "HRV_MSEn"
[75] "HRV_CMSEn"                    "HRV_RCMSEn"
[77] "HRV_CD"                       "HRV_HFD"
[79] "HRV_KFD"                      "HRV_LZC"
[81] "HRV_DFA_alpha1"               "HRV_MFDFA_alpha1_Width"
[83] "HRV_MFDFA_alpha1_Peak"        "HRV_MFDFA_alpha1_Mean"
[85] "HRV_MFDFA_alpha1_Max"         "HRV_MFDFA_alpha1_Delta"
[87] "HRV_MFDFA_alpha1_Asymmetry"   "HRV_MFDFA_alpha1_Fluctuation"
[89] "HRV_MFDFA_alpha1_Increment"
```

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

```
# * There are some imputed ethnicity set to "e". We will exclude them at this
↪ time.
data <- data %>%
    filter(ethnicity != "e")

# * We also need to manually relevel the categorical variables
data <- data %>%
    mutate(
        # Set "Never" (0) as baseline for smoking
        smoking = factor(smoking,
            levels = c("0", "1", "2", "-3"),
            labels = c("Never", "Previous", "Current", "Prefer not to
             ↪ answer")
        ),

        # Set "No" (0) as baseline for diabetes
        diabetes = factor(diabetes,
            levels = c("0", "1", "-1", "-3"),
```

```
            labels = c("No", "Yes", "Do not know", "Prefer not to answer")
        ),

        # Ensure other categorical variables are properly factored
        ethnicity = factor(ethnicity,
            levels = c("1", "2", "3", "4", "5", "6"),
            labels = c("White", "Mixed", "Asian/Asian British", "Black/Black
             ↪  British", "Chinese", "Other")
        ),
        education = factor(education,
            levels = c("1", "2", "3", "4", "5", "6", "-7", "-3"),
            labels = c(
                "College/University degree", "A levels/AS levels",
                "O levels/GCSEs", "CSEs", "NVQ/HND/HNC",
                "Other professional", "None of the above",
                "Prefer not to answer"
            )
        ),
        activity = factor(activity,
            levels = c("0", "1", "2"),
            labels = c("Low", "Moderate", "High")
        ),
        sex = factor(sex,
            levels = c("0", "1"),
            labels = c("Female", "Male")
        ),
        hypertension_treatment = factor(hypertension_treatment,
            levels = c("0", "1"),
            labels = c("No", "Yes")
        )
    )
```

```
# * 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

```
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)$coefficients[1:dim(summary(cox_model_single)$coefficients)[1],
 ↪  5]

    if (determine_type(predictor) == "categorical") {
        # exclude -1, -3, -7 in names
        return(
            data.frame(
                predictor = names(coef),
                hr = hr,
```

```r
                lower_ci = lower_ci,
                upper_ci = upper_ci,
                p_value = p_value
            )
        )
    } else {
        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(
        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(0.2, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2),
        clip = c(0.2, 3.2),
        zero = 1
```

```r
  ) %>%
  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
diabetesYes	2.41	(2.04,2.85)	0
hypertension_treatmentYes	2.1	(1.95,2.25)	0
educationNone of the above	2.04	(1.85,2.26)	0
age	1.98	(1.9,2.06)	0
educationPrefer not to answer	1.86	(1.36,2.54)	0
sexMale	1.8	(1.69,1.93)	0
educationNVQ/HND/HNC	1.66	(1.46,1.9)	0
diabetesDo not know	1.55	(0.81,2.98)	0.188
educationOther professional	1.54	(1.33,1.79)	0
systolic_bp	1.42	(1.38,1.47)	0
smokingPrefer not to answer	1.4	(0.79,2.47)	0.248
smokingPrevious	1.28	(1.19,1.38)	0
smokingCurrent	1.27	(1.14,1.42)	0
BMI	1.24	(1.2,1.28)	0
educationO levels/GCSEs	1.21	(1.1,1.33)	0
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	(1.08,1.16)	0
HRV_AI	1.11	(1.07,1.16)	0
HRV_PSS	1.1	(1.07,1.14)	0
HRV_ApEn	1.1	(1.06,1.14)	0
HRV_CMSEn	1.1	(1.06,1.14)	0
educationA levels/AS levels	1.09	(0.97,1.22)	0.144
HRV_VHF	1.09	(1.06,1.13)	0
HRV_C2d	1.09	(1.05,1.13)	0
HRV_Cd	1.09	(1.05,1.13)	0
HRV_RCMSEn	1.09	(1.05,1.13)	0
diabetesPrefer not to answer	1.06	(0.26,4.22)	0.939
total_chol	1.06	(1.03,1.1)	0
HRV_C1d	1.06	(1.02,1.09)	0.002
HRV_pNN20	1.05	(1.02,1.09)	0.003
HRV_MFDFA_alpha1_Asymmetry	1.04	(1.01,1.08)	0.011
HRV_MSEn	1.03	(1,1.07)	0.065
HRV_pNN50	1.02	(0.98,1.05)	0.315
HRV_KFD	1.02	(1,1.04)	0.051
HRV_ShanEn	1.01	(0.98,1.04)	0.593
HRV_MFDFA_alpha1_Width	1.01	(0.98,1.05)	0.411
HRV_HF	1	(0.97,1.03)	0.913
HRV_TP	1	(0.97,1.04)	0.931
HRV_MFDFA_alpha1_Fluctuation	1	(0.97,1.04)	0.901
HRV_MFDFA_alpha1_Increment	1	(0.97,1.04)	0.771
HRV_LFHF	0.99	(0.95,1.03)	0.622
HRV_HFn	0.99	(0.96,1.03)	0.732
HRV_LnHF	0.99	(0.96,1.03)	0.674
activityModerate	0.98	(0.88,1.09)	0.662
HRV_S	0.98	(0.94,1.02)	0.277
HRV_MFDFA_alpha1_Mean	0.98	(0.95,1.01)	0.188
activityHigh	0.96	(0.86,1.07)	0.435
HRV_TINN	0.96	(0.92,1)	0.029
HRV_FuzzyEn	0.96	(0.93,0.99)	0.009
HRV_C1a	0.95	(0.91,0.98)	0.002
educationCSEs	0.94	(0.79,1.1)	0.424
ethnicityMixed	0.93	(0.8,1.09)	0.401
HRV_CVSD	0.93	(0.9,0.96)	0
HRV_MadNN	0.93	(0.89,0.96)	0
HRV_MCVNN	0.93	(0.9,0.96)	0
HRV_IQRNN	0.93	(0.89,0.96)	0
HRV_HTI	0.93	(0.9,0.97)	0
HRV_LF	0.93	(0.9,0.96)	0
HRV_SD1d	0.93	(0.89,0.96)	0
HRV_RMSSD	0.92	(0.89,0.96)	0
HRV_SDSD	0.92	(0.89,0.96)	0
HRV_SD1	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
ethnicityAsian/Asian British	0.91	(0.78,1.05)	0.2
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_SDNN	0.9	(0.87,0.94)	0
HRV_MedianNN	0.9	(0.86,0.93)	0
HRV_Prc80NN	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	(0.86,0.93)	0
HRV_SD2	0.89	(0.86,0.93)	0
HRV_CSI_Modified	0.89	(0.85,0.93)	0
HRV_SD2a	0.89	(0.85,0.92)	0
HRV_LZC	0.89	(0.86,0.91)	0
HRV_DFA_alpha1	0.88	(0.85,0.91)	0
max_workload	0.87	(0.84,0.9)	0
HRV_CVNN	0.87	(0.84,0.91)	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_CSI	0.85	(0.82,0.88)	0
HRV_CVI	0.85	(0.82,0.88)	0
HRV_SDRMSSD	0.84	(0.81,0.87)	0
hdl_chol	0.82	(0.79,0.85)	0
max_heart_rate	0.78	(0.75,0.81)	0
ethnicityOther	0.66	(0.46,0.94)	0.023
ethnicityBlack/Black British	0.51	(0.31,0.85)	0.01
ethnicityChinese	0.48	(0.25,0.92)	0.028

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
age                      2.66e-01  1 0.60618
sex                      6.93e+00  1 0.00848
ethnicity                2.22e+00  5 0.81765
BMI                      5.50e-02  1 0.81457
smoking                  3.37e+00  3 0.33740
diabetes                 4.14e+00  3 0.24646
systolic_bp              4.69e-01  1 0.49347
hypertension_treatment   1.14e+01  1 0.00073
total_chol               1.18e+00  1 0.27710
hdl_chol                 5.43e+00  1 0.01980
education                2.37e+00  7 0.93622
activity                 1.74e+00  2 0.41794
max_workload             3.06e+00  1 0.08017
max_heart_rate           2.94e+00  1 0.08641
HRV_MeanNN               5.04e+00  1 0.02475
HRV_SDNN                 3.01e+00  1 0.08268
HRV_RMSSD                3.28e+00  1 0.07025
HRV_SDSD                 3.31e+00  1 0.06896
HRV_CVNN                 1.30e+00  1 0.25512
HRV_CVSD                 1.39e+00  1 0.23887
HRV_MedianNN             6.72e+00  1 0.00953
HRV_MadNN                3.38e+00  1 0.06608
HRV_MCVNN                1.13e-01  1 0.73656
HRV_IQRNN                3.91e+00  1 0.04796
```

HRV_SDRMSSD	1.40e-01	1	0.70822
HRV_Prc20NN	2.04e+00	1	0.15357
HRV_Prc80NN	5.40e+00	1	0.02015
HRV_pNN50	1.41e-01	1	0.70695
HRV_pNN20	1.55e+00	1	0.21270
HRV_MinNN	9.02e-01	1	0.34233
HRV_MaxNN	2.00e+00	1	0.15730
HRV_HTI	8.59e-01	1	0.35395
HRV_TINN	1.91e+00	1	0.16651
HRV_LF	1.59e+00	1	0.20714
HRV_HF	2.34e-02	1	0.87843
HRV_VHF	1.57e-01	1	0.69194
HRV_LFHF	2.24e+00	1	0.13460
HRV_LFn	1.49e+00	1	0.22222
HRV_HFn	1.49e+00	1	0.22202
HRV_LnHF	1.78e-01	1	0.67305
HRV_SD2	3.03e+00	1	0.08164
HRV_SD1SD2	4.66e-02	1	0.82916
HRV_S	2.38e+00	1	0.12251
HRV_CSI	2.42e-02	1	0.87644
HRV_CVI	2.29e+00	1	0.12985
HRV_CSI_Modified	2.13e+00	1	0.14419
HRV_PIP	5.69e-01	1	0.45052
HRV_IALS	2.98e-01	1	0.58500
HRV_PSS	5.59e-01	1	0.45484
HRV_PAS	1.70e+00	1	0.19165
HRV_GI	1.69e+00	1	0.19373
HRV_SI	2.31e+00	1	0.12882
HRV_AI	1.03e+00	1	0.31130
HRV_PI	2.65e-01	1	0.60664
HRV_C1d	3.50e-01	1	0.55392
HRV_SD1d	3.44e+00	1	0.06361
HRV_SD1a	3.17e+00	1	0.07516
HRV_C2d	2.96e-01	1	0.58615
HRV_SD2d	2.50e+00	1	0.11379
HRV_SD2a	3.32e+00	1	0.06860
HRV_Cd	6.29e-01	1	0.42789
HRV_SDNNd	3.06e+00	1	0.08027
HRV_SDNNa	3.44e+00	1	0.06371
HRV_ApEn	3.01e+00	1	0.08288
HRV_ShanEn	5.80e-01	1	0.44644
HRV_FuzzyEn	6.57e-01	1	0.41778
HRV_MSEn	4.54e-01	1	0.50032

```
HRV_CMSEn                      3.19e+00  1 0.07390
HRV_RCMSEn                     1.87e+00  1 0.17105
HRV_CD                         5.79e-04  1 0.98081
HRV_HFD                        1.13e-01  1 0.73644
HRV_KFD                        8.23e-01  1 0.36421
HRV_LZC                        3.40e-06  1 0.99853
HRV_DFA_alpha1                 1.86e-02  1 0.89149
HRV_MFDFA_alpha1_Width         3.97e-03  1 0.94978
HRV_MFDFA_alpha1_Peak          7.05e-01  1 0.40126
HRV_MFDFA_alpha1_Mean          1.54e-02  1 0.90124
HRV_MFDFA_alpha1_Max           2.80e-02  1 0.86715
HRV_MFDFA_alpha1_Delta         4.98e-02  1 0.82347
HRV_MFDFA_alpha1_Asymmetry     1.94e-02  1 0.88926
HRV_MFDFA_alpha1_Fluctuation 4.39e-03  1 0.94720
HRV_MFDFA_alpha1_Increment     7.86e-04  1 0.97764
GLOBAL                         1.26e+02 97 0.02429
```

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

```
                                  chisq df        p
age                            2.66e-01  1 0.60618
sex                            6.93e+00  1 0.00848
ethnicity                      2.22e+00  5 0.81765
BMI                            5.50e-02  1 0.81457
smoking                        3.37e+00  3 0.33740
diabetes                       4.14e+00  3 0.24646
systolic_bp                    4.69e-01  1 0.49347
hypertension_treatment         1.14e+01  1 0.00073
total_chol                     1.18e+00  1 0.27710
hdl_chol                       5.43e+00  1 0.01980
education                      2.37e+00  7 0.93622
activity                       1.74e+00  2 0.41794
max_workload                   3.06e+00  1 0.08017
max_heart_rate                 2.94e+00  1 0.08641
HRV_MeanNN                     5.04e+00  1 0.02475
HRV_SDNN                       3.01e+00  1 0.08268
HRV_RMSSD                      3.28e+00  1 0.07025
HRV_SDSD                       3.31e+00  1 0.06896
HRV_CVNN                       1.30e+00  1 0.25512
HRV_CVSD                       1.39e+00  1 0.23887
HRV_MedianNN                   6.72e+00  1 0.00953
HRV_MadNN                      3.38e+00  1 0.06608
```

HRV_MCVNN	1.13e-01	1	0.73656
HRV_IQRNN	3.91e+00	1	0.04796
HRV_SDRMSSD	1.40e-01	1	0.70822
HRV_Prc20NN	2.04e+00	1	0.15357
HRV_Prc80NN	5.40e+00	1	0.02015
HRV_pNN50	1.41e-01	1	0.70695
HRV_pNN20	1.55e+00	1	0.21270
HRV_MinNN	9.02e-01	1	0.34233
HRV_MaxNN	2.00e+00	1	0.15730
HRV_HTI	8.59e-01	1	0.35395
HRV_TINN	1.91e+00	1	0.16651
HRV_LF	1.59e+00	1	0.20714
HRV_HF	2.34e-02	1	0.87843
HRV_VHF	1.57e-01	1	0.69194
HRV_LFHF	2.24e+00	1	0.13460
HRV_LFn	1.49e+00	1	0.22222
HRV_HFn	1.49e+00	1	0.22202
HRV_LnHF	1.78e-01	1	0.67305
HRV_SD2	3.03e+00	1	0.08164
HRV_SD1SD2	4.66e-02	1	0.82916
HRV_S	2.38e+00	1	0.12251
HRV_CSI	2.42e-02	1	0.87644
HRV_CVI	2.29e+00	1	0.12985
HRV_CSI_Modified	2.13e+00	1	0.14419
HRV_PIP	5.69e-01	1	0.45052
HRV_IALS	2.98e-01	1	0.58500
HRV_PSS	5.59e-01	1	0.45484
HRV_PAS	1.70e+00	1	0.19165
HRV_GI	1.69e+00	1	0.19373
HRV_SI	2.31e+00	1	0.12882
HRV_AI	1.03e+00	1	0.31130
HRV_PI	2.65e-01	1	0.60664
HRV_C1d	3.50e-01	1	0.55392
HRV_SD1d	3.44e+00	1	0.06361
HRV_SD1a	3.17e+00	1	0.07516
HRV_C2d	2.96e-01	1	0.58615
HRV_SD2d	2.50e+00	1	0.11379
HRV_SD2a	3.32e+00	1	0.06860
HRV_Cd	6.29e-01	1	0.42789
HRV_SDNNd	3.06e+00	1	0.08027
HRV_SDNNa	3.44e+00	1	0.06371
HRV_ApEn	3.01e+00	1	0.08288
HRV_ShanEn	5.80e-01	1	0.44644

```
HRV_FuzzyEn                    6.57e-01  1 0.41778
HRV_MSEn                       4.54e-01  1 0.50032
HRV_CMSEn                      3.19e+00  1 0.07390
HRV_RCMSEn                     1.87e+00  1 0.17105
HRV_CD                         5.79e-04  1 0.98081
HRV_HFD                        1.13e-01  1 0.73644
HRV_KFD                        8.23e-01  1 0.36421
HRV_LZC                        3.40e-06  1 0.99853
HRV_DFA_alpha1                 1.86e-02  1 0.89149
HRV_MFDFA_alpha1_Width         3.97e-03  1 0.94978
HRV_MFDFA_alpha1_Peak          7.05e-01  1 0.40126
HRV_MFDFA_alpha1_Mean          1.54e-02  1 0.90124
HRV_MFDFA_alpha1_Max           2.80e-02  1 0.86715
HRV_MFDFA_alpha1_Delta         4.98e-02  1 0.82347
HRV_MFDFA_alpha1_Asymmetry     1.94e-02  1 0.88926
HRV_MFDFA_alpha1_Fluctuation 4.39e-03  1 0.94720
HRV_MFDFA_alpha1_Increment     7.86e-04  1 0.97764
GLOBAL                         1.26e+02 97 0.02429
```

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

```r
# * LASSO doesn't allow missing values
set.seed(1234)
# x <- as.matrix(data_complete %>% select(-c(time, event)))
# * We need to explicitly use model.matrix for categorical variables
x <- model.matrix(~ . - 1 - time - event, data = data_complete)
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
)
```

```r
# plot(cox_model_lasso.cv)  # Plot partial likelihood deviance vs log(lambda)
print(cox_model_lasso.cv$lambda.min)
```

```
[1] 0.007299261
```

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

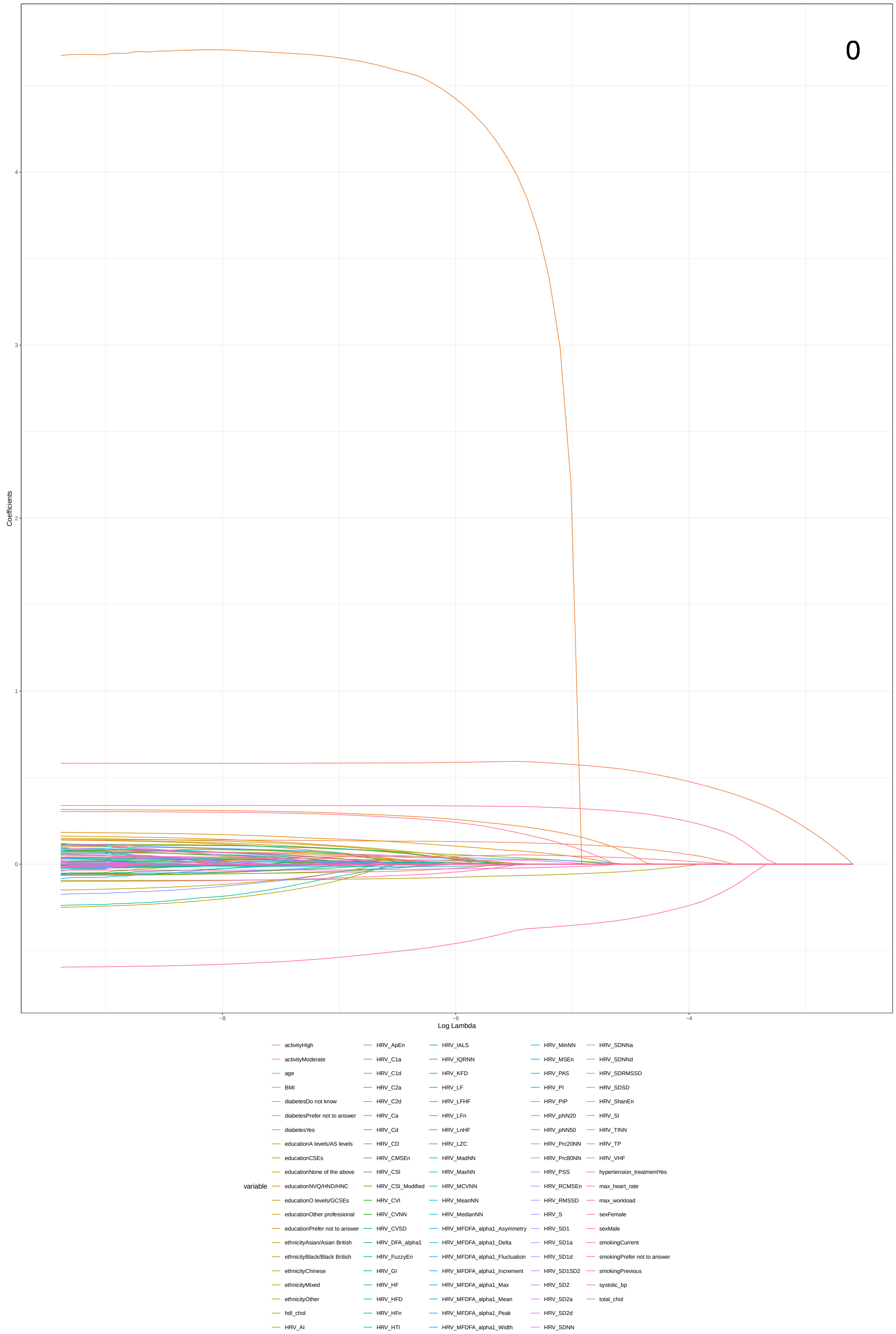
```
[1] 0.01536424
```

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
4
3
2
1
0
−8
−6
−4
Log Lambda
variable
activityHigh
activityModerate
age
BMI
diabetesDo not know
diabetesPrefer not to answer
diabetesYes
educationA levels/AS levels
educationCSEs
educationNone of the above
educationNVQ/HND/HNC
educationO levels/GCSEs
educationOther professional
educationPrefer not to answer
ethnicityAsian/Asian British
ethnicityBlack/Black British
ethnicityChinese
ethnicityMixed
ethnicityOther
hdl_chol
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
hypertension_treatmentYes
max_heart_rate
max_workload
sexFemale
sexMale
smokingCurrent
smokingPrefer not to answer
smokingPrevious
systolic_bp
total_chol

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) ~ age + sex + ethnicity + BMI +
    smoking + diabetes + systolic_bp + hypertension_treatment +
    total_chol + hdl_chol + education + activity + max_workload +
    max_heart_rate + 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_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_PI + HRV_C1d + HRV_Cd + HRV_MSEn +
    HRV_CMSEn + HRV_RCMSEn + HRV_CD + HRV_HFD + HRV_KFD + HRV_LZC +
    HRV_DFA_alpha1 + HRV_MFDFA_alpha1_Width + HRV_MFDFA_alpha1_Peak +
    HRV_MFDFA_alpha1_Mean + HRV_MFDFA_alpha1_Max + HRV_MFDFA_alpha1_Delta +
    HRV_MFDFA_alpha1_Asymmetry + HRV_MFDFA_alpha1_Fluctuation +
    HRV_MFDFA_alpha1_Increment, data = data_complete)

  n= 19577, number of events= 2389

                                  coef  exp(coef)   se(coef)      z Pr(>|z|)
age                          5.861e-01  1.797e+00  2.952e-02 19.857  < 2e-16
sexMale                      5.807e-01  1.787e+00  7.001e-02  8.294  < 2e-16
ethnicityMixed               7.598e-02  1.079e+00  9.744e-02  0.780 0.435535
ethnicityAsian/Asian British  1.575e-01  1.171e+00  9.237e-02  1.705 0.088232
ethnicityBlack/Black British -1.608e-01  8.514e-01  2.916e-01 -0.552 0.581192
ethnicityChinese            -2.680e-01  7.649e-01  4.108e-01 -0.652 0.514081
ethnicityOther              -6.822e-02  9.341e-01  2.157e-01 -0.316 0.751817
BMI                          1.402e-01  1.151e+00  2.367e-02  5.923 3.15e-09
```

smokingPrevious	7.236e-02	1.075e+00	4.425e-02	1.635	0.101979
smokingCurrent	3.037e-01	1.355e+00	7.310e-02	4.155	3.25e-05
smokingPrefer not to answer	-4.663e-02	9.544e-01	4.118e-01	-0.113	0.909826
diabetesYes	3.111e-01	1.365e+00	1.110e-01	2.804	0.005049
diabetesDo not know	7.142e-02	1.074e+00	4.498e-01	0.159	0.873826
diabetesPrefer not to answer	4.548e+00	9.444e+01	1.017e+00	4.473	7.72e-06
systolic_bp	3.863e-02	1.039e+00	2.428e-02	1.591	0.111572
hypertension_treatmentYes	3.431e-01	1.409e+00	4.874e-02	7.039	1.93e-12
total_chol	4.060e-02	1.041e+00	2.236e-02	1.815	0.069482
hdl_chol	-9.409e-02	9.102e-01	2.720e-02	-3.459	0.000543
educationA levels/AS levels	1.034e-01	1.109e+00	6.931e-02	1.493	0.135539
educationO levels/GCSEs	4.498e-02	1.046e+00	5.905e-02	0.762	0.446242
educationCSEs	1.485e-01	1.160e+00	1.062e-01	1.398	0.162208
educationNVQ/HND/HNC	1.515e-01	1.164e+00	8.229e-02	1.841	0.065624
educationOther professional	1.638e-01	1.178e+00	9.074e-02	1.805	0.071106
educationNone of the above	1.882e-01	1.207e+00	6.596e-02	2.853	0.004334
educationPrefer not to answer	1.013e-01	1.107e+00	2.618e-01	0.387	0.698824
activityModerate	1.502e-02	1.015e+00	6.166e-02	0.244	0.807579
activityHigh	-6.190e-02	9.400e-01	6.215e-02	-0.996	0.319258
max_workload	-9.637e-02	9.081e-01	3.574e-02	-2.697	0.007002
max_heart_rate	-3.739e-02	9.633e-01	2.873e-02	-1.302	0.193082
HRV_MeanNN	-3.274e-01	7.208e-01	6.490e-01	-0.505	0.613892
HRV_SDNN	4.413e-01	1.555e+00	1.251e+00	0.353	0.724220
HRV_RMSSD	1.434e+01	1.694e+06	1.420e+01	1.010	0.312453
HRV_SDSD	-1.479e+01	3.767e-07	1.382e+01	-1.070	0.284428
HRV_CVNN	-3.884e-01	6.782e-01	4.208e-01	-0.923	0.356016
HRV_CVSD	1.983e-01	1.219e+00	3.377e-01	0.587	0.557003
HRV_MedianNN	-3.754e-02	9.632e-01	2.144e-01	-0.175	0.860989
HRV_MadNN	2.104e-01	1.234e+00	2.546e-01	0.826	0.408621
HRV_MCVNN	-1.526e-01	8.585e-01	1.197e-01	-1.274	0.202553
HRV_IQRNN	-2.739e-02	9.730e-01	1.280e-01	-0.214	0.830566
HRV_SDRMSSD	3.496e-01	1.419e+00	2.816e-01	1.241	0.214467
HRV_Prc20NN	-1.499e-01	8.608e-01	1.087e-01	-1.380	0.167725
HRV_Prc80NN	-9.533e-02	9.091e-01	1.995e-01	-0.478	0.632817
HRV_pNN50	4.890e-02	1.050e+00	4.150e-02	1.178	0.238651
HRV_pNN20	-8.245e-02	9.209e-01	5.237e-02	-1.574	0.115380
HRV_MinNN	5.338e-02	1.055e+00	3.980e-02	1.341	0.179797
HRV_MaxNN	-1.761e-01	8.386e-01	1.859e-01	-0.947	0.343538
HRV_HTI	1.305e-01	1.139e+00	3.749e-02	3.480	0.000501
HRV_TINN	5.706e-02	1.059e+00	4.097e-02	1.393	0.163682
HRV_LF	-7.462e-02	9.281e-01	6.416e-02	-1.163	0.244851
HRV_HF	-6.530e-02	9.368e-01	1.428e-01	-0.457	0.647555
HRV_TP	7.072e-02	1.073e+00	1.687e-01	0.419	0.675053

HRV_LFHF	-2.764e-01	7.585e-01	2.072e-01	-1.334	0.182289
HRV_LFn	1.841e-01	1.202e+00	9.215e-02	1.998	0.045738
HRV_HFn	4.285e-02	1.044e+00	6.931e-02	0.618	0.536384
HRV_LnHF	6.735e-02	1.070e+00	9.797e-02	0.688	0.491751
HRV_SD1	NA	NA	0.000e+00	NA	NA
HRV_SD2	4.632e-01	1.589e+00	8.422e-01	0.550	0.582345
HRV_SD1SD2	9.770e-02	1.103e+00	1.050e-01	0.930	0.352338
HRV_S	1.631e-01	1.177e+00	4.368e-01	0.374	0.708757
HRV_CSI	-1.375e-01	8.715e-01	2.928e-01	-0.470	0.638577
HRV_CVI	2.846e-01	1.329e+00	3.482e-01	0.818	0.413598
HRV_CSI_Modified	-1.545e-01	8.568e-01	3.617e-01	-0.427	0.669261
HRV_PIP	3.169e-01	1.373e+00	3.694e-01	0.858	0.390989
HRV_IALS	-1.771e-01	8.377e-01	3.459e-01	-0.512	0.608663
HRV_PSS	-2.206e-02	9.782e-01	4.386e-02	-0.503	0.615002
HRV_PAS	-1.499e-02	9.851e-01	3.487e-02	-0.430	0.667333
HRV_GI	1.542e-01	1.167e+00	6.457e-02	2.388	0.016922
HRV_PI	-6.698e-03	9.933e-01	3.431e-02	-0.195	0.845211
HRV_C1d	-5.709e-02	9.445e-01	5.303e-02	-1.077	0.281673
HRV_Cd	3.131e-02	1.032e+00	4.398e-02	0.712	0.476536
HRV_MSEn	-5.214e-03	9.948e-01	3.538e-02	-0.147	0.882822
HRV_CMSEn	4.081e-02	1.042e+00	8.227e-02	0.496	0.619850
HRV_RCMSEn	-1.187e-02	9.882e-01	7.948e-02	-0.149	0.881305
HRV_CD	-2.824e-02	9.722e-01	3.967e-02	-0.712	0.476579
HRV_HFD	8.009e-02	1.083e+00	5.314e-02	1.507	0.131811
HRV_KFD	1.246e-02	1.013e+00	9.643e-03	1.292	0.196207
HRV_LZC	-3.662e-03	9.963e-01	3.549e-02	-0.103	0.917835
HRV_DFA_alpha1	-9.290e-02	9.113e-01	8.978e-02	-1.035	0.300776
HRV_MFDFA_alpha1_Width	-4.519e-02	9.558e-01	1.937e-01	-0.233	0.815563
HRV_MFDFA_alpha1_Peak	-1.569e-02	9.844e-01	5.952e-02	-0.264	0.792121
HRV_MFDFA_alpha1_Mean	1.268e-01	1.135e+00	1.958e-01	0.648	0.517093
HRV_MFDFA_alpha1_Max	6.365e-02	1.066e+00	1.198e-01	0.531	0.595162
HRV_MFDFA_alpha1_Delta	-5.828e-02	9.434e-01	1.235e-01	-0.472	0.636991
HRV_MFDFA_alpha1_Asymmetry	-2.123e-02	9.790e-01	3.818e-02	-0.556	0.578234
HRV_MFDFA_alpha1_Fluctuation	-3.179e-03	9.968e-01	2.017e-01	-0.016	0.987424
HRV_MFDFA_alpha1_Increment	-4.321e-02	9.577e-01	2.575e-01	-0.168	0.866733

age	***
sexMale	***
ethnicityMixed	
ethnicityAsian/Asian British	.
ethnicityBlack/Black British	
ethnicityChinese	
ethnicityOther	

```
BMI                           ***
smokingPrevious
smokingCurrent                ***
smokingPrefer not to answer
diabetesYes                   **
diabetesDo not know
diabetesPrefer not to answer  ***
systolic_bp
hypertension_treatmentYes     ***
total_chol                    .
hdl_chol                      ***
educationA levels/AS levels
educationO levels/GCSEs
educationCSEs
educationNVQ/HND/HNC          .
educationOther professional   .
educationNone of the above    **
educationPrefer not to answer
activityModerate
activityHigh
max_workload                  **
max_heart_rate
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_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_PI
HRV_C1d
HRV_Cd
HRV_MSEn
HRV_CMSEn
HRV_RCMSEn
HRV_CD
HRV_HFD
HRV_KFD
HRV_LZC
HRV_DFA_alpha1
HRV_MFDFA_alpha1_Width
HRV_MFDFA_alpha1_Peak
HRV_MFDFA_alpha1_Mean
HRV_MFDFA_alpha1_Max
HRV_MFDFA_alpha1_Delta
HRV_MFDFA_alpha1_Asymmetry
HRV_MFDFA_alpha1_Fluctuation
HRV_MFDFA_alpha1_Increment
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                               exp(coef) exp(-coef) lower .95 upper .95
age                            1.797e+00  5.565e-01 1.696e+00 1.904e+00
sexMale                        1.787e+00  5.595e-01 1.558e+00 2.050e+00
ethnicityMixed                 1.079e+00  9.268e-01 8.914e-01 1.306e+00
```

ethnicityAsian/Asian British	1.171e+00	8.543e-01	9.767e-01	1.403e+00
ethnicityBlack/Black British	8.514e-01	1.174e+00	4.808e-01	1.508e+00
ethnicityChinese	7.649e-01	1.307e+00	3.419e-01	1.711e+00
ethnicityOther	9.341e-01	1.071e+00	6.120e-01	1.426e+00
BMI	1.151e+00	8.692e-01	1.098e+00	1.205e+00
smokingPrevious	1.075e+00	9.302e-01	9.857e-01	1.172e+00
smokingCurrent	1.355e+00	7.381e-01	1.174e+00	1.564e+00
smokingPrefer not to answer	9.544e-01	1.048e+00	4.259e-01	2.139e+00
diabetesYes	1.365e+00	7.326e-01	1.098e+00	1.697e+00
diabetesDo not know	1.074e+00	9.311e-01	4.448e-01	2.593e+00
diabetesPrefer not to answer	9.444e+01	1.059e-02	1.287e+01	6.929e+02
systolic_bp	1.039e+00	9.621e-01	9.911e-01	1.090e+00
hypertension_treatmentYes	1.409e+00	7.096e-01	1.281e+00	1.551e+00
total_chol	1.041e+00	9.602e-01	9.968e-01	1.088e+00
hdl_chol	9.102e-01	1.099e+00	8.629e-01	9.601e-01
educationA levels/AS levels	1.109e+00	9.017e-01	9.681e-01	1.270e+00
educationO levels/GCSEs	1.046e+00	9.560e-01	9.317e-01	1.174e+00
educationCSEs	1.160e+00	8.620e-01	9.420e-01	1.429e+00
educationNVQ/HND/HNC	1.164e+00	8.594e-01	9.903e-01	1.367e+00
educationOther professional	1.178e+00	8.489e-01	9.860e-01	1.407e+00
educationNone of the above	1.207e+00	8.285e-01	1.061e+00	1.374e+00
educationPrefer not to answer	1.107e+00	9.037e-01	6.624e-01	1.849e+00
activityModerate	1.015e+00	9.851e-01	8.996e-01	1.146e+00
activityHigh	9.400e-01	1.064e+00	8.322e-01	1.062e+00
max_workload	9.081e-01	1.101e+00	8.467e-01	9.740e-01
max_heart_rate	9.633e-01	1.038e+00	9.106e-01	1.019e+00
HRV_MeanNN	7.208e-01	1.387e+00	2.020e-01	2.572e+00
HRV_SDNN	1.555e+00	6.432e-01	1.340e-01	1.804e+01
HRV_RMSSD	1.694e+06	5.904e-07	1.388e-06	2.067e+18
HRV_SDSD	3.767e-07	2.655e+06	6.508e-19	2.180e+05
HRV_CVNN	6.782e-01	1.475e+00	2.973e-01	1.547e+00
HRV_CVSD	1.219e+00	8.201e-01	6.290e-01	2.364e+00
HRV_MedianNN	9.632e-01	1.038e+00	6.327e-01	1.466e+00
HRV_MadNN	1.234e+00	8.103e-01	7.493e-01	2.033e+00
HRV_MCVNN	8.585e-01	1.165e+00	6.789e-01	1.086e+00
HRV_IQRNN	9.730e-01	1.028e+00	7.571e-01	1.250e+00
HRV_SDRMSSD	1.419e+00	7.050e-01	8.168e-01	2.464e+00
HRV_Prc20NN	8.608e-01	1.162e+00	6.957e-01	1.065e+00
HRV_Prc80NN	9.091e-01	1.100e+00	6.148e-01	1.344e+00
HRV_pNN50	1.050e+00	9.523e-01	9.681e-01	1.139e+00
HRV_pNN20	9.209e-01	1.086e+00	8.310e-01	1.020e+00
HRV_MinNN	1.055e+00	9.480e-01	9.757e-01	1.140e+00
HRV_MaxNN	8.386e-01	1.193e+00	5.825e-01	1.207e+00

```
HRV_HTI                         1.139e+00  8.777e-01 1.059e+00 1.226e+00
HRV_TINN                        1.059e+00  9.445e-01 9.770e-01 1.147e+00
HRV_LF                          9.281e-01  1.077e+00 8.184e-01 1.052e+00
HRV_HF                          9.368e-01  1.067e+00 7.080e-01 1.239e+00
HRV_TP                          1.073e+00  9.317e-01 7.711e-01 1.494e+00
HRV_LFHF                        7.585e-01  1.318e+00 5.054e-01 1.139e+00
HRV_LFn                         1.202e+00  8.319e-01 1.003e+00 1.440e+00
HRV_HFn                         1.044e+00  9.581e-01 9.112e-01 1.196e+00
HRV_LnHF                        1.070e+00  9.349e-01 8.828e-01 1.296e+00
HRV_SD1                                NA         NA        NA        NA
HRV_SD2                         1.589e+00  6.293e-01 3.050e-01 8.280e+00
HRV_SD1SD2                      1.103e+00  9.069e-01 8.975e-01 1.355e+00
HRV_S                           1.177e+00  8.495e-01 5.001e-01 2.771e+00
HRV_CSI                         8.715e-01  1.147e+00 4.909e-01 1.547e+00
HRV_CVI                         1.329e+00  7.523e-01 6.718e-01 2.630e+00
HRV_CSI_Modified                8.568e-01  1.167e+00 4.217e-01 1.741e+00
HRV_PIP                         1.373e+00  7.284e-01 6.656e-01 2.832e+00
HRV_IALS                        8.377e-01  1.194e+00 4.253e-01 1.650e+00
HRV_PSS                         9.782e-01  1.022e+00 8.976e-01 1.066e+00
HRV_PAS                         9.851e-01  1.015e+00 9.200e-01 1.055e+00
HRV_GI                          1.167e+00  8.571e-01 1.028e+00 1.324e+00
HRV_PI                          9.933e-01  1.007e+00 9.287e-01 1.062e+00
HRV_C1d                         9.445e-01  1.059e+00 8.513e-01 1.048e+00
HRV_Cd                          1.032e+00  9.692e-01 9.466e-01 1.125e+00
HRV_MSEn                        9.948e-01  1.005e+00 9.282e-01 1.066e+00
HRV_CMSEn                       1.042e+00  9.600e-01 8.865e-01 1.224e+00
HRV_RCMSEn                      9.882e-01  1.012e+00 8.457e-01 1.155e+00
HRV_CD                          9.722e-01  1.029e+00 8.994e-01 1.051e+00
HRV_HFD                         1.083e+00  9.230e-01 9.762e-01 1.202e+00
HRV_KFD                         1.013e+00  9.876e-01 9.936e-01 1.032e+00
HRV_LZC                         9.963e-01  1.004e+00 9.294e-01 1.068e+00
HRV_DFA_alpha1                  9.113e-01  1.097e+00 7.642e-01 1.087e+00
HRV_MFDFA_alpha1_Width          9.558e-01  1.046e+00 6.538e-01 1.397e+00
HRV_MFDFA_alpha1_Peak           9.844e-01  1.016e+00 8.760e-01 1.106e+00
HRV_MFDFA_alpha1_Mean           1.135e+00  8.809e-01 7.735e-01 1.666e+00
HRV_MFDFA_alpha1_Max            1.066e+00  9.383e-01 8.427e-01 1.348e+00
HRV_MFDFA_alpha1_Delta          9.434e-01  1.060e+00 7.406e-01 1.202e+00
HRV_MFDFA_alpha1_Asymmetry      9.790e-01  1.021e+00 9.084e-01 1.055e+00
HRV_MFDFA_alpha1_Fluctuation    9.968e-01  1.003e+00 6.713e-01 1.480e+00
HRV_MFDFA_alpha1_Increment      9.577e-01  1.044e+00 5.782e-01 1.586e+00

Concordance= 0.72  (se = 0.005 )
Likelihood ratio test= 1514  on 85 df,   p=<2e-16
```

```
Wald test            = 1411  on 85 df,   p=<2e-16
Score (logrank) test = 1665  on 85 df,   p=<2e-16
```

Summary of Variable Selection

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

```
# 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)]

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

```
# * We need to handle the categorical variables for correct displaying
get_base_variable_name <- function(variable_name) {
    if
     ↪  (grepl("^(sex|ethnicity|education|smoking|diabetes|activity|hypertension_treatment)[A
     ↪  variable_name)) {
        return(sub("([a-z_]+)[A-Z].*", "\\1", variable_name))
    }
    return(variable_name)
}
```

```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 (determine_type(variable) == "categorical") {
        if (variable %in% sapply(variable_names_univariate,
         ↪  get_base_variable_name)) {
            variable_selection_matrix[variable, "univariate"] <- 1
        }
        if (variable %in% sapply(variable_names_multivariate,
         ↪  get_base_variable_name)) {
            variable_selection_matrix[variable, "multivariate"] <- 1
        }
        if (variable %in% sapply(variable_names_lasso,
         ↪  get_base_variable_name)) {
            variable_selection_matrix[variable, "lasso"] <- 1
        }
        if (variable %in% sapply(variable_names_step,
         ↪  get_base_variable_name)) {
            variable_selection_matrix[variable, "stepwise"] <- 1
        }
    } else {
        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
        }
    }
}
```

```
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, "")
) %>%
    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
BMI	*	*	*	*	4
age	*	*	*	*	4
hdl_chol	*	*	*	*	4
hypertension_treatment	*	*	*	*	4
sex	*	*	*	*	4
HRV_HTI	*	*		*	3
diabetes	*	*		*	3
education	*	*		*	3
max_workload	*	*		*	3
smoking	*	*		*	3
HRV_ApEn	*	*			2
HRV_GI	*			*	2
HRV_LFn	*			*	2
systolic_bp	*		*		2
HRV_AI	*				1
HRV_C1a	*				1
HRV_C1d	*				1
HRV_C2a	*				1
HRV_C2d	*				1
HRV_CD	*				1
HRV_CMSEn	*				1
HRV_CSI	*				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_IQRNN	*				1
HRV_LF	*				1
HRV_LZC	*				1
HRV_MCVNN	*				1
HRV_MFDFA_alpha1_Asymmetry	*				1
HRV_MFDFA_alpha1_Delta	*				1
HRV_MFDFA_alpha1_Max	*				1
HRV_MFDFA_alpha1_Peak	*				1
HRV_MadNN	*				1
HRV_MaxNN	*				1

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

Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
HRV_MeanNN	*				1
HRV_MedianNN	*				1
HRV_MinNN	*				1
HRV_PAS	*				1
HRV_PI	*				1
HRV_PIP	*				1
HRV_PSS	*				1
HRV_Prc20NN	*				1
HRV_Prc80NN	*				1
HRV_RCMSEn	*				1
HRV_RMSSD	*				1
HRV_SD1	*				1
HRV_SD1SD2	*				1
HRV_SD1a	*				1
HRV_SD1d	*				1
HRV_SD2	*				1
HRV_SD2a	*				1
HRV_SD2d	*				1
HRV_SDNN	*				1
HRV_SDNNa	*				1
HRV_SDNNd	*				1
HRV_SDRMSSD	*				1
HRV_SDSD	*				1
HRV_SI	*				1
HRV_ShanEn		*			1
HRV_TINN	*				1
HRV_VHF	*				1
HRV_pNN20	*				1
ethnicity	*				1
max_heart_rate	*				1
total_chol	*				1
HRV_HF					0
HRV_HFn					0
HRV_KFD					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_TP					0
HRV_pNN50					0
activity					0

Note:

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