# Graphical displays for subgroup analysis in clinical trials

## The prostate cancer dataset

To illustrate the proposed methods, we use a prostate carcinoma dataset from a clinical trial of University Hospital Freiburg, 2011. The data has been analyzed in the literature before, and Rosenkranz (2016) used it to illustrate subgroups analysis by model selection. The dataset consist of 475 subjects randomized to a placebo group and three dose levels of diethyl stilbestrol. For this analysis, the control arm consist on the placebo and the lowest dose level of treatment, while the active treatment arm consist of the higher doses are of treatment.

```r
library(sas7bdat)
library(survival)
library(ggplot2)
library(dplyr)

# Read the SAS dataset into R
prca0<-read.sas7bdat('adv_prostate_ca.sas7bdat')
# Select the variables that we use for the analysis
prca <- prca0[,c("SURVTIME","CENS","RX","BM","HX","STAGE","PF", "AGE", "WT")]
names(prca)<- c("survtime","cens","rx","bm",
                "hx","stage","pf","age", "wt")
head(prca)
```

```
##   survtime cens rx bm hx stage pf age  wt
## 1     72.5    0  0  0  0     3  0  75  76
## 2     40.5    1  1  0  1     3  0  69 102
## 3     20.5    1  0  0  1     3  1  75  94
## 4     65.5    0  0  0  0     3  0  67  99
## 5     24.5    1  0  0  0     3  0  71  98
## 6     46.5    1  0  0  0     3  0  75 100
```

```r
# Create subgroups for Age and Weight and Stage
prca$age1 <- 1 * (prca$age > 65 & prca$age <= 75)
prca$age2 <- 1 * (prca$age > 75)
prca$wt1  <- 1 * (prca$wt > 90 & prca$wt <= 110)
prca$wt2  <- 1 * (prca$wt > 110)
# Create subgroups for Age and Weight and Stage
prca$agegroup <- 1 + (1 * (prca$age > 65) + 1 * (prca$age > 75))
prca$wtgroup  <- 1 + (1 * (prca$wt > 90) + 1 * (prca$wt > 110))

formula = Surv(survtime, cens) ~ rx
reg <- survival::coxph(formula, data=prca, ties="breslow")
survival::coxph(formula,
                data = prca, ties="breslow") -> reg
exp(reg$coefficients)
```

```
##        rx
## 0.8385597
```

```r
p.val <- survdiff(Surv(survtime, cens)~rx, data=prca)
lrt.p.val <- 1 - pchisq(p.val$chisq, length(p.val$n) - 1)
```

The p-value of the log-rank test for the difference between treatments was 0.103 and the Kaplan-Meier curves by treatment are shown in Figure S1.

We are interested now in identifying subgroups of patients that may benefit from the treatment using the predicted individual treatment effect (PITE). There are $K = 6$ variables to consider: existence of bone metastasis (bm), disease stage (3 or 4), performance (pf), history of cardiovascular events (hx), age, and weight (wt).

## Level plot

```
# Calculate HR for subgroups
prca %>%
  group_by(agegroup, wtgroup) %>%
  do(data.frame(hr = exp(survival::coxph(formula, data = ., ties="breslow")$coefficients),
                # broom::tidy(survival::coxph(formula, data = ., ties="breslow")),
                n  = nrow(.))) -> df1
# Calculate HR for subgroups. Marginally for subgroups degined by age
prca %>%
  group_by(agegroup) %>%
  do(data.frame(hr = exp(survival::coxph(formula, data = ., ties="breslow")$coefficients),
                # broom::tidy(survival::coxph(formula, data = ., ties="breslow")),
                n  = nrow(.))) -> df2
# Calculate HR for subgroups. Marginally for subgroups degined by wt
prca %>%
  group_by(wtgroup) %>%
  do(data.frame(hr = exp(survival::coxph(formula, data = ., ties="breslow")$coefficients),
                # broom::tidy(survival::coxph(formula, data = ., ties="breslow")),
                n  = nrow(.))) -> df3
```

```
range(prca$age)
```

```
## [1] 48 89
```

```
range(prca$wt)
```

```
## [1]  69 152
```

```
agegroups = c("[48, 65]", "(65, 75]", "(75, 89]")
wtgroups  = c("[69, 90]", "(90, 110]", "(110, 152]")

ggplot() +
  geom_tile(aes(x = agegroup, y = wtgroup, fill = hr), df1, color = "white") +
  geom_text(aes(x = agegroup, y = wtgroup, label = n), df1, color = "white") +
  geom_tile(aes(x = agegroup, y = -0.25, fill = hr),   df2, color = "white", height = 0.35) +
  geom_tile(aes(x = -0.25, y = wtgroup, fill = hr),    df3, color = "white", width = 0.35) +
  ggtitle("Hazard ratio across subgroups (N = 475)") +
  coord_fixed() +
  theme_bw() +
  theme(panel.grid = element_blank(),
        panel.border = element_blank(),
        axis.title = element_text(face = "bold", hjust = 0.63),
        plot.title = element_text(face = "bold"),
        legend.title.align = 0.5,
        legend.title = element_text(face = "bold", angle = 90),
```

```r
      axis.text = element_blank(),
      axis.ticks = element_blank()) +
  guides(fill = guide_colourbar(title.position = "left")) +
  scale_fill_gradientn(name = "Hazard Ratio", breaks = seq(0,2,0.5), limits=c(0,2),
                       colours=c("yellow", "red", "black")) +
  # scale_fill_continuous(name = "Hazard Ratio", breaks = seq(0,2,0.5), limits=c(0,2)) +
  scale_x_continuous(name = "Age", expand = c(0,0.05)) +
  scale_y_continuous(name = "Weight",expand = c(0,0.05)) +
  annotate(x=0.25, y=seq(1,3,1), label = wtgroups, angle = 90, size = 3, geom="text") +
  annotate(x=0.4, xend=0.4, y=0.5, yend=3.5, lwd=0.75, geom="segment") +
  annotate(x=0.35, xend=0.4, y=seq(0.5,3.5,1), yend=seq(0.5,3.5,1), lwd=0.75, geom="segment") +
  annotate(x=seq(1,3,1), y=0.25, label = agegroups, size = 3, geom="text") +
  annotate(x=0.5, xend=3.5, y=0.4, yend=0.4, lwd=0.75, geom="segment") +
  annotate(x=seq(0.5,3.5,1), xend=seq(0.5,3.5,1), y=0.35, yend=0.4, lwd=0.75, geom="segment") -> p1
p1
```



**Hazard ratio across subgroups (N = 475)**
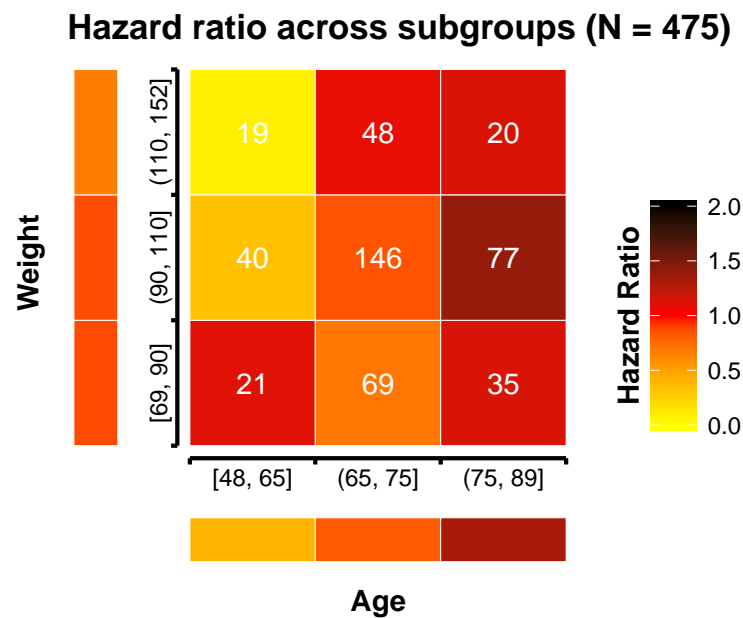
Contour Plots

Venn Diagram

# Bar Chart

```
## Bar Plots -----
prca %>%
  group_by(agegroup, wtgroup) %>%
  do(data.frame(hr = exp(survival::coxph(formula, data = ., ties="breslow")$coefficients),
                broom::tidy(survival::coxph(formula, data = ., ties="breslow")),
                n  = nrow(.))) -> df1
df1 %>%
  ungroup() %>%
  mutate(weight = sqrt(n) / sqrt(nrow(prca)),
         cumsum = cumsum(weight),
         new_x = cumsum - weight/2) -> df1

agegroups = c("1" = "[48, 65]", "2" = "(65, 75]", "3" = "(75, 89]")
wtgroups  = c("[69, 90]", "(90, 110]", "(110, 152]")

ggplot(df1) +
  geom_errorbar(aes(x = new_x, ymin = hr, ymax = hr + std.error),
                width = .1) +
  geom_bar(aes(x = new_x, y = hr, fill = factor(wtgroup)), width = df1$weight-0.05,
           stat = "identity", color = "black") +
  facet_grid(~ agegroup, scales = "free_x", space = "free_x",
             labeller = labeller(agegroup = agegroups)) +
  ggtitle("Hazard ratio across subgroups (N = 475)") +
  scale_x_continuous(name = "Weight", breaks = df1$new_x, labels = rep(wtgroups, 3),
                     sec.axis = sec_axis(~ ., name = "Age", labels = NULL)) +
  scale_y_continuous(name = "Hazard ratio", limits = c(0, 2), expand = c(0,0)) +
  theme_bw() +
  theme(panel.grid = element_blank(),
        legend.position = "none",
        strip.background = element_blank(),
        plot.title = element_text(face = "bold", hjust = 0.5),
        axis.title = element_text(face = "bold"),
        axis.ticks.x = element_blank(),
        axis.text = element_text(angle = 90, hjust = 1))
```

Hazard ratio across subgroups (N = 475)
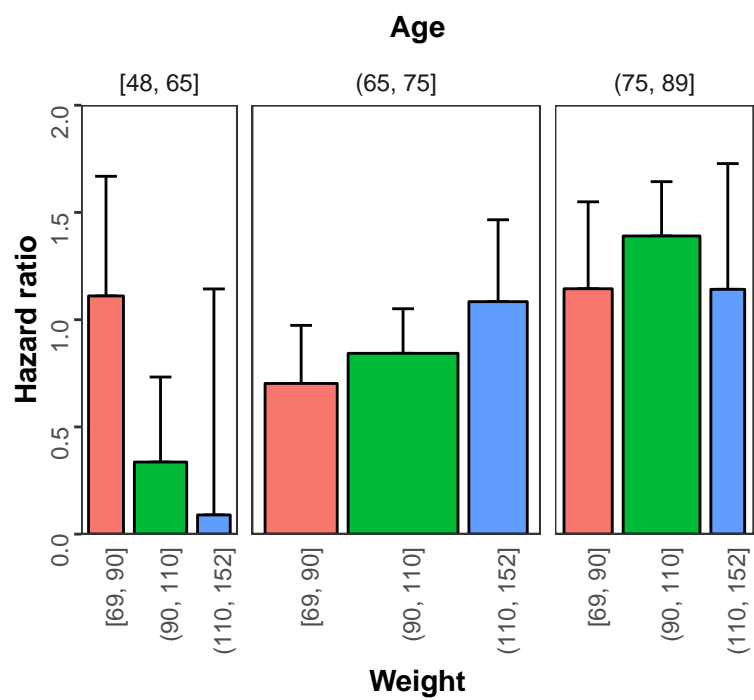
# Forest Plots

```
## Forest Plot -----
# Create subgroups for Age and Weight and Stage
prca$agegroup <- 1 * (prca$age > 75)
prca$wtgroup  <- 1 * (prca$wt > 100)
prca$stage <- prca$stage - 3

prca %>%
  do(data.frame(subgroup = 2,
                hr = exp(survival::coxph(formula, data = ., ties="breslow")$coefficients),
                broom::tidy(survival::coxph(formula, data = ., ties="breslow")),
                n  = nrow(.))) %>%
  mutate(variable = "Overall",
         hr.ll = exp(conf.low),
         hr.ul = exp(conf.high)) %>%
  group_by(subgroup) -> overall
# overall

hrSubgr <- function(data, var){
  data %>%
    group_by_(var) %>%
    do(data.frame(hr = exp(survival::coxph(formula, data = ., ties="breslow")$coefficients),
                  broom::tidy(survival::coxph(formula, data = ., ties="breslow")),
                  n  = nrow(.))) %>%
    mutate(variable = var,
           hr.ll = exp(conf.low),
           hr.ul = exp(conf.high)) %>%
    rename_(subgroup = var)%>%
    ungroup()
}

bind_rows(overall,
  hrSubgr(prca, "agegroup"),
  hrSubgr(prca, "wtgroup"),
  hrSubgr(prca, "pf"),
  hrSubgr(prca, "bm"),
  hrSubgr(prca, "hx"),
  hrSubgr(prca, "stage")) -> fp.dat
fp.dat %>%
  mutate(variable = factor(variable,
                           levels = c("Overall", "agegroup", "wtgroup", "pf", "bm", "hx", "stage")),
         label = sprintf("%s, %s, %s, %s, %s", variable, subgroup, hr, hr.ll, hr.ul)) -> fp.dat


fp.dat %>%
  ggplot() +
  geom_errorbar(aes(ymin=hr.ll, ymax=hr.ul, x=subgroup), width = 0) +
  geom_point(aes(y=hr, x=subgroup, size = n, fill = as.factor(subgroup)), shape = 22) +
  geom_hline(yintercept = fp.dat$hr[1]) +
  geom_hline(yintercept = 1, linetype = 2) +
  scale_x_continuous(name = "",
                     labels = c("a", "b"),
```

```
                        breaks = c(0,1),
                        expand = c(0.5,0.5)) +
    scale_y_continuous(name = "Hazard Ratio") +
    coord_flip() +
    facet_grid(variable ~ ., switch = "y",
               scales = "free_y") +
    theme_bw() +
    theme(panel.grid = element_blank(),
          panel.border = element_blank(),
          panel.spacing = unit(0, units = "cm"),
          legend.position = "none",
          strip.background = element_blank(),
          strip.text.y = element_text(face = "bold", angle = 180, hjust = 0),
          strip.placement = "outside",
          plot.title = element_text(face = "bold", hjust = 0.5),
          axis.line.x = element_line(),
          axis.title = element_text(face = "bold"))
```
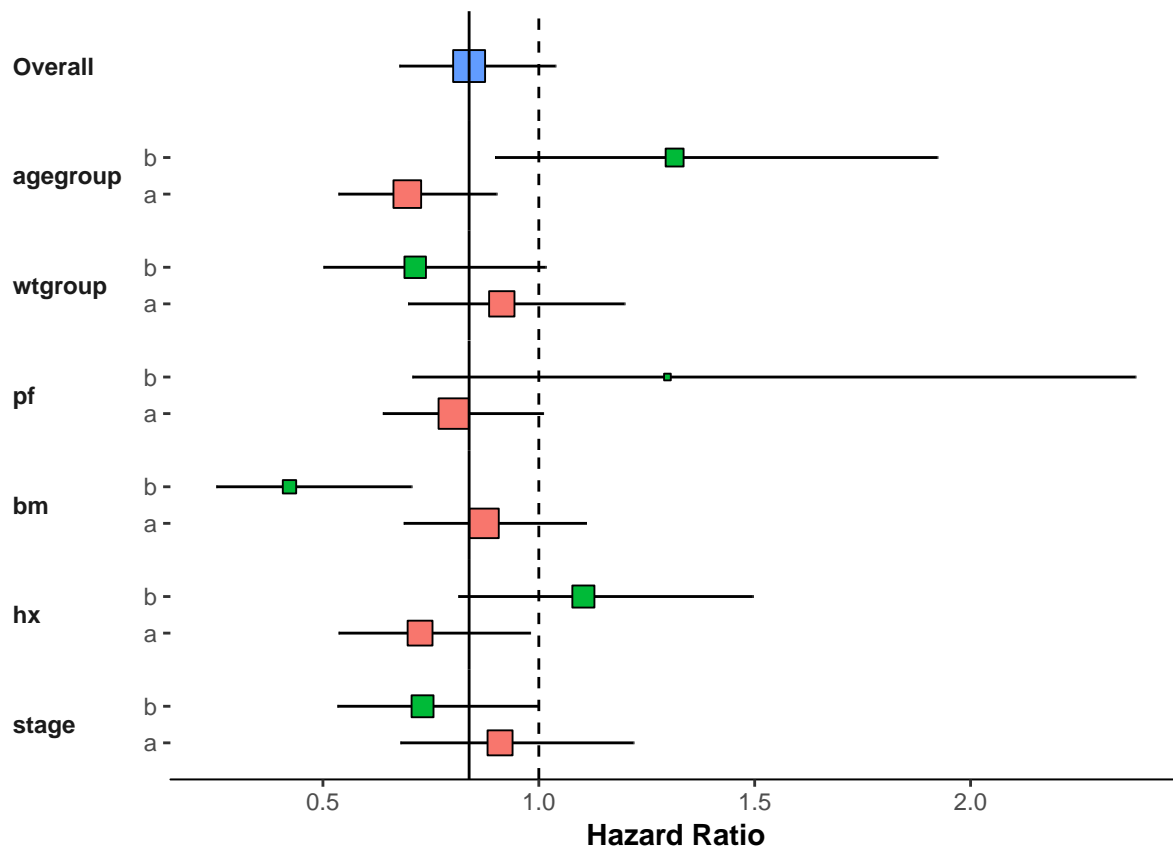


```
## Using the forestplot package
library(forestplot)
tabletext <- cbind(c("Biomarker","\n",as.character(fp.dat$variable)),
                   c("Subgroup","\n",as.character(fp.dat$subgroup)),
                   c("Estimate","\n",round(fp.dat$hr,2)),
                   c("n","\n",fp.dat$n))
forestplot(labeltext=tabletext, graph.pos=3,
           mean=c(NA,NA,drop(fp.dat$hr)),
           lower=c(NA,NA,drop(fp.dat$hr.ll)),
```
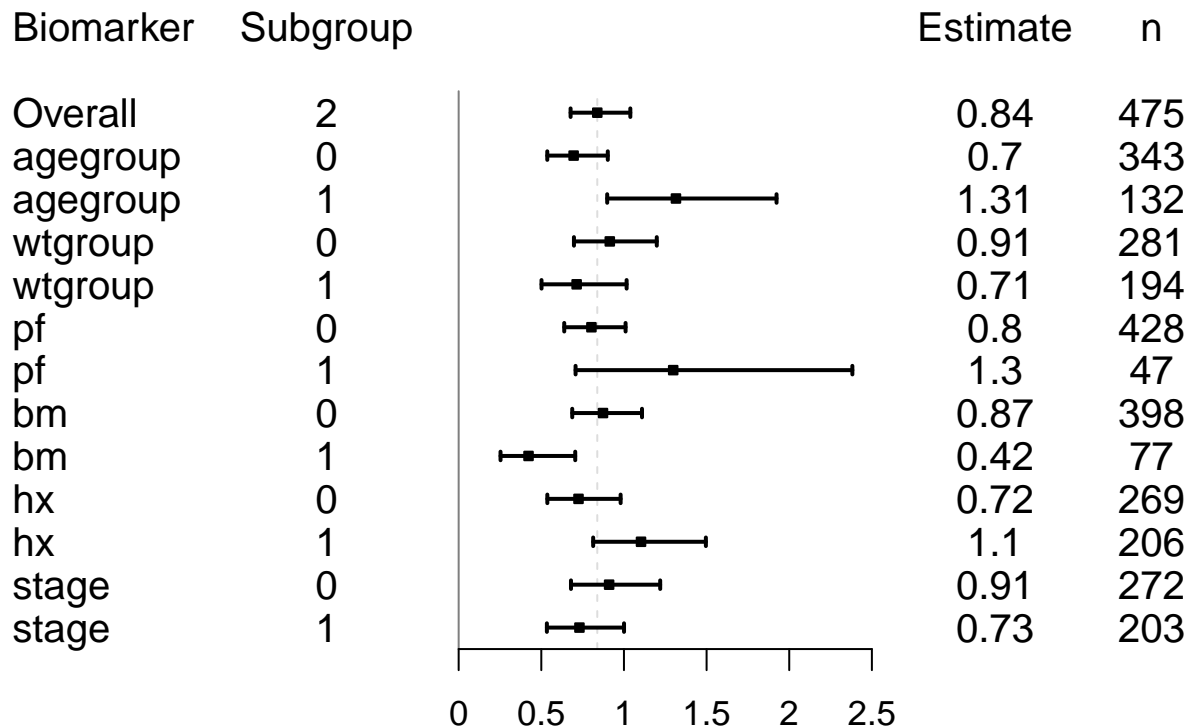
```
            upper=c(NA,NA,drop(fp.dat$hr.ul)),
            title="Treatment Effect",
            # hrzl_lines=sapply(indexes, function(x)
            #   gpar(lwd=1/length(variables.notrt)*600,
            #       lineend="butt",
            #       columns=c(1:4), col="#99999922"),simplify=FALSE, USE.NAMES=TRUE),
            # "31" = gpar(lwd=60, lineend="butt", columns=c(2:6), col="#99999922")),
            txt_gp=fpTxtGp(label=gpar(cex=1.25),
                        ticks=gpar(cex=1.1),
                        xlab=gpar(cex = 1.2),
                        title=gpar(cex = 1.2)),grid = fp.dat$hr[1],
          col=fpColors(box="black", lines="black", zero = "gray50"),
          zero=0, cex=0.9, lineheight = "auto", boxsize=0.2, colgap=unit(6,"mm"),
          lwd.ci=2, ci.vertices=TRUE, ci.vertices.height = 0.1)
```
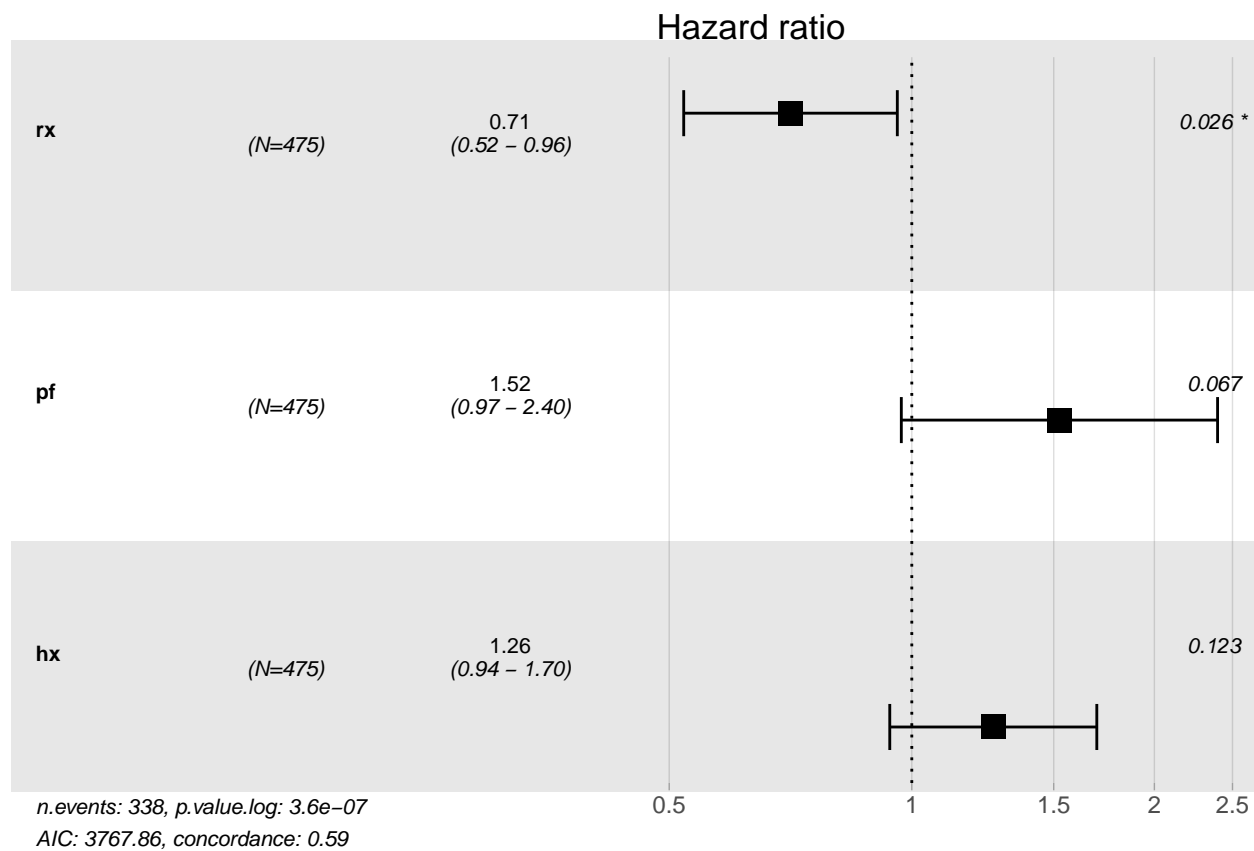


**Treatment Effect**

| Biomarker | Subgroup | | Estimate | n |
|-----------|----------|---|----------|---|
| Overall | 2 | | 0.84 | 475 |
| agegroup | 0 | | 0.7 | 343 |
| agegroup | 1 | | 1.31 | 132 |
| wtgroup | 0 | | 0.91 | 281 |
| wtgroup | 1 | | 0.71 | 194 |
| pf | 0 | | 0.8 | 428 |
| pf | 1 | | 1.3 | 47 |
| bm | 0 | | 0.87 | 398 |
| bm | 1 | | 0.42 | 77 |
| hx | 0 | | 0.72 | 269 |
| hx | 1 | | 1.1 | 206 |
| stage | 0 | | 0.91 | 272 |
| stage | 1 | | 0.73 | 203 |

```
## Using the survminer package
library(survival)
library(survminer)
model <- coxph( Surv(survtime, cens) ~ rx + pf + hx + rx*pf + rx * hx, data = prca )
ggforest(model)
```

Hazard ratio

| | | | | | |
|---|---|---|---|---|---|
| **rx** | (N=475) | 0.71 (0.52 – 0.96) | | | 0.026 * |
| **pf** | (N=475) | 1.52 (0.97 – 2.40) | | | 0.067 |
| **hx** | (N=475) | 1.26 (0.94 – 1.70) | | | 0.123 |

n.events: 338, p.value.log: 3.6e−07
AIC: 3767.86, concordance: 0.59

## Tree Plot

## Galbraith Plot

## L'Abbe' Plot

## STEPP approaches