OCM_modelvalidation

Elizabeth Chase

7/27/2020

This script demonstrates how the model was validated in PLCO. Unfortunately, the PLCO data cannot be shared, and therefore this script cannot be run. However, we provide the code for transparency.

Validation

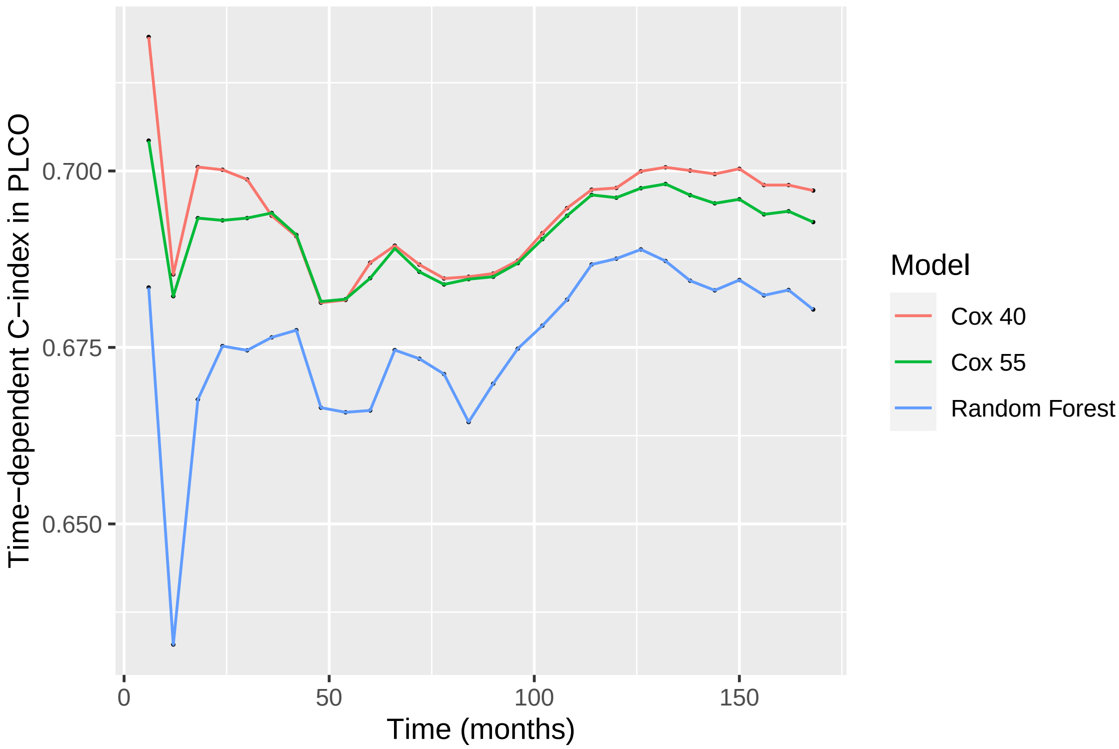
Now we validate each model. First, we look at C-index:

```
load("cox_55.RData")
load("cox_40.RData")
load("rforest_40.RData")

times <- seq(6, 168, by=6)
plco_clean$age_ctr_40 <- plco_clean$age - 60.34387 #scaling the age 40+ cohort
plco_clean$age_ctr_55 <- plco_clean$age - 68.35788 #scaling the age 55+ cohort

cox55_plco <- pec::cindex(object=cox_55, Surv(permth_exm, mortstat) ~ age_ctr_55 +
                            race2 + educ + marital2 + emphysema + diabetic +
                            stroke + smoker + underweight + overweight2 + obese2 +
                            pc + age_ctr_55*diabetic + age_ctr_55*educ +
                            age_ctr_55*marital2 + race2*educ, data = plco_clean,
                          eval.times = times)

cox40_plco <- pec::cindex(object=cox_40, formula = Surv(permth_exm, mortstat) ~
                            age_ctr_40 + diabetic + educ + hypertension + marital2 +
                            underweight + overweight2 + obese2 + smoker + stroke +
                            age_ctr_40*diabetic + age_ctr_40*educ +
                            age_ctr_40*hypertension + age_ctr_40*stroke + pc,
                          data=plco_clean, eval.times = times)

forest_plco <- pec::cindex(object=rforest_40, formula = Surv(permth_exm, mortstat) ~
                             age + arthritis + bronch + diabetic + educ +
                             emphysema + hypertension + single + sep + mi_chd +
                             underweight + overweight_ex + obese + liver + black +
                             other + smoker + stroke + pc, data=plco_nf, eval.times = times)

valid_perf <- data.frame("Time" = rep(seq(6, 168, by=6), 3),
                         "Model" = c(rep("Cox 55", 28), rep("Cox 40", 28),
                                     rep("Random Forest", 28)),
                         "C" = c(cox55_plco$AppCindex$coxph, cox40_plco$AppCindex$coxph,
                                 forest_plco$AppCindex$rfsrc))

performance_plot <- ggplot(data=valid_perf, aes(x=Time, y=C)) + geom_point(size=0.1) +
```

Table 1: Time-dependent C-index in PLCO

Model	Year5	Year10	Year14
Cox 55	0.685	0.696	0.693
Cox 40	0.687	0.698	0.697
Random Forest	0.666	0.688	0.680

```
  geom_line(data=valid_perf, aes(group=Model, color=Model)) + xlab("Time (months)") +
  ylab("Time-dependent C-index in PLCO")

performance_plot
```



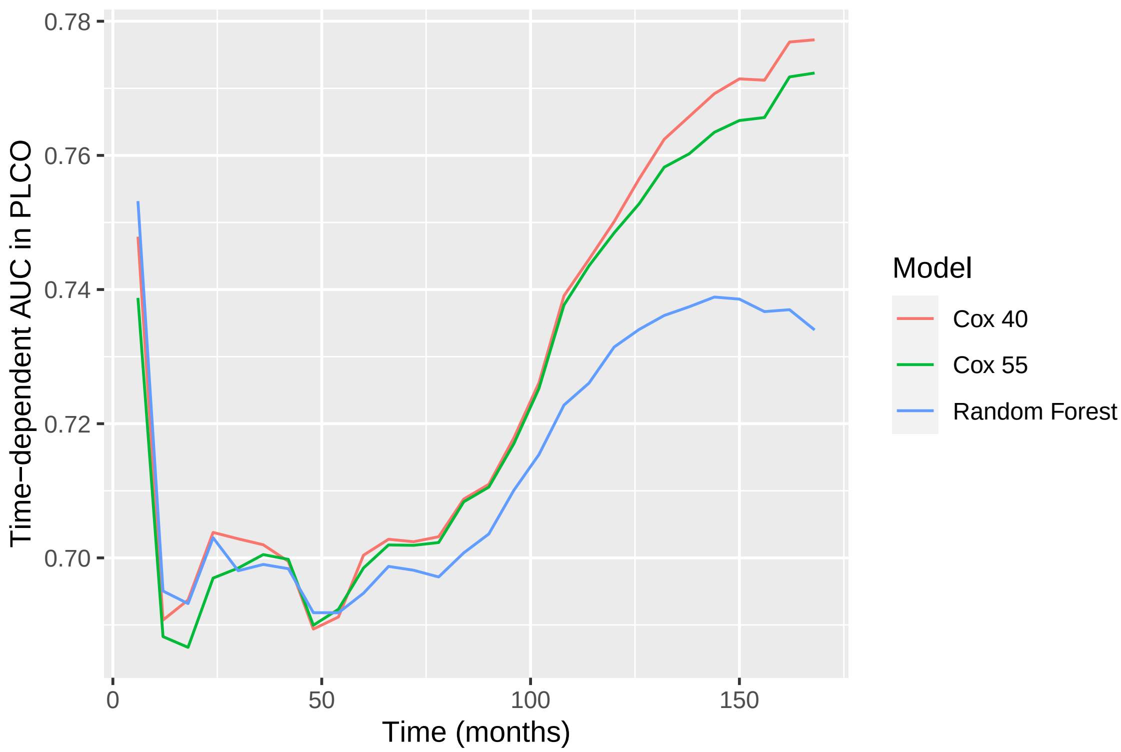

```
subperf <- data.frame("Model" = c("Cox 55", "Cox 40", "Random Forest"),
                      "Year5" = round(c(cox55_plco$AppCindex$coxph[10],
                                        cox40_plco$AppCindex$coxph[10],
                                        forest_plco$AppCindex$rfsrc[10]), digits=3),
                      "Year10" = round(c(cox55_plco$AppCindex$coxph[20],
                                         cox40_plco$AppCindex$coxph[20],
                                         forest_plco$AppCindex$rfsrc[20]), digits=3),
                      "Year14" = round(c(cox55_plco$AppCindex$coxph[28],
                                         cox40_plco$AppCindex$coxph[28],
                                         forest_plco$AppCindex$rfsrc[28]), digits=3))

kable(subperf, caption = "Time-dependent C-index in PLCO")
```

Based on C-index, our strongest performing model is the Cox model fit to men ages 40+. Now we look at time-dependent AUC:

```r
cox40predict <-
  predict(object = cox_40,
          newdata = plco_clean,
          type = "lp")
cox55predict <-
  predict(object = cox_55,
          newdata = plco_clean,
          type = "lp")
rforestpredict <-
  predict(object=rforest_40,
          newdata = plco_nf)

cox40roc <-
  timeROC(
    T = plco_clean$permth_exm,
    delta = plco_clean$mortstat,
    marker = cox40predict,
    cause = 1,
    weighting = "marginal",
    times = seq(0, 168, by = 6)
  )
cox55roc <-
  timeROC(
    T = plco_clean$permth_exm,
    delta = plco_clean$mortstat,
    marker = cox55predict,
    cause = 1,
    weighting = "marginal",
    times = seq(0, 168, by = 6)
  )
rforestroc <-
  timeROC(
    T = plco_nf$permth_exm,
    delta = plco_nf$mortstat,
    marker = rforestpredict$predicted,
    cause = 1,
    weighting = "marginal",
    times = seq(0, 168, by = 6)
  )

aucdat <-
  data.frame("Time" = rep(seq(6, 168, by = 6), 3),
             "Model" = c(rep("Cox 55", 28), rep("Cox 40", 28),
                                     rep("Random Forest", 28)),
             "AUC" = c(cox55roc$AUC[-1], cox40roc$AUC[-1],
                       rforestroc$AUC[-1]))

aucplot <-
  ggplot(data = aucdat, aes(x = Time, y = AUC, group = Model, color=Model)) +
  geom_line() + xlab("Time (months)") + ylab("Time-dependent AUC in PLCO")

aucplot
```

Table 2: Time-Dependent AUC in PLCO

Model	Year5	Year10	Year14
Cox 55	0.698	0.748	0.772
Cox 40	0.700	0.750	0.777
Random Forest	0.695	0.731	0.734




```
auc_table <- data.frame("Model" = c("Cox 55", "Cox 40", "Random Forest"),
                        "Year5" = round(c(cox55roc$AUC[11],
                                          cox40roc$AUC[11],
                                          rforestroc$AUC[11]), digits=3),
                        "Year10" = round(c(cox55roc$AUC[21],
                                           cox40roc$AUC[21],
                                           rforestroc$AUC[21]), digits=3),
                        "Year14" = round(c(cox55roc$AUC[29],
                                           cox40roc$AUC[29],
                                           rforestroc$AUC[29]), digits=3))

kable(auc_table, caption="Time-Dependent AUC in PLCO")
```

Based on time-dependent AUC, the Cox model fit to men ages 40+ is still the strongest performing model. We will now focus our attention on the Cox ages 40+ model for the rest of our model validation/performance assessment.

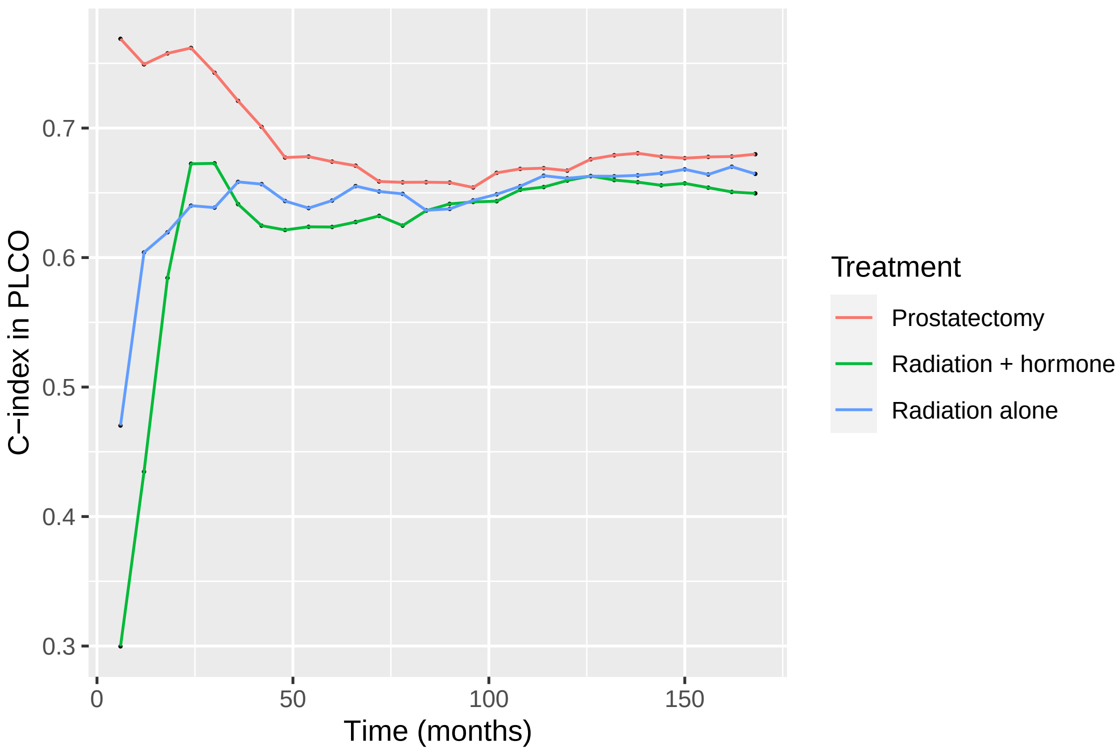
We look at C-index and time-dependent AUC stratified by treatment group:

```
cox40_prostatectomy <- pec::cindex(object=cox_40, formula = Surv(permth_exm, mortstat) ~
                                   age_ctr_40 + diabetic + educ + hypertension +
                                   marital2 + underweight + overweight2 + obese2 +
                                   smoker + stroke + age_ctr_40*diabetic +
```

```
                                     age_ctr_40*educ + age_ctr_40*hypertension +
                                     age_ctr_40*stroke + pc,
                                   data=plco_clean[plco_clean$primary_tx=="Prostatectomy",],
                                   eval.times = times)

cox40_radiation <- pec::cindex(object=cox_40, formula = Surv(permth_exm, mortstat) ~
                                 age_ctr_40 + diabetic + educ + hypertension +
                                 marital2 + underweight + overweight2 + obese2 +
                                 smoker + stroke + age_ctr_40*diabetic +
                                 age_ctr_40*educ + age_ctr_40*hypertension +
                                 age_ctr_40*stroke + pc,
                               data=plco_clean[plco_clean$primary_tx=="Radiation alone",],
                               eval.times = times)

cox40_rtadt <- pec::cindex(object=cox_40, formula = Surv(permth_exm, mortstat) ~
                             age_ctr_40 + diabetic + educ + hypertension +
                             marital2 + underweight + overweight2 + obese2 +
                             smoker + stroke + age_ctr_40*diabetic +
                             age_ctr_40*educ + age_ctr_40*hypertension +
                             age_ctr_40*stroke + pc,
                           data=plco_clean[plco_clean$primary_tx=="Radiation + hormone",],
                           eval.times = times)

valid_perf_trt <- data.frame("Time" = rep(seq(6, 168, by=6), 3),
                             "Treatment" = c(rep("Prostatectomy", 28),
                                             rep("Radiation alone", 28),
                                             rep("Radiation + hormone", 28)),
                             "C" = c(cox40_prostatectomy$AppCindex$coxph,
                                     cox40_radiation$AppCindex$coxph,
                                     cox40_rtadt$AppCindex$coxph))

performance_plot_trt <- ggplot(data=valid_perf_trt, aes(x=Time, y=C)) +
  geom_point(size=0.1) + geom_line(data=valid_perf_trt, aes(group=Treatment,
                                                            color=Treatment)) +
  xlab("Time (months)") + ylab("C-index in PLCO")

performance_plot_trt
```

Table 3: PLCO Time-Dependent C-index, by Treatment

Treatment	Year5	Year10	Year14
Prostatectomy	0.674	0.667	0.680
Radiation alone	0.644	0.661	0.665
Radiation + hormone	0.624	0.659	0.650



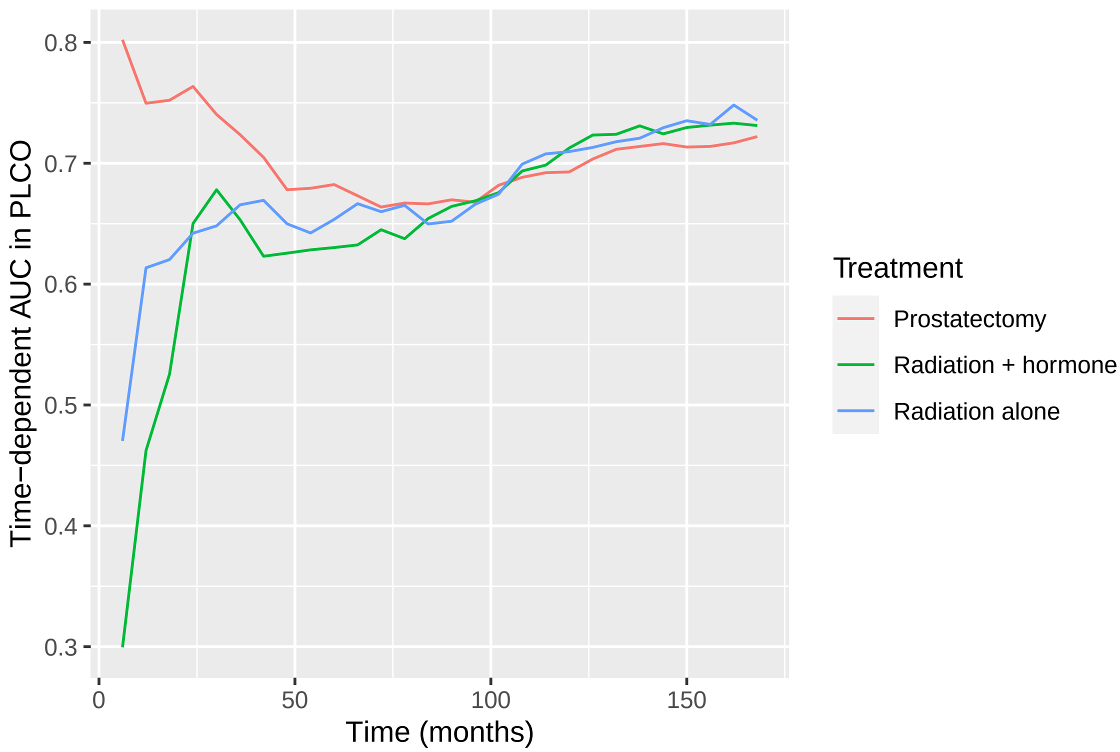

```
subperf_trt <- data.frame("Treatment" = c("Prostatectomy", "Radiation alone",
                                          "Radiation + hormone"),
                          "Year5" = round(c(cox40_prostatectomy$AppCindex$coxph[10],
                                            cox40_radiation$AppCindex$coxph[10],
                                            cox40_rtadt$AppCindex$coxph[10]), digits=3),
                          "Year10" = round(c(cox40_prostatectomy$AppCindex$coxph[20],
                                             cox40_radiation$AppCindex$coxph[20],
                                             cox40_rtadt$AppCindex$coxph[20]), digits=3),
                          "Year14" = round(c(cox40_prostatectomy$AppCindex$coxph[28],
                                             cox40_radiation$AppCindex$coxph[28],
                                             cox40_rtadt$AppCindex$coxph[28]), digits=3))

kable(subperf_trt, caption = "PLCO Time-Dependent C-index, by Treatment")
```

```
cox40predict_prostatectomy <- predict(object=cox_40, newdata =
                                        plco_clean[plco_clean$primary_tx=="Prostatectomy",],
                                      type="lp")

cox40predict_radiation <- predict(object=cox_40, newdata =
                                    plco_clean[plco_clean$primary_tx=="Radiation alone",],
                                  type="lp")
```

```r
cox40predict_rtadt <- predict(object=cox_40, newdata =
                                plco_clean[plco_clean$primary_tx=="Radiation + hormone",],
                              type="lp")

prostatectomy_roc <- timeROC(T = plco_clean$permth_exm[plco_clean$primary_tx=="Prostatectomy"],
                             delta = plco_clean$mortstat[plco_clean$primary_tx=="Prostatectomy"],
                             marker = cox40predict_prostatectomy, cause = 1,
                             weighting="marginal", times = seq(0, 168, by=6))

radiation_roc <- timeROC(T = plco_clean$permth_exm[plco_clean$primary_tx=="Radiation alone"],
                         delta = plco_clean$mortstat[plco_clean$primary_tx=="Radiation alone"],
                         marker = cox40predict_radiation, cause = 1,
                         weighting="marginal", times = seq(0, 168, by=6))

rtadt_roc <- timeROC(T = plco_clean$permth_exm[plco_clean$primary_tx=="Radiation + hormone"],
                     delta = plco_clean$mortstat[plco_clean$primary_tx=="Radiation + hormone"],
                     marker = cox40predict_rtadt, cause = 1, weighting="marginal",
                     times = seq(0, 168, by=6))

aucdat_trt <- data.frame("Time" = rep(seq(6, 168, by=6), 3),
                         "AUC" = c(prostatectomy_roc$AUC[-1], radiation_roc$AUC[-1],
                                   rtadt_roc$AUC[-1]),
                         "Treatment" = c(rep("Prostatectomy", 28),
                                         rep("Radiation alone", 28),
                                         rep("Radiation + hormone", 28)))

aucplot_trt <- ggplot(data=aucdat_trt, aes(x=Time, y=AUC, group=Treatment,
                                           color=Treatment)) + geom_line() +
  xlab("Time (months)") + ylab("Time-dependent AUC in PLCO")

aucplot_trt
```

Table 4: Time-Dependent AUC in PLCO, by Treatment

Years	Prostatectomy	Radiation	Radiation_hormone
5	0.682	0.654	0.630
10	0.693	0.710	0.713
14	0.722	0.736	0.731




```
auc_table_trt <- data.frame("Years" = c(5, 10, 14),
                            "Prostatectomy" = round(c(prostatectomy_roc$AUC[11],
                                                      prostatectomy_roc$AUC[21],
                                                      prostatectomy_roc$AUC[29]),
                                                    digits = 3),
                            "Radiation" = round(c(radiation_roc$AUC[11],
                                                  radiation_roc$AUC[21],
                                                  radiation_roc$AUC[29]),
                                                digits = 3),
                            "Radiation_hormone" = round(c(rtadt_roc$AUC[11],
                                                          rtadt_roc$AUC[21],
                                                          rtadt_roc$AUC[29]),
                                                        digits = 3))

rownames(auc_table_trt) <- NULL
kable(auc_table_trt, caption="Time-Dependent AUC in PLCO, by Treatment")
```

```
cutTime <- 14*12   #14 yr cutoff point

plco_clean$pc_time_14yr <- ifelse(plco_clean$permth_exm > cutTime, cutTime,
                                  plco_clean$permth_exm)
```

```r
plco_clean$pc_status <- ifelse(plco_clean$permth_exm >= cutTime, 0,
                               plco_clean$mortstat2)
plco_clean$linpred <- cox40predict

dat <- plco_clean[, c("pc_status", "pc_time_14yr", "linpred")]
dat <- dat[complete.cases(dat), ]

eve.ocm <- dat$pc_status == 1
eve.pcsm <- dat$pc_status == 2
eve.cens <- dat$pc_status == 0

fstatus <- rep(0,times = nrow(dat))
fstatus[which(eve.ocm)] <- 1
fstatus[which(eve.pcsm)] <- 2
ftime <- dat$pc_time_14yr

fg_covariates <- dat$linpred

finegray_pc <- crr(ftime = ftime, fstatus = fstatus, cov1 = fg_covariates)
summary(finegray_pc)
## Competing Risks Regression
##
## Call:
## crr(ftime = ftime, fstatus = fstatus, cov1 = fg_covariates)
##
##                  coef exp(coef) se(coef)    z p-value
## fg_covariates1 1.02      2.76   0.0346 29.3       0
##
##                  exp(coef) exp(-coef) 2.5% 97.5%
## fg_covariates1      2.76      0.362 2.58  2.95
##
## Num. cases = 8220
## Pseudo Log-likelihood = -17432
## Pseudo likelihood ratio test = 936  on 1 df,

finegray_predictions <- predict(finegray_pc, cov1 = fg_covariates)

predict_5 <- finegray_predictions[finegray_predictions[,1]==60,-1]
predict_10 <- finegray_predictions[finegray_predictions[,1]==120,-1]
predict_14 <- finegray_predictions[finegray_predictions[,1]==167,-1]

group_5 <- case_when(
  predict_5 < 0.2 ~ 0.1,
  predict_5 >= 0.2 & predict_5 < 0.4 ~ 0.3,
  predict_5 >= 0.4 & predict_5 < 0.6 ~ 0.5,
  predict_5 >= 0.6 & predict_5 < 0.8 ~ 0.7,
  predict_5 >= 0.8 ~ 0.9
)

group_10 <- case_when(
  predict_10 < 0.2 ~ 0.1,
  predict_10 >= 0.2 & predict_10 < 0.4 ~ 0.3,
  predict_10 >= 0.4 & predict_10 < 0.6 ~ 0.5,
```

```r
  predict_10 >= 0.6 & predict_10 < 0.8 ~ 0.7,
  predict_10 >= 0.8 ~ 0.9
)

group_14 <- case_when(
  predict_14 < 0.2 ~ 0.1,
  predict_14 >= 0.2 & predict_14 < 0.4 ~ 0.3,
  predict_14 >= 0.4 & predict_14 < 0.6 ~ 0.5,
  predict_14 >= 0.6 & predict_14 < 0.8 ~ 0.7,
  predict_14 >= 0.8 ~ 0.9
)

cuminc_5 <- cuminc(ftime, fstatus, group = group_5)
cuminc_10 <- cuminc(ftime, fstatus, group = group_10)
cuminc_14 <- cuminc(ftime, fstatus, group = group_14)

calibration5_data <- data.frame("Year" = rep("5 Years", 3),
                                "Group" = c(0.1, 0.3, 0.5),
                                "Est" = c(cuminc_5$`0.1 1`$est[cuminc_5$`0.1 1`$time==60][1],
                                          cuminc_5$`0.3 1`$est[cuminc_5$`0.3 1`$time==59][1],
                                          cuminc_5$`0.5 1`$est[cuminc_5$`0.5 1`$time==57][1]),
                                "Var" = c(cuminc_5$`0.1 1`$var[cuminc_5$`0.1 1`$time==60][1],
                                          cuminc_5$`0.3 1`$var[cuminc_5$`0.3 1`$time==59][1],
                                          cuminc_5$`0.5 1`$var[cuminc_5$`0.5 1`$time==57][1]))

calibration10_data <- data.frame("Year" = rep("10 Years", 4),
                                 "Group" = c(0.1, 0.3, 0.5, 0.7),
                                 "Est" = c(cuminc_10$`0.1 1`$est[cuminc_10$`0.1 1`$time==120][1],
                                           cuminc_10$`0.3 1`$est[cuminc_10$`0.3 1`$time==120][1],
                                           cuminc_10$`0.5 1`$est[cuminc_10$`0.5 1`$time==119][1],
                                           cuminc_10$`0.7 1`$est[cuminc_10$`0.7 1`$time==118][1]),
                                 "Var" = c(cuminc_10$`0.1 1`$var[cuminc_10$`0.1 1`$time==120][1],
                                           cuminc_10$`0.3 1`$var[cuminc_10$`0.3 1`$time==120][1],
                                           cuminc_10$`0.5 1`$var[cuminc_10$`0.5 1`$time==119][1],
                                           cuminc_10$`0.7 1`$var[cuminc_10$`0.7 1`$time==118][1]))

calibration14_data <- data.frame("Year" = rep("10 Years", 5),
                                 "Group" = c(0.1, 0.3, 0.5, 0.7, 0.9),
                                 "Est" = c(cuminc_14$`0.1 1`$est[cuminc_14$`0.1 1`$time==168][1],
                                           cuminc_14$`0.3 1`$est[cuminc_14$`0.3 1`$time==168][1],
                                           cuminc_14$`0.5 1`$est[cuminc_14$`0.5 1`$time==168][1],
                                           cuminc_14$`0.7 1`$est[cuminc_14$`0.7 1`$time==168][1],
                                           cuminc_14$`0.9 1`$est[cuminc_14$`0.9 1`$time==157][1]),
                                 "Var" = c(cuminc_14$`0.1 1`$var[cuminc_14$`0.1 1`$time==168][1],
                                           cuminc_14$`0.3 1`$var[cuminc_14$`0.3 1`$time==168][1],
                                           cuminc_14$`0.5 1`$var[cuminc_14$`0.5 1`$time==168][1],
                                           cuminc_14$`0.7 1`$var[cuminc_14$`0.7 1`$time==168][1],
                                           cuminc_14$`0.9 1`$var[cuminc_14$`0.9 1`$time==157][1]))

calibration5_data$Upper <- calibration5_data$Est + 1.96*sqrt(calibration5_data$Var)
calibration10_data$Upper <- calibration10_data$Est + 1.96*sqrt(calibration10_data$Var)
calibration14_data$Upper <- calibration14_data$Est + 1.96*sqrt(calibration14_data$Var)
calibration5_data$Lower <- calibration5_data$Est - 1.96*sqrt(calibration5_data$Var)
```

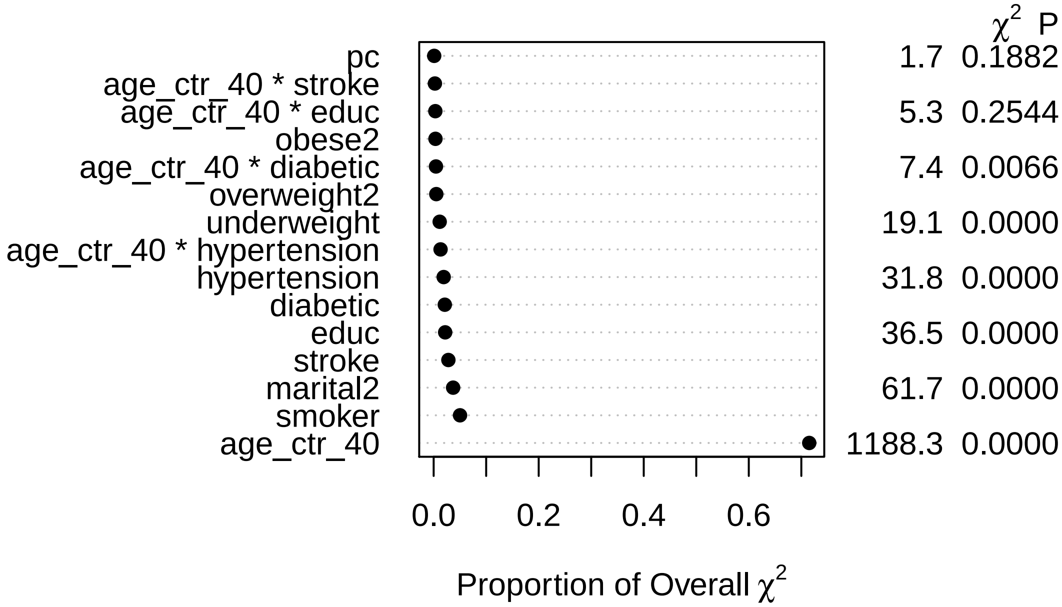
```r
calibration10_data$Lower <- calibration10_data$Est - 1.96*sqrt(calibration10_data$Var)
calibration14_data$Lower <- calibration14_data$Est - 1.96*sqrt(calibration14_data$Var)

calibration5_data$Upper[calibration5_data$Upper>1] <- 1
calibration5_data$Lower[calibration5_data$Lower<0] <- 0
calibration10_data$Upper[calibration10_data$Upper>1] <- 1
calibration10_data$Lower[calibration10_data$Lower<0] <- 0
calibration14_data$Upper[calibration14_data$Upper>1] <- 1
calibration14_data$Lower[calibration14_data$Lower<0] <- 0

myplot <- ggplot() + geom_line(data = calibration5_data, aes(x=Group, y=Est),
                               color = "darkblue") +
  geom_errorbar(data = calibration5_data, aes(x = Group, ymin = Lower,
                                              ymax = Upper, width = 0.05),
                color = "darkblue") + geom_line(data = calibration10_data,
                                                aes(x=Group, y=Est),
                                                color = "mediumpurple4") +
  geom_errorbar(data = calibration10_data, aes(x = Group, ymin = Lower,
                                               ymax = Upper, width = 0.05),
                color = "mediumpurple4") + geom_line(data = calibration14_data,
                                                     aes(x=Group, y=Est),
                                                     color = "gray74") +
  geom_errorbar(data = calibration14_data, aes(x = Group, ymin = Lower,
                                               ymax = Upper), color = "gray74",
                width = 0.05) + scale_x_continuous(name = "Predicted Risk of OCM",
                                                   limits = c(0, 1),
                                                   breaks = seq(0, 1, by=0.2)) +
  scale_y_continuous(name = "Observed Risk of OCM", limits = c(0, 1),
                     breaks = seq(0, 1.2, by=0.2)) + theme_bw() +
  geom_abline(slope = 1, intercept = 0)  +
  labs(caption = "Black line: perfect calibration. Blue line: calibration at
      5 years. Purple line: calibration at 10 years. Gray line: calibration
      at 14 years.")

#myplot

#save(list=c("plco_clean", "plco_nf"), file = "~/Box/PLCO Data/Prostate/plco_data_clean.RData")
```

Variable Importance

We assess the variable importance of our Cox model:

```r
load("/Users/ecchase/Desktop/Research/Matt Prostate OCM/PCOtherCause/Data/nhanes_data_clean.RData")
f <-
  cph(
    Surv(permth_exm, mortstat) ~ age_ctr_40 + diabetic + educ + hypertension +
      marital2 + underweight + overweight2 + obese2 + smoker + stroke +
      age_ctr_40 * diabetic + age_ctr_40 * educ + age_ctr_40 * hypertension +
      age_ctr_40 * stroke + pc,
    data = mydata[mydata$inmodel4 == 1, ],
    x = TRUE,
    y = TRUE
  )
```

```r
plot(anova(f), what='proportion chisq')
```

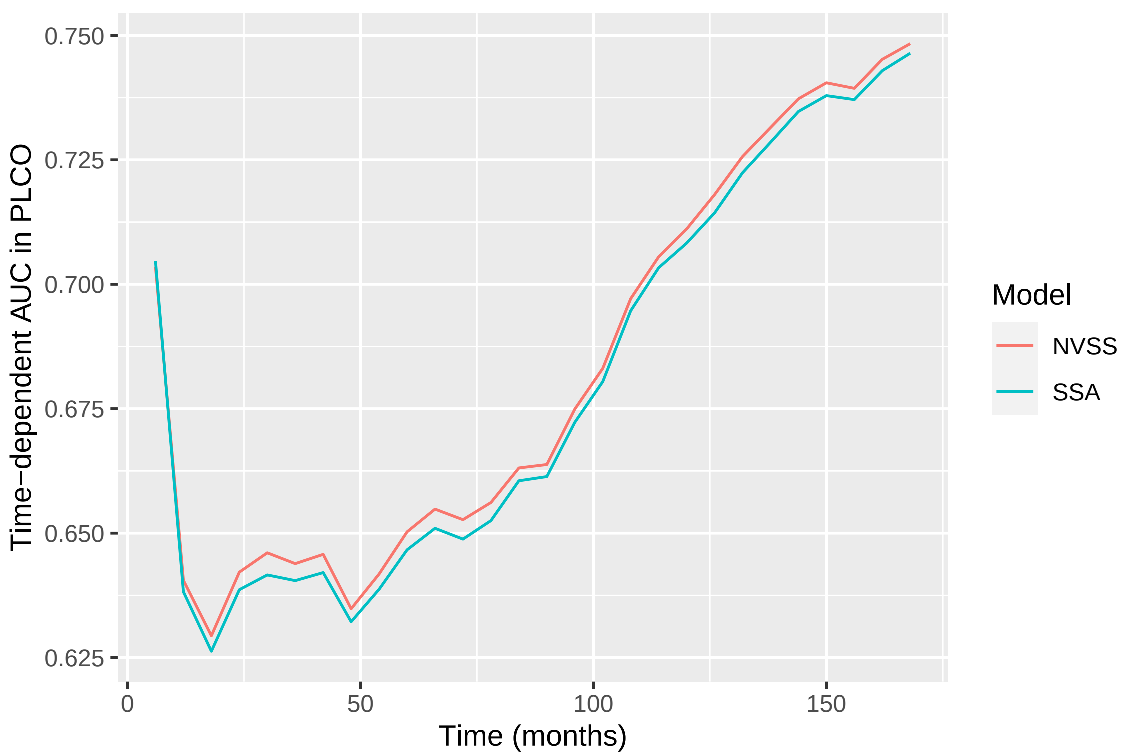


Comparison to SSA and NVSS Life Tables

We compare to the SSA and NVSS life tables:

```r
ssa <- read_csv("/Users/ecchase/Desktop/Research/Matt Prostate OCM/PCOtherCause/Data/ssa_2000.csv")
ssa$age <- seq(0, 119, by = 1)

nvss_black <- read_csv("/Users/ecchase/Desktop/Research/Matt Prostate OCM/PCOtherCause/Data/nvss_blackma
nvss_white <- read_csv("/Users/ecchase/Desktop/Research/Matt Prostate OCM/PCOtherCause/Data/nvss_whitema
nvss_other <- read_csv("/Users/ecchase/Desktop/Research/Matt Prostate OCM/PCOtherCause/Data/nvss_allmale
nvss_other <- nvss_other[-102,]

nvss <- data.frame("Age" = seq(0, 100, by = 1), "White" = nvss_white$Expectancy, "Black" = nvss_black$Ex

plco_clean$ssa <- NA
plco_clean$nvss <- NA
for (i in 1:nrow(plco_clean)){
  plco_clean$ssa[i] <- ssa$Expectancy[which(ssa$age==plco_clean$age[i])]
  if (plco_clean$race2[i]=="NHW"){
    plco_clean$nvss[i] <- nvss$White[which(nvss$Age==plco_clean$age[i])]
  } else if (plco_clean$race2[i]=="NHB"){
    plco_clean$nvss[i] <- nvss$Black[which(nvss$Age==plco_clean$age[i])]
  } else if (plco_clean$race2[i]=="Other"){
    plco_clean$nvss[i] <- nvss$Other[which(nvss$Age==plco_clean$age[i])]
  }
}

ssa_roc <-
  timeROC(
    T = plco_clean$permth_exm,
    delta = plco_clean$mortstat,
```

```r
    marker = 1/plco_clean$ssa,
    cause = 1,
    weighting = "marginal",
    times = seq(0, 168, by = 6)
  )

nvss_roc <-
  timeROC(
    T = plco_clean$permth_exm,
    delta = plco_clean$mortstat,
    marker = 1/plco_clean$nvss,
    cause = 1,
    weighting = "marginal",
    times = seq(0, 168, by = 6)
  )

aucdat <-
  data.frame("Time" = rep(seq(6, 168, by = 6), 2),
             "Model" = c(rep("SSA", 28), rep("NVSS", 28)),
             "AUC" = c(ssa_roc$AUC[-1], nvss_roc$AUC[-1]))

aucplot <-
  ggplot(data = aucdat, aes(x = Time, y = AUC, group = Model, color=Model)) +
  geom_line() + xlab("Time (months)") + ylab("Time-dependent AUC in PLCO")

aucplot
```



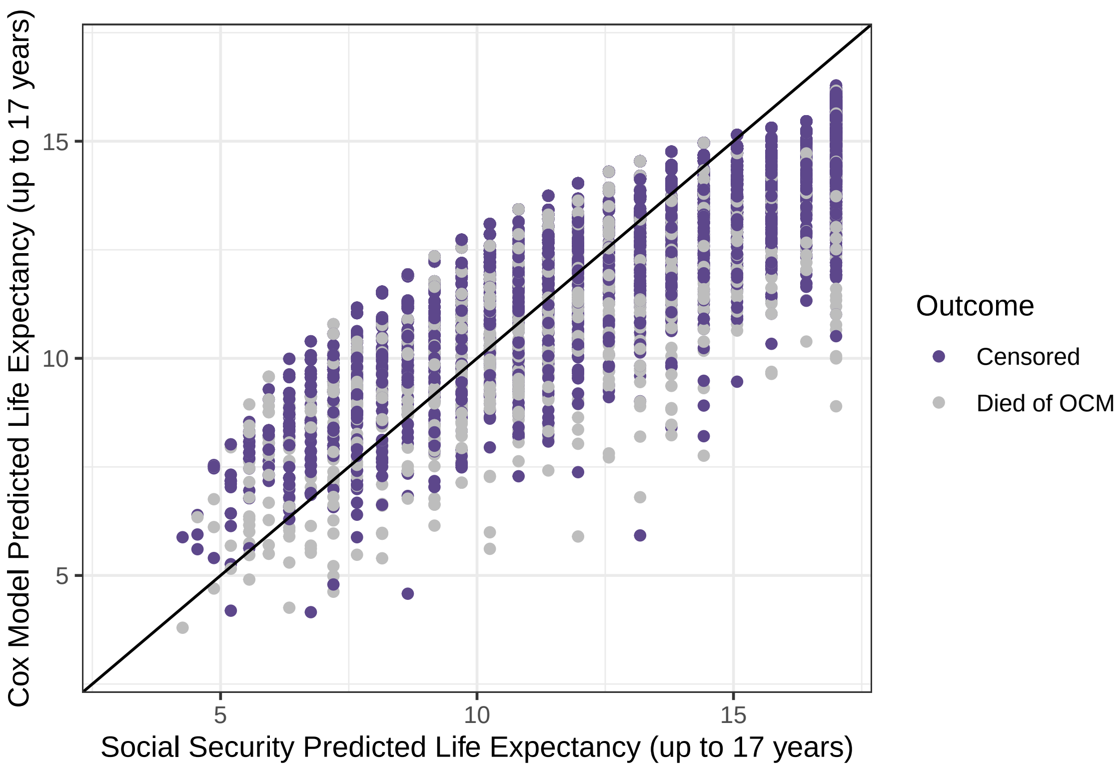

```r
auc_table <- data.frame("Model" = c("SSA", "NVSS"),
                        "Year5" = round(c(ssa_roc$AUC[11],
```

Table 5: Time-Dependent AUC in PLCO

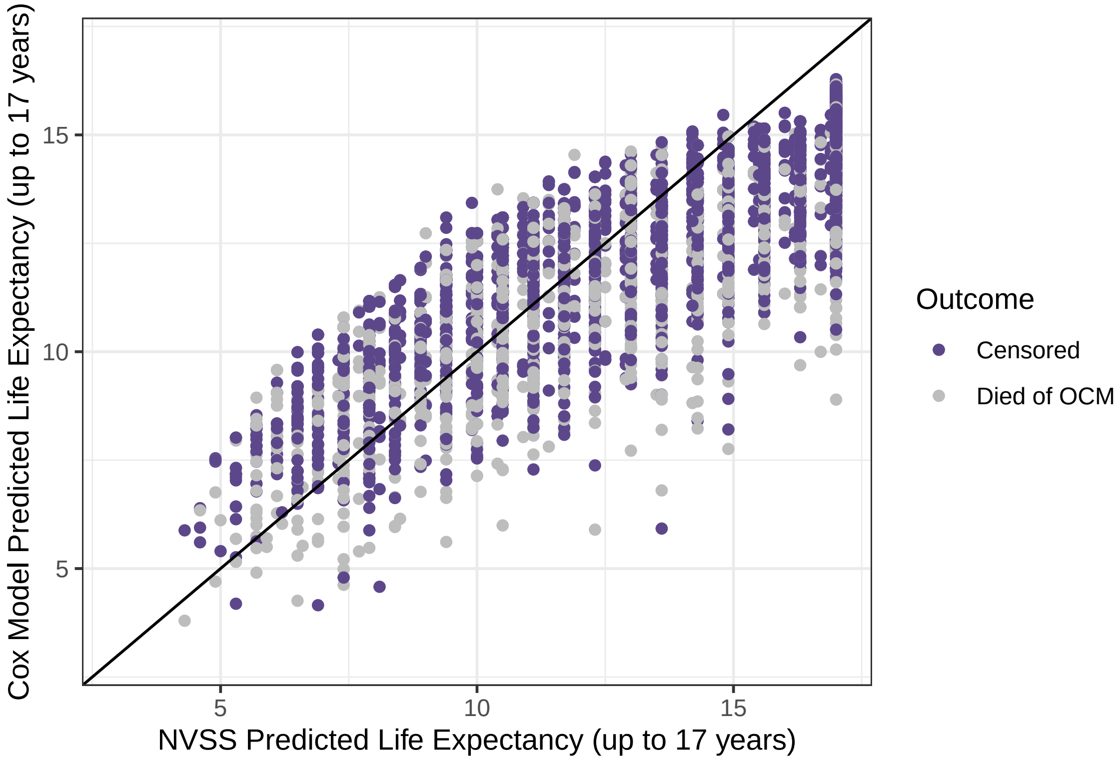
Model	Year5	Year10	Year14
SSA	0.647	0.708	0.746
NVSS	0.650	0.711	0.748

```
                                nvss_roc$AUC[11]), digits=3),
                  "Year10" = round(c(ssa_roc$AUC[21],
                                 nvss_roc$AUC[21]), digits=3),
                  "Year14" = round(c(ssa_roc$AUC[29],
                                 nvss_roc$AUC[29]), digits=3))

kable(auc_table, caption="Time-Dependent AUC in PLCO")
```

```
km_predictions <- survfit(cox_40, plco_clean)
incs2 <- km_predictions$time[2:length(km_predictions$time)] - km_predictions$time[1:(length(km_predicti
rmst2 <- (incs2 %*% (km_predictions$surv[-nrow(km_predictions$surv),]))/12

expectancy_dat <- data.frame("Cox_40" = rmst2[1,], "SSA" = plco_clean$ssa, "NVSS" = plco_clean$nvss, "A

expectancy_dat$SSA[expectancy_dat$SSA > 17] <- 17
expectancy_dat$NVSS[expectancy_dat$NVSS > 17] <- 17

comparison_plot <- ggplot(data=expectancy_dat, aes(x = SSA, y = Cox_40, color = factor(Outcome))) + geom

comparison_plot
```



```
comparison_plot2 <- ggplot(data=expectancy_dat, aes(x = NVSS, y = Cox_40, color = factor(Outcome))) + ge

comparison_plot2
```



PCSM Prediction

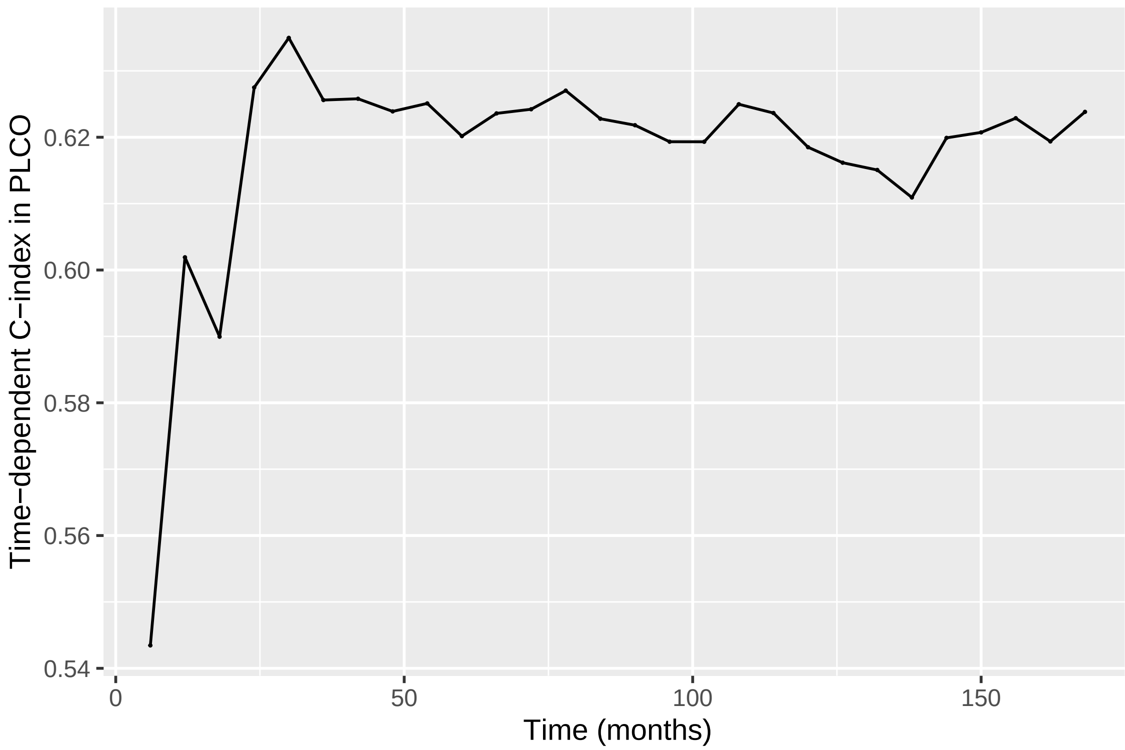
We get the C-index for the linear predictor for PCSM:

```
#Indicator for death from prostate cancer (other cause mortality censored)
plco_clean$mortstat <- case_when(
  plco_clean$mortstat2==2 ~ 1,
  plco_clean$mortstat2==1 | plco_clean$mortstat2==0 ~ 0
)

times <- seq(6, 168, by=6)

cox40_plco_pcsm <- pec::cindex(object=cox_40, formula = Surv(permth_exm, mortstat) ~
                            age_ctr_40 + diabetic + educ + hypertension + marital2 +
                            underweight + overweight2 + obese2 + smoker + stroke +
                            age_ctr_40*diabetic + age_ctr_40*educ +
                            age_ctr_40*hypertension + age_ctr_40*stroke + pc,
                          data=plco_clean, eval.times = times)

valid_perf <- data.frame("Time" = seq(6, 168, by=6),
                         "C" = cox40_plco_pcsm$AppCindex$coxph)

performance_plot <- ggplot(data=valid_perf, aes(x=Time, y=C)) + geom_point(size=0.1) +
  geom_line(data=valid_perf) + xlab("Time (months)") + ylab("Time-dependent C-index in PLCO")
```

Table 6: Time-dependent C-index for PCSM

Model	Year5	Year10	Year14
Cox 40	0.62	0.619	0.624

```
performance_plot
```



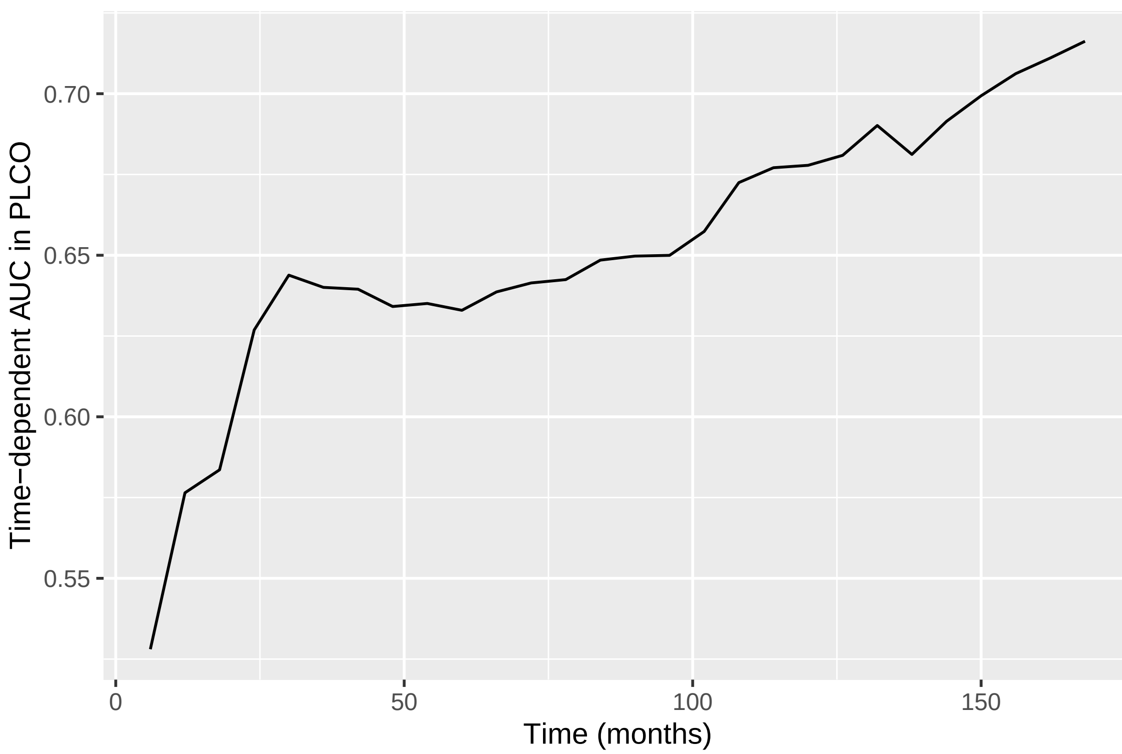

```
subperf <- data.frame("Model" = "Cox 40",
                      "Year5" = round(cox40_plco_pcsm$AppCindex$coxph[10], digits=3),
                      "Year10" = round(cox40_plco_pcsm$AppCindex$coxph[20], digits=3),
                      "Year14" = round(cox40_plco_pcsm$AppCindex$coxph[28], digits=3))

kable(subperf, caption = "Time-dependent C-index for PCSM")
```

```
cox40predict <-
  predict(object = cox_40,
          newdata = plco_clean,
          type = "lp")

cox40roc <-
  timeROC(
    T = plco_clean$permth_exm,
    delta = plco_clean$mortstat,
    marker = cox40predict,
    cause = 1,
    weighting = "marginal",
    times = seq(0, 168, by = 6)
```

Table 7: Time-Dependent AUC for PCSM

	Model	Year5	Year10	Year14
t=60	Cox 40	0.633	0.678	0.716

```
  )

aucdat <-
  data.frame("Time" = seq(6, 168, by = 6),
             "AUC" = cox40roc$AUC[-1])

aucplot <-
  ggplot(data = aucdat, aes(x = Time, y = AUC)) + geom_line() +
  xlab("Time (months)") + ylab("Time-dependent AUC in PLCO")

aucplot
```




```
auc_table <- data.frame("Model" = "Cox 40",
                        "Year5" = round(cox40roc$AUC[11], digits=3),
                        "Year10" = round(cox40roc$AUC[21], digits=3),
                        "Year14" = round(cox40roc$AUC[29], digits=3))

kable(auc_table, caption="Time-Dependent AUC for PCSM")
```