Distributions: Clopidogrel, Warfarin, Statin

February 15, 2016

Contents

List of Tables

List of Figures

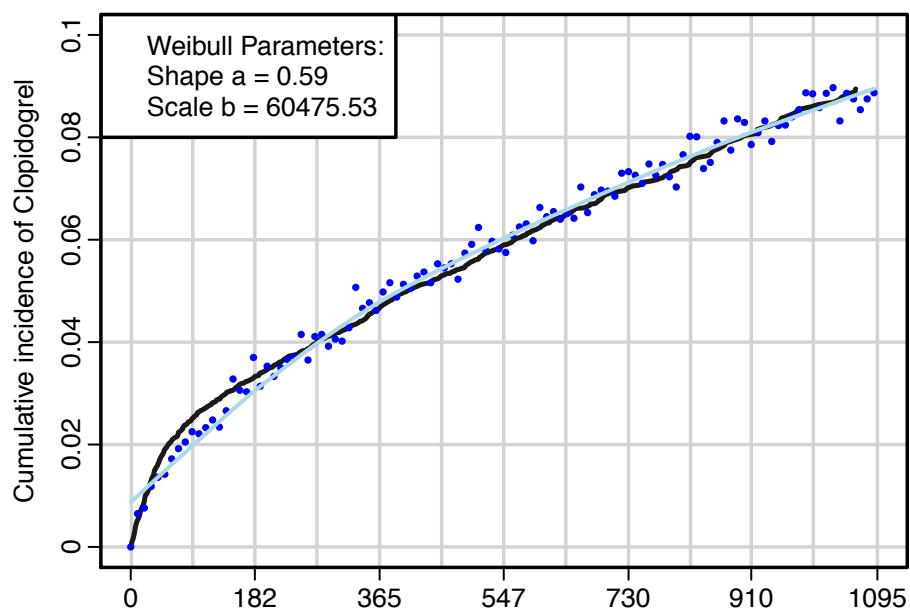Figure 1: Kaplan Meier curves and Weibull distribution approximation: time to antiplatelet medication exposure

Statin Naive Cohort (baseline, event/n = 2873/19497)
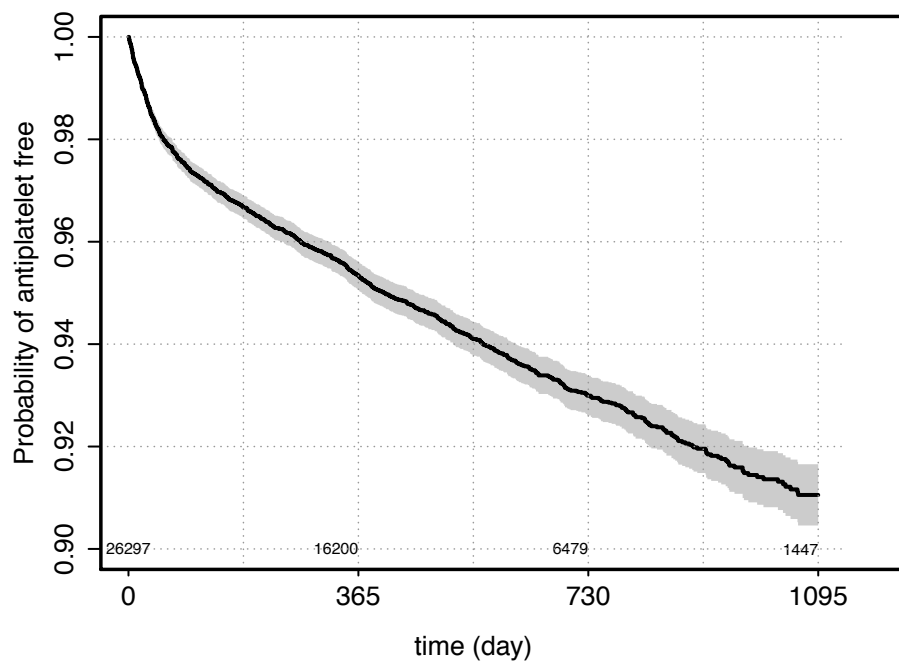

Figure 2: Kaplan Meier curves and Weibull distribution approximation: time to statin

Statin Naive Cohort (baseline, event/n = 1293/19497)
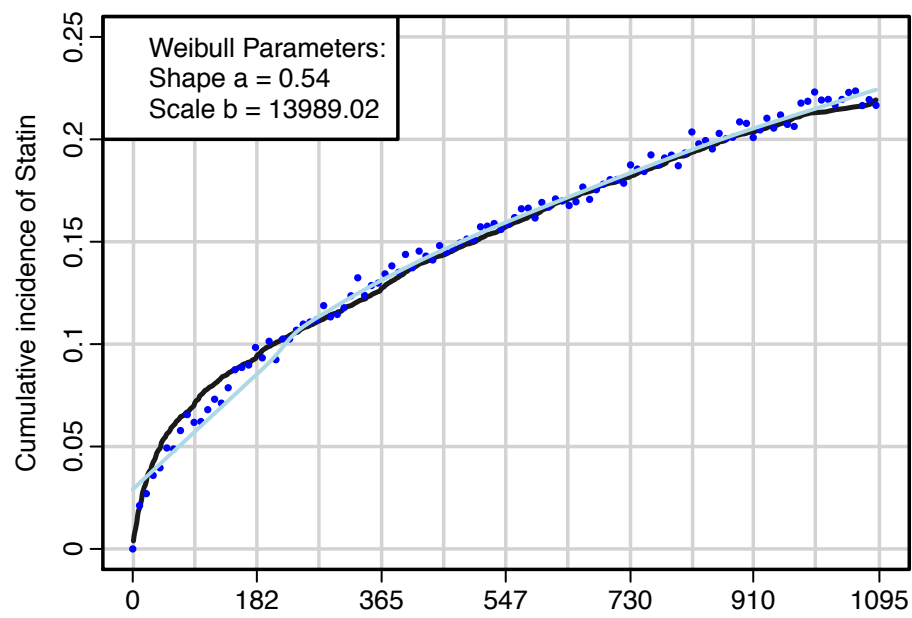

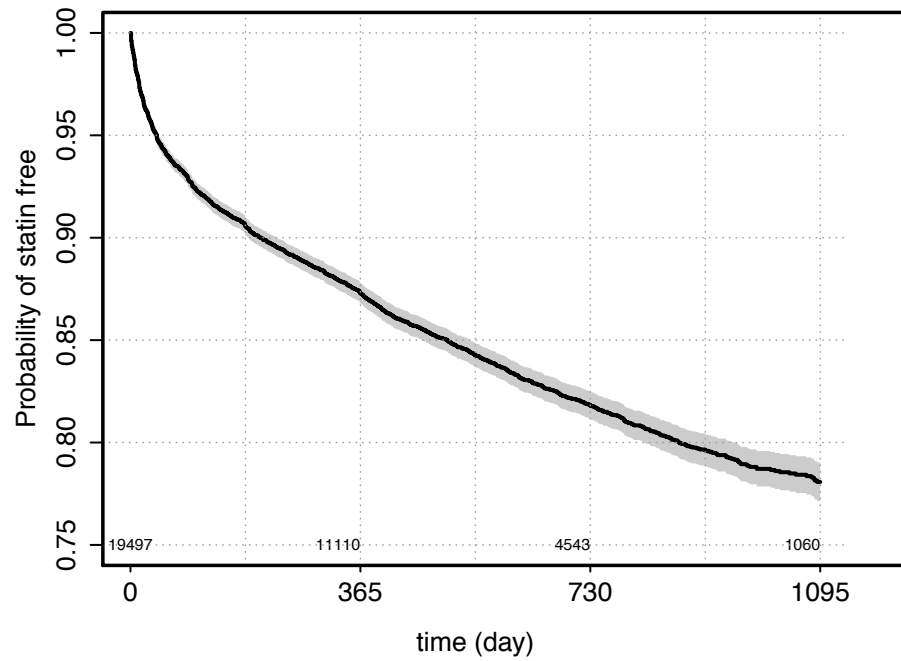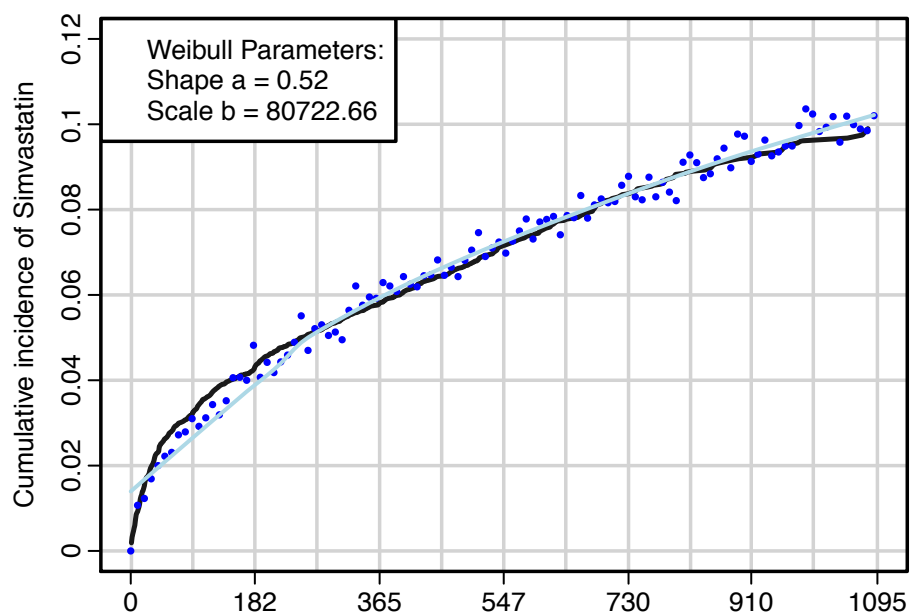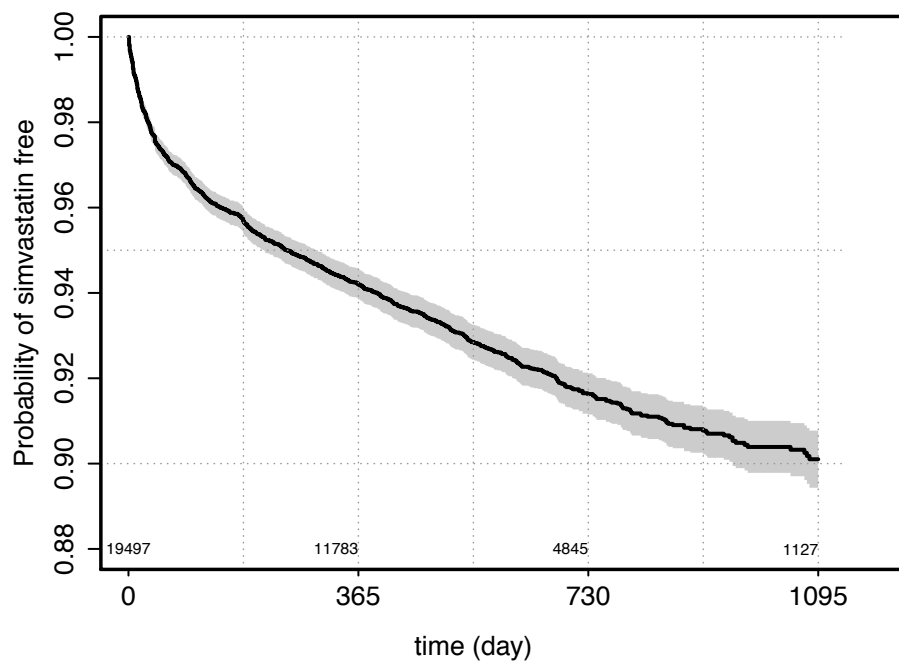Figure 3: Kaplan Meier curves and Weibull distribution approximation: time to simvastatin

4

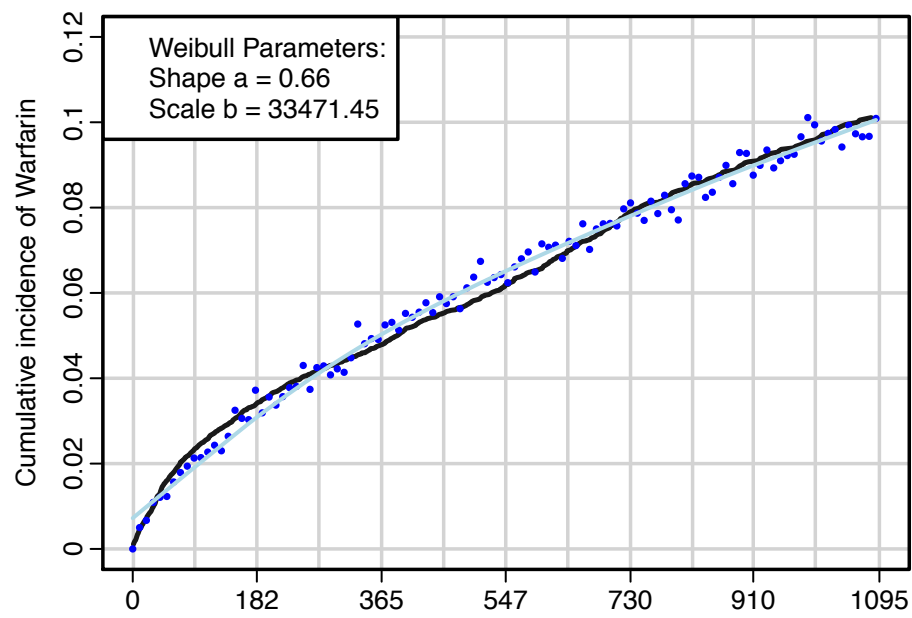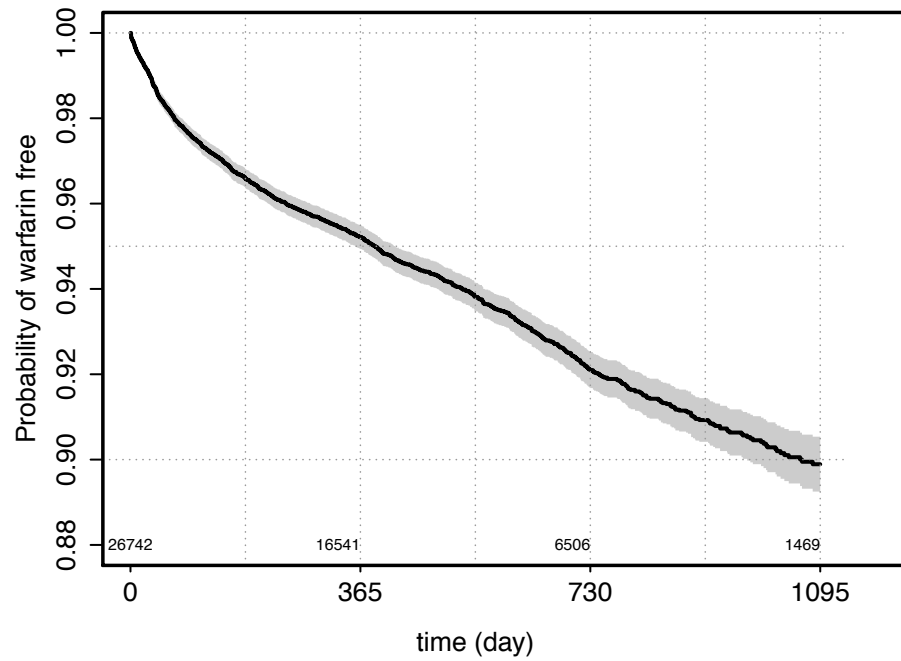## Warfarin Naive Cohort (baseline, event/n = 1594/26742)

Figure 4: Kaplan Meier curves and Weibull distribution approximation: time to warfarin

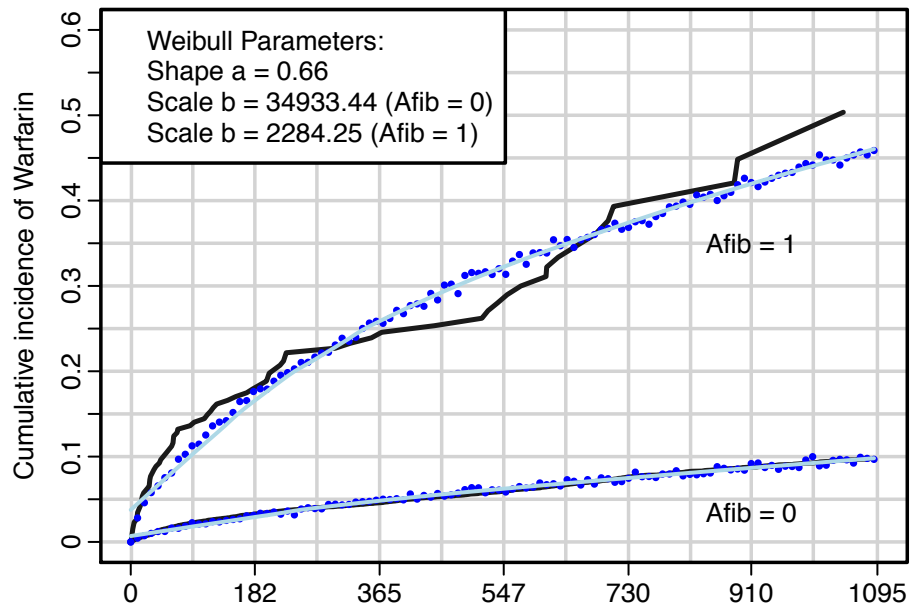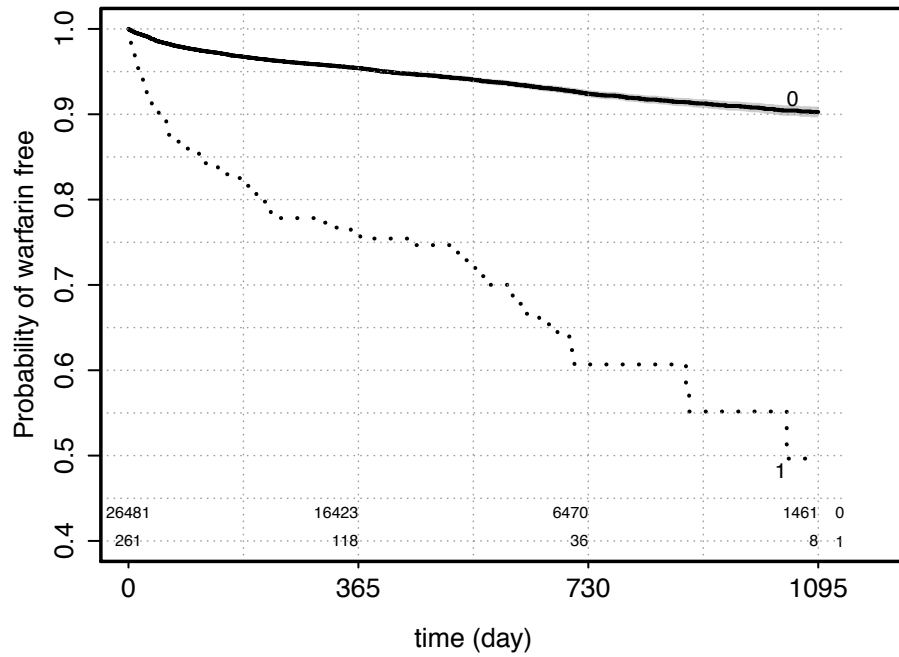## Warfarin Naive Cohort Stratified by BL Afib

Figure 5: Kaplan Meier curves and Weibull distribution approximation: time to warfarin

Source Code for this Report

```
% Usage: R CMD Sweave template3.Rnw = Sweave template3
%         rubber -d template3 or pdflatex template3
% To get .R file: R CMD Stangle template3.Rnw = Stangle template3
\documentclass{article}
\usepackage{relsize,setspace,graphicx,fancyhdr,longtable,lscape,booktabs,color,url, caption}  % used by latex(describe( ))
\usepackage{needspace}
\usepackage{url}                  % used in bibliography
\usepackage[superscript,nomove]{cite} % use if \cite is used and superscripts wanted
% Remove nomove if you want superscripts after punctuation in citations
\usepackage{lscape}               % for landscape mode tables
\usepackage{moreverb}             %for verbatimtabinput
\usepackage[utf8]{inputenc}
\usepackage{moreverb}             %for verbatimtabinput
\usepackage[pdftex]{lscape}       %allows tables to be landscape
% \usepackage{hyperref}           %hyperlinks to sections/tables
\textwidth 6.75in                 % set dimensions before fancyhdr
\textheight 9.25in
\topmargin -.875in
\oddsidemargin -.125in
\evensidemargin -.125in
\usepackage{fancyhdr}             % this and next line are for fancy headers/footers
%\pagestyle{fancy}
\newcommand{\bc}{\begin{center}}  % abbreviate
\newcommand{\ec}{\end{center}}
\newcommand{\code}[1]{{\smaller\texttt{#1}}}
\newcommand{\R}{{\normalfont\textsf{R}}{}}
\captionsetup{width=15cm}
\usepackage{Sweavel}


\SweaveOpts{keep.source=TRUE}
% To produce both postscript and pdf graphics, remove the eps and pdf
% parameters in the next line.  Set default plot size to 6x4 in.
\SweaveOpts{prefix.string=graphics/plot, eps = FALSE, pdf = TRUE}
\SweaveOpts{width=6, height=4}

\title{\textbf{Distributions:  Clopidogrel, Warfarin, Statin}}
% \author{Yaping Shi, MS and Jonathan Schildcrout, PhD\\\smaller Department of Biostatistics\\\smaller Vanderbilt University School of Medicine}

\begin{document}
\SweaveOpts{concordance=FALSE}
\maketitle
\tableofcontents
\listoftables
\listoffigures     % not used unless figure environments used
\clearpage
%\rhead{\scriptsize {\em Health Literacy Project \\
%  \today}

<<echo=F>>=
spar <- function(mar=c(3.25+bot-.45*multi,3.5+left,.5+top+.25*multi,.5+rt),
                 lwd = if(multi)1 else 1.75,
                 mgp = if(multi) c(1.5, .365, 0) else c(2.4-.4, 0.475, 0),
                 tcl = if(multi)-0.25 else -0.4,
                 bot=0, left=0, top=0, rt=0, ps=14,
                 mfrow=NULL, ...)
  {
    multi <- length(mfrow) > 0
    par(mar=mar, lwd=lwd, mgp=mgp, tcl=tcl, ...)
    if(multi) par(mfrow=mfrow)
  }
options(SweaveHooks=list(fig=spar)) # run spar() before every plot
options(prompt=' ',continue=' ')  # remove prompt characters at start of lines
require(rms)
@

<<echo=FALSE,results=hide>>=
plot.weib <- function(med='Clopidogrel',XLIM=1095,YLIM=0.2,title='',km.ff,risk.weib=risk.weib,inc=0.1,weib.fit=ff) {

plot(km.ff$time[km.ff$time<=XLIM], (1-km.ff$surv[km.ff$time<=XLIM]),
     xlim=c(0, XLIM), ylim=c(0, YLIM), xlab='',
     ylab=paste("Cumulative incidence of ", med, sep=""), type="l", axes=FALSE, cex=2)
abline(v=c(1, 91, 182, 273, 365, 456, 547, 638, 730, 821, 910, 1001, 1095),
       h=seq(0, 1, inc), col="lightgray")

lines(km.ff$time[km.ff$time<=XLIM], (1-km.ff$surv[km.ff$time<=XLIM]),
      col=grey(0.1), lty=1, lwd=2.5) # KM curve
# lines(seq(0,1095,10), risk.weib, col='blue', lty=3, lwd=2.5) # psm-Weibull
points(seq(0,1095,10), risk.weib, col='blue', pch=20,cex=0.5)

tt <- seq(0,1095,10)
lo <- loess(risk.weib ~ tt)
lines(x=tt,y=lowess(predict(lo),f=1/3)$y,col='lightblue',lwd=2) # lowess smoother

a <- round(1/ff$scale,2)
b <- round(exp(ff$linear.predictors[1]),2)

legend('topleft',
       legend=paste0(c("Weibull Parameters:","Shape a = ","Scale b = "),c('',a,b)),
       cex=1,bty='o',bg='white')


axis(1,at=c(0,182,365,547,730,910,1095),labels=c(0,182,365,547,730,910,1095))
axis(2,at=seq(0,1,inc),labels=seq(0,1,inc))
mtext(title, side=3, line=1, cex=1)
box()
}
@

<<echo=FALSE>>=
library(rms)
setwd("/home/shiy3/Projects/RIGHT")
# load("RIGHT_data.Rdata")
load("RIGHT_data10272014.Rdata")
# load('WCS_KM_Dist_Generation_Data02122016.Rdata')

# impl.0: merged data
# impl: original demo data
```

```
# impl.bl: baseline data after merging

impl.0$time.since.nearest.evt[impl.0$time.since.nearest.evt==Inf] <- 9999
impl.0$time.since.nearest.evt[is.na(impl.0$time.since.nearest.evt)] <- 9999
impl.bl$time.since.nearest.evt[impl.bl$time.since.nearest.evt==Inf] <- 9999
impl.bl$time.since.nearest.evt[is.na(impl.bl$time.since.nearest.evt)] <- 9999


# data with last observation
first <- !duplicated(impl.0$id)
last <- c(first[2:length(first)], TRUE)
impl.last <- impl.0[last,]
impl.last$all <- "All"


# Antiplt cohort: clop. naive at baseline
antiplt.dat <- subset(impl.0, antiplt.popn==1)
first <- !duplicated(antiplt.dat$id)
last <- c(first[2:length(first)], TRUE)
antiplt.dat.bl <- antiplt.dat[first,] # baseline data
antiplt.dat.last <- antiplt.dat[last,] # last observed data
# antiplt.status
antiplt.dat.bl$race2 <- relevel(factor(antiplt.dat.bl$race2), ref="White")
antiplt.dat.bl$hgba1c.adj <- antiplt.dat.bl$hgba1c
antiplt.dat.bl$hgba1c.adj[antiplt.dat.bl$hgba1c.test==0] <- 5+rnorm(sum(antiplt.dat.bl$hgba1c.test==0),0,0.5)
antiplt.dat.bl$hgba1c.test <- factor(antiplt.dat.bl$hgba1c.test)

antiplt.dat.bl$ldl.test <- factor(ifelse(is.na(antiplt.dat.bl$ldl),0,1))
antiplt.dat.bl$ldl[antiplt.dat.bl$ldl.test=='0'] <- 100
antiplt.dat.bl$ldl.adj <- antiplt.dat.bl$ldl
antiplt.dat.bl$ldl.adj[antiplt.dat.bl$ldl.test=='0'] <- 100+rnorm(sum(antiplt.dat.bl$ldl.test=='0'),0,5)


antiplt.dat.bl$hdl.test <- factor(ifelse(is.na(antiplt.dat.bl$hdl),0,1))
antiplt.dat.bl$hdl[antiplt.dat.bl$hdl.test=='0'] <- 50
antiplt.dat.bl$hdl.adj <- antiplt.dat.bl$hdl
antiplt.dat.bl$hdl.adj[antiplt.dat.bl$hdl.test=='0']   <- 50+rnorm(sum(antiplt.dat.bl$hdl.test=='0'),0,5)

antiplt.dat.bl$triglycerides.test <- factor(ifelse(is.na(antiplt.dat.bl$triglycerides),0,1))
antiplt.dat.bl$triglycerides[antiplt.dat.bl$triglycerides.test=='0'] <- 110
antiplt.dat.bl$triglycerides.adj <- antiplt.dat.bl$triglycerides
antiplt.dat.bl$triglycerides.adj[antiplt.dat.bl$triglycerides.test=='0'] <- 110+rnorm(sum(antiplt.dat.bl$triglycerides.test=='0'),0,5)

antiplt.dat.bl$lipids.test <- apply(cbind(antiplt.dat.bl$triglycerides.test,antiplt.dat.bl$ldl.test,antiplt.dat.bl$hdl.test), 1, max)


# Statin cohort
anystat.dat <- subset(impl.0, anystatin.popn==1)
first <- !duplicated(anystat.dat$id)
last <- c(first[2:length(first)], TRUE)
anystat.dat.bl <- anystat.dat[first,]
anystat.dat.last <- anystat.dat[last,]

anystat.dat.bl$simvastat.status <- as.numeric(anystat.dat.bl$simvastat.since.mh>0)
anystat.dat.bl$simvastat.rel.day <- anystat.dat.bl$simvastat.since.mh
anystat.dat.bl$simvastat.rel.day[anystat.dat.bl$simvastat.rel.day<0] <- anystat.dat.bl$obs.since.mh[anystat.dat.bl$simvastat.rel.day<0]

# table(as.numeric(anystat.dat.bl$anystat.since.mh>0))
# anystat.since.mh <- anystat.dat.bl$anystat.since.mh
# anystat.since.mh[anystat.dat.bl$anystat.since.mh==Inf] <- anystat.dat.bl$obs.since.mh[anystat.dat.bl$anystat.since.mh==Inf]
# range(anystat.since.mh)
# range(anystat.since.mh-anystat.dat.bl$anystat.rel.day)

anystat.dat.bl$race2 <- relevel(factor(anystat.dat.bl$race2), ref="White")
anystat.dat.bl$hgba1c.adj <- anystat.dat.bl$hgba1c
anystat.dat.bl$hgba1c.adj[anystat.dat.bl$hgba1c.test==0] <- 5+rnorm(sum(anystat.dat.bl$hgba1c.test==0),0,0.5)
anystat.dat.bl$hgba1c.test <- factor(anystat.dat.bl$hgba1c.test)

anystat.dat.bl$ldl.test <- factor(ifelse(is.na(anystat.dat.bl$ldl),0,1))
anystat.dat.bl$ldl[anystat.dat.bl$ldl.test=='0'] <- 100
anystat.dat.bl$ldl.adj <- anystat.dat.bl$ldl
anystat.dat.bl$ldl.adj[anystat.dat.bl$ldl.test=='0'] <- 100+rnorm(sum(anystat.dat.bl$ldl.test=='0'),0,5)

anystat.dat.bl$hdl.test <- factor(ifelse(is.na(anystat.dat.bl$hdl),0,1))
anystat.dat.bl$hdl[anystat.dat.bl$hdl.test=='0'] <- 50
anystat.dat.bl$hdl.adj <- anystat.dat.bl$hdl
anystat.dat.bl$hdl.adj[anystat.dat.bl$hdl.test=='0'] <- 50+rnorm(sum(anystat.dat.bl$hdl.test=='0'),0,5)

anystat.dat.bl$triglycerides.test <- factor(ifelse(is.na(anystat.dat.bl$triglycerides),0,1))
anystat.dat.bl$triglycerides[anystat.dat.bl$triglycerides.test=='0'] <- 110
anystat.dat.bl$triglycerides.adj <- anystat.dat.bl$triglycerides
anystat.dat.bl$triglycerides.adj[anystat.dat.bl$triglycerides.test=='0'] <- 110+rnorm(sum(anystat.dat.bl$triglycerides.test=='0'),0,5)
anystat.dat.bl$lipids.test <- apply(cbind(anystat.dat.bl$triglycerides.test,anystat.dat.bl$ldl.test,anystat.dat.bl$hdl.test), 1, max)


warf.dat <- subset(impl.0, warf.popn==1)
first <- !duplicated(warf.dat$id)
last <- c(first[2:length(first)], TRUE)
warf.dat.bl <- warf.dat[first,]
warf.dat.last <- warf.dat[last,]
# warf.status

# save(warf.dat.bl,anystat.dat.bl,antiplt.dat.bl,file='WCS_KM_Dist_Generation_Data02122016.Rdata')

# C0.anystat <- 1-rcorr.cens(predict(fit0.anystat),s0.anystat)[1]

@


%' \clearpage
%' % \begin{landscape}
%' \begin{figure}[!htbp]
%'   \centering
%'   \label{KM.antiplt}
%' <<echo=FALSE,fig=TRUE,width=5,height=4>>=
%' par(mfrow=c(1,1),mar=c(4,4,3,1), cex=0.8)
%' # par(mar=c(4,4,3,1), cex=0.8)
%' s <- with(antiplt.dat.bl, Surv(antiplt.rel.day, antiplt.status))
%' km.v <- npsurv(s ~ 1, data=antiplt.dat.bl)
%' # km.v <- survfit(s ~ 1)
%' # fun <- function(x) 1-x
%'
%' survplot(km.v, n.risk=T, xlim=c(0,1700), ylim=c(0.8,1), time.inc=365, conf="band",
%'         ylab="Probability of antiplatelet free", xlab="time (day)",
```

```
%'          title="Antiplatelet Naive Cohort (baseline)", adj.n.risk=0.5,
%'          levels.only=TRUE, y.n.risk=0.7, cex.n.risk=0.7, lwd=2,
%'          label.curves=F, conf.int=T, col.fill=grey(0.8))
%'
%' abline(v=365*c(0.5,1,1.5,2,2.5,3,3.5,4), lty=3, lwd=0.8, col=grey(0.6))
%' abline(h=seq(0.7,1,0.05), lty=3, lwd=0.8, col=grey(0.6))
%' # table(antiplt.dat.bl$antiplt.status,useNA='ifany')
%' mtext("Antiplatelet Naive Cohort (baseline, event/n = 1449/26297)",side=3,line=2,cex=1)
%' # legend("topright", paste("Log rank ",ifelse(p.v<0.001,"p < 0.001",paste("p = ",round(p.v,3), sep="")),sep=""), pch=NA, cex=0.8, bty="n", horiz=T)
%' box()
%' @
%' \caption{Kaplan Meier curves: time to antiplatelet medication exposure}
%' \end{figure}
%' % \end{landscape}


\clearpage
% \section{Antiplatelet}
<<echo=FALSE,results=hide>>=
dd <- datadist(antiplt.dat.bl)
options(datadist="dd")
# Approximate the survival distribution using Weibull regression
# This is a special case that only intercept was fitted
s <- with(antiplt.dat.bl, Surv(antiplt.rel.day, antiplt.status))
ff <- psm(s ~ 1 , data=antiplt.dat.bl, dist="weibull")
# lp.weib <- predict(ff)
lp.weib <- ff$linear.predictors
scale.weib <- ff$scale

km.ff  <- summary(survfit(s ~ 1, data=antiplt.dat.bl))

## for person 1, generate 10000 observations from a weibull (check parameterization!!!)
# rweibull(10000, shape=1/scale.weib, scale=exp(lp.weib))
## calculate the distribution of predicted risks at 2 years
# set.seed(12345)
# risk.weib.6mon  <- mean(rweibull(10000,shape=1/scale.weib,scale=exp(lp.weib[1]))<182)
# risk.weib.12mon <- mean(rweibull(10000,shape=1/scale.weib,scale=exp(lp.weib[1]))<365)
# risk.weib.18mon <- mean(rweibull(10000,shape=1/scale.weib,scale=exp(lp.weib[1]))<547)
# risk.weib.24mon <- mean(rweibull(10000,shape=1/scale.weib,scale=exp(lp.weib[1]))<730)
# risk.weib.30mon <- mean(rweibull(10000,shape=1/scale.weib,scale=exp(lp.weib[1]))<913)
# risk.weib.36mon <- mean(rweibull(10000,shape=1/scale.weib,scale=exp(lp.weib[1]))<1095)

set.seed(12345)
risk.weib <- risk.km <- NULL
for (tt in seq(0,1095,10))   {
risk.weib[tt/10+1] <- mean( rweibull(10000,shape=1/scale.weib[1],
                                    scale=exp(lp.weib[1])) < tt)
risk.km[tt/10+1] <- 1-rev(km.ff$surv[km.ff$time<tt])[1]
}
@


\clearpage
\begin{figure}[!htbp]
  \centering
  \label{antiplt}
<<echo=FALSE,fig=TRUE,width=5,height=8>>=
par(mfrow=c(2,1),mar=c(4,4,3,1), cex=0.8)

# par(mar=c(4,4,3,1), cex=0.8)
s <- with(antiplt.dat.bl, Surv(antiplt.rel.day, antiplt.status))
km.v <- npsurv(s ~ 1, data=antiplt.dat.bl)
# km.v <- survfit(s ~ 1)
# fun <- function(x) 1-x
survplot(km.v, n.risk=T, xlim=c(0,1095), ylim=c(0.9,1), time.inc=365, conf="band",
        ylab="Probability of antiplatelet free", xlab="time (day)",
        title="Antiplatelet Naive Cohort (baseline)", adj.n.risk=0.5,
        levels.only=TRUE, y.n.risk=0.9, cex.n.risk=0.65, lwd=2,
        label.curves=F, conf.int=T, col.fill=grey(0.8))
abline(v=365*c(0.5,1,1.5,2,2.5,3,3.5,4), lty=3, lwd=0.8, col=grey(0.6))
abline(h=seq(0.7,1,0.02), lty=3, lwd=0.8, col=grey(0.6))
# table(antiplt.dat.bl$antiplt.status,useNA='ifany')
mtext("Antiplatelet Naive Cohort (baseline, event/n = 1449/26297)",side=3,line=2,cex=1)
# legend("topright", paste("Log rank ",ifelse(p.v<0.001,"p < 0.001",paste("p = ",round(p.v,3), sep="")),sep=""), pch=NA, cex=0.8, bty="n", horiz=T)
box()

# check appropriateness of a Weibull model
# surv.km <- log(-log(1 - risk.km[-1]))
# x <- log(seq(0,1095,10)[-1])
# plot(surv.km, x)
# med   <- 'Clopidogrel'
# XLIM  <- 1095
# title <- ''
plot.weib(med='Clopidogrel',XLIM=1095,title='',km.ff=km.ff,
        risk.weib=risk.weib,YLIM=0.1,inc=0.02,weib.fit=ff)
@
\caption{Kaplan Meier curves and Weibull distribution approximation: time to antiplatelet medication exposure}
\end{figure}


\clearpage
% \section{Statin}
<<echo=FALSE, results=hide>>=
dd <- datadist(anystat.dat.bl)
options(datadist="dd")

s <- with(anystat.dat.bl, Surv(anystat.rel.day, anystat.status))
ff <- psm(s ~ 1 , data=anystat.dat.bl, dist="weibull")

lp.weib <- ff$linear.predictors
scale.weib <- ff$scale

km.ff  <- summary(survfit(s ~ 1, data=anystat.dat.bl))

set.seed(12345)
risk.weib <- risk.km <- NULL
for (tt in seq(0,1095,10))   {
risk.weib[tt/10+1] <- mean( rweibull(10000,shape=1/scale.weib[1],
                                    scale=exp(lp.weib[1])) < tt)
risk.km[tt/10+1] <- 1-rev(km.ff$surv[km.ff$time<tt])[1]
}
```

```
@

\clearpage
\begin{figure}[!htbp]
  \centering
  \label{statin}
<<echo=FALSE,fig=TRUE,width=5,height=8>>=
par(mfrow=c(2,1),mar=c(4,4,3,1), cex=0.8)
s <- with(anystat.dat.bl, Surv(anystat.rel.day, anystat.status))
km.v <- npsurv(s ~ 1, data=anystat.dat.bl)
# km.v <- survfit(s ~ 1)
# fun <- function(x) 1-x
survplot(km.v, n.risk=T, xlim=c(0,1095), ylim=c(0.75,1), time.inc=365, conf="band",
         ylab="Probability of statin free", xlab="time (day)",
         title="Statin Naive Cohort (baseline)", adj.n.risk=0.5,
         levels.only=TRUE, y.n.risk=0.75, cex.n.risk=0.65, lwd=2,
         label.curves=F, conf.int=T, col.fill=grey(0.8))
abline(v=365*c(0.5,1,1.5,2,2.5,3,3.5,4), lty=3, lwd=0.8, col=grey(0.6))
abline(h=seq(0.7,1,0.05), lty=3, lwd=0.8, col=grey(0.6))
# table(anystat.dat.bl$anystat.status,useNA='a')
mtext("Statin Naive Cohort (baseline, event/n = 2873/19497)",side=3,line=2,cex=1)
# legend("topright", paste("Log rank ",ifelse(p.v<0.001,"p < 0.001",paste("p = ",round(p.v,3), sep="")),sep=""), pch=NA, cex=0.8, bty="n", horiz=T)
box()
# check appropriateness of a Weibull model
# surv.km <- log(-log(1 - risk.km[-1]))
# x <- log(seq(0,1095,10)[-1])
# plot(surv.km, x)
plot.weib(med='Statin',XLIM=1095,title='',km.ff=km.ff,
         risk.weib=risk.weib,YLIM=0.25,inc=0.05,weib.fit=ff)
@
\caption{Kaplan Meier curves and Weibull distribution approximation: time to statin}
\end{figure}


\clearpage
<<echo=FALSE, results=hide>>=
s <- with(anystat.dat.bl, Surv(simvastat.rel.day, simvastat.status))
ff <- psm(s ~ 1 , data=anystat.dat.bl, dist="weibull")

lp.weib <- ff$linear.predictors
scale.weib <- ff$scale

km.ff  <- summary(survfit(s ~ 1, data=anystat.dat.bl))

set.seed(12345)
risk.weib <- risk.km <- NULL
for (tt in seq(0,1095,10))   {
risk.weib[tt/10+1] <- mean( rweibull(10000,shape=1/scale.weib[1],
                                     scale=exp(lp.weib[1])) < tt)
risk.km[tt/10+1] <- 1-rev(km.ff$surv[km.ff$time<tt])[1]
}
@

\begin{figure}[!htbp]
  \centering
  \label{simva}
<<echo=FALSE,fig=TRUE,width=5,height=8>>=
par(mfrow=c(2,1),mar=c(4,4,3,1), cex=0.8)
s <- with(anystat.dat.bl, Surv(simvastat.rel.day, simvastat.status))
km.v <- npsurv(s ~ 1, data=anystat.dat.bl)
# km.v <- survfit(s ~ 1)
# fun <- function(x) 1-x
survplot(km.v, n.risk=T, xlim=c(0,1095), ylim=c(0.88,1), time.inc=365, conf="band",
         ylab="Probability of simvastatin free", xlab="time (day)",
         title="Statin Naive Cohort (baseline)", adj.n.risk=0.5,
         levels.only=TRUE, y.n.risk=0.88, cex.n.risk=0.65, lwd=2,
         label.curves=F, conf.int=T, col.fill=grey(0.8))
abline(v=365*c(0.5,1,1.5,2,2.5,3,3.5,4), lty=3, lwd=0.8, col=grey(0.6))
abline(h=seq(0.7,1,0.05), lty=3, lwd=0.8, col=grey(0.6))
# table(anystat.dat.bl$simvastat.status,useNA='a')
mtext("Statin Naive Cohort (baseline, event/n = 1293/19497)",side=3,line=2,cex=1)
# legend("topright", paste("Log rank ",ifelse(p.v<0.001,"p < 0.001",paste("p = ",round(p.v,3), sep="")),sep=""), pch=NA, cex=0.8, bty="n", horiz=T)
box()
# check appropriateness of a Weibull model
# surv.km <- log(-log(1 - risk.km[-1]))
# x <- log(seq(0,1095,10)[-1])
# plot(surv.km, x)
plot.weib(med='Simvastatin',XLIM=1095,title='',km.ff=km.ff,
         risk.weib=risk.weib,YLIM=0.12,inc=0.02,weib.fit=ff)
@
\caption{Kaplan Meier curves and Weibull distribution approximation: time to simvastatin}
\end{figure}


\clearpage
% \section{Warfarin}
<<echo=FALSE, results=hide>>=
dd <- datadist(warf.dat.bl)
options(datadist="dd")

s <- with(warf.dat.bl, Surv(warf.rel.day, warf.status))
ff <- psm(s ~ 1 , data=warf.dat.bl, dist="weibull")

lp.weib <- ff$linear.predictors
scale.weib <- ff$scale

km.ff  <- summary(survfit(s ~ 1, data=warf.dat.bl))

set.seed(12345)
risk.weib <- risk.km <- NULL
for (tt in seq(0,1095,10))   {
risk.weib[tt/10+1] <- mean( rweibull(10000,shape=1/scale.weib[1],
                                     scale=exp(lp.weib[1])) < tt)
risk.km[tt/10+1] <- 1-rev(km.ff$surv[km.ff$time<tt])[1]
}
@

\clearpage
\begin{figure}[!htbp]
  \centering
```

```
    \label{warfarin}
<<echo=FALSE,fig=TRUE,width=5,height=8>>=
par(mfrow=c(2,1),mar=c(4,4,3,1), cex=0.8)
s <- with(warf.dat.bl, Surv(warf.rel.day, warf.status))
km.v <- npsurv(s ~ 1, data=warf.dat.bl)
# km.v <- survfit(s ~ 1)
# fun <- function(x) 1-x
survplot(km.v, n.risk=T, xlim=c(0,1095), ylim=c(0.88,1), time.inc=365, conf="band",
         ylab="Probability of warfarin free", xlab="time (day)",
         title="Warfarin Naive Cohort (baseline)", adj.n.risk=0.5,
         levels.only=TRUE, y.n.risk=0.88, cex.n.risk=0.65, lwd=2,
         label.curves=F, conf.int=T, col.fill=grey(0.8))
abline(v=365*c(0.5,1,1.5,2,2.5,3,3.5,4), lty=3, lwd=0.8, col=grey(0.6))
abline(h=seq(0.7,1,0.05), lty=3, lwd=0.8, col=grey(0.6))
# table(warf.dat.bl$warf.status,useNA='a')
mtext("Warfarin Naive Cohort (baseline, event/n = 1594/26742)",side=3,line=2,cex=1)
# legend("topright", paste("Log rank ",ifelse(p.v<0.001,"p < 0.001",paste("p = ",round(p.v,3), sep="")),sep=""), pch=NA, cex=0.8, bty="n", horiz=T)
box()
# check appropriateness of a Weibull model
# surv.km <- log(-log(1 - risk.km[-1]))
# x <- log(seq(0,1095,10)[-1])
# plot(surv.km, x)
plot.weib(med='Warfarin',XLIM=1095,title='',km.ff=km.ff,
          risk.weib=risk.weib,YLIM=0.12,inc=0.02,weib.fit=ff)
@
\caption{Kaplan Meier curves and Weibull distribution approximation: time to warfarin}
\end{figure}


<<echo=FALSE, results=hide>>=
s <- with(warf.dat.bl, Surv(warf.rel.day, warf.status))
ff <- psm(s ~ afib , data=warf.dat.bl, dist="weibull")

lp.weib <- unique(ff$linear.predictors)
scale.weib <- ff$scale

km.ff  <- summary(survfit(s ~ afib, data=warf.dat.bl))
# km.ff
set.seed(12345)
risk.weib.0 <- risk.km.0 <- risk.weib.1 <- risk.km.1 <- NULL
for (tt in seq(0,1095,10))   {
risk.weib.0[tt/10+1] <- mean( rweibull(10000,shape=1/scale.weib[1],
                                         scale=exp(lp.weib[1])) < tt)
risk.weib.1[tt/10+1] <- mean( rweibull(10000,shape=1/scale.weib[1],
                                         scale=exp(lp.weib[2])) < tt)
# risk.km.0[tt/10+1] <- 1-rev(km.ff$surv[km.ff$time<tt])[1]
}
@


\clearpage
\begin{figure}[!htbp]
  \centering
  \label{warfarin}
<<echo=FALSE,fig=TRUE,width=5,height=8>>=
par(mfrow=c(2,1),mar=c(4,4,3,1), cex=0.8)
s <- with(warf.dat.bl, Surv(warf.rel.day, warf.status))
km.v <- npsurv(s ~ afib, data=warf.dat.bl)
# km.v <- survfit(s ~ 1)
# fun <- function(x) 1-x
survplot(km.v, n.risk=T, xlim=c(0,1095), ylim=c(0.4,1), time.inc=365, conf="band",
         ylab="Probability of warfarin free", xlab="time (day)",
         title="Warfarin Naive Cohort (baseline)", adj.n.risk=0.5,
         levels.only=TRUE, y.n.risk=0.4, cex.n.risk=0.65, lwd=2,
         label.curves=T, conf.int=T, col.fill=grey(0.8))
abline(v=365*c(0.5,1,1.5,2,2.5,3,3.5,4), lty=3, lwd=0.8, col=grey(0.6))
abline(h=seq(0.4,1,0.05), lty=3, lwd=0.8, col=grey(0.6))
# table(warf.dat.bl$warf.status,useNA='a')
mtext("Warfarin Naive Cohort Stratified by BL Afib",side=3,line=2,cex=1)
# legend("topright", paste("Log rank ",ifelse(p.v<0.001,"p < 0.001",paste("p = ",round(p.v,3), sep="")),sep=""), pch=NA, cex=0.8, bty="n", horiz=T)
box()
# check appropriateness of a Weibull model
# surv.km <- log(-log(1 - risk.km[-1]))
# x <- log(seq(0,1095,10)[-1])
# plot(surv.km, x)
plot.weib1 <- function(med='Warfarin',XLIM=1095,YLIM=0.6,title='',km.ff=ff,
                       risk.weib.0=risk.weib.0,risk.weib.1=risk.weib.1,
                       inc=0.1,weib.fit=ff) {
km.ff$time0 <- km.ff$time[1:607]
km.ff$time1 <- km.ff$time[608:675]
km.ff$surv0 <- km.ff$surv[1:607]
km.ff$surv1 <- km.ff$surv[608:675]

plot(km.ff$time0[km.ff$time0<=XLIM], (1-km.ff$surv0[km.ff$time0<=XLIM]),
     xlim=c(0, XLIM), ylim=c(0, YLIM), xlab='',
     ylab=paste("Cumulative incidence of ", med, sep=""), type="l", axes=FALSE, cex=2)
abline(v=c(1, 91, 182, 273, 365, 456, 547, 638, 730, 821, 910, 1001, 1095),
       h=seq(0, 1, inc), col="lightgray")

lines(km.ff$time0[km.ff$time0<=XLIM], (1-km.ff$surv0[km.ff$time0<=XLIM]),
      col=grey(0.1), lty=1, lwd=2.5) # KM curve
lines(km.ff$time1[km.ff$time1<=XLIM], (1-km.ff$surv1[km.ff$time1<=XLIM]),
      col=grey(0.1), lty=1, lwd=2.5) # KM curve
# lines(seq(0,1095,10), risk.weib.0, col='blue', lty=3, lwd=2.5) # psm-Weibull
# lines(seq(0,1095,10), risk.weib.1, col='blue', lty=3, lwd=2.5)
points(seq(0,1095,10), risk.weib.0, col='blue', pch=20,cex=0.5)
points(seq(0,1095,10), risk.weib.1, col='blue', pch=20,cex=0.5)
text(c(910,910),y=c(0.035,0.35),label=c('Afib = 0','Afib = 1'),cex=1)

tt <- seq(0,1095,10)
lo <- loess(risk.weib.0 ~ tt)
lines(x=tt,y=lowess(predict(lo),f=1/3)$y,col='lightblue',lwd=2) # lowess smoother
lo <- loess(risk.weib.1 ~ tt)
lines(x=tt,y=lowess(predict(lo),f=1/5)$y,col='lightblue',lwd=2) # lowess smoother

a <- round(1/ff$scale,2)
b <- round(exp(unique(ff$linear.predictors)),2)

legend('topleft',
       legend=paste0(c("Weibull Parameters:","Shape a = ",
                       "Scale b = ","Scale b = "),
```

```
                    c('',a,b[1],b[2]), c('','',' (Afib = 0)',' (Afib = 1)')),
        cex=1,bty='o',bg='white')

axis(1,at=c(0,182,365,547,730,910,1095),labels=c(0,182,365,547,730,910,1095))
axis(2,at=seq(0,1,inc),labels=seq(0,1,inc))
mtext(title, side=3, line=1, cex=1)
box()
}

plot.weib1(med='Warfarin',XLIM=1095,title='',km.ff=km.ff,
        risk.weib.0=risk.weib.0,risk.weib.1=risk.weib.1,
        YLIM=0.6,inc=0.05,weib.fit=ff)
@
\caption{Kaplan Meier curves and Weibull distribution approximation: time to warfarin}
\end{figure}

%' <<echo=FALSE, results=hide>>=
%' ## population: intersect of antiplt and statin cohort
%' # DAT <- rbind(antiplt.dat.bl, anystat.dat.bl)
%' # DAT <- DAT[order(DAT$id),]
%' # dup <- duplicated(DAT$id)
%' # DAT <- DAT[dup,]
%'
%' ## population: statin cohort
%' # DAT <- anystat.dat.bl
%'
%' ## population: statin cohort with antiplt free
%' # DAT <- anystat.dat.bl
%' # DAT <- subset(DAT, antiplt.since.mh>0)
%'
%' #######################################
%' bl.risk <- fit1$surv.summary[,,1][4]
%' 1-bl.risk^(log(0.6)/log(bl.risk))
%'
%' bl.risk^exp(log(0.6)-.029)
%' dim(antiplt.dat.bl)
%'
%' ## Fit the cox model
%' fit1 <- cph(s0 ~ rcs(age.mh,5) + rcs(bmi,5) + gender + race2 + dm.right + coronary +
%' afib + htn + ath + chf + cereb+any.acute.event + appt.Card.count + stat + warf,
%' data=antiplt.dat.bl, surv=TRUE, time.inc=365, x=TRUE, y=TRUE)
%'
%' ## linear predictor from cox model
%' antiplt.dat.bl$lp.cox <- predict(fit1)
%'
%' ## weibull regression of linear predictor from Cox
%' fit2 <- psm(s0 ~ lp.cox , data=antiplt.dat.bl, dist="weibull")
%'
%' ## linear predictor from weibull
%' antiplt.dat.bl$lp.weib <- predict(fit2)
%'
%' ## scale parameter from weibull
%' antiplt.dat.bl$scale.weib <- fit2$scale
%'
%' ## Create a new dataset called dat
%' dat <- antiplt.dat.bl
%'
%' ## look at the linear predictor function for the weibull model
%' fit2
%'
%' ## for person 1, here is how I generate 10000 observations from a weibull
%' ## (please check parameterization)
%' ## you only need to generate it once per person
%' # rweibull(10000, shape=1/dat$scale.weib[1], scale=exp(dat$lp.weib[1]))
%'
%' ## Here is how I tested...
%' ## i.e.,  calculate the distribution of predicted risks at 2 years
%' # tmp <- NULL
%' # for (i in 1:length(antiplt.dat.bl[,1])){
%' # tmp <- c(tmp, mean(rweibull(10000, shape=1/dat$scale.weib[i],
%' #                               scale=exp(dat$lp.weib[i]))<730))
%' # }
%' # hist(tmp, nclass=200)
%'
%' # dat <- antiplt.dat.bl
%' # write.csv(dat, file="Data4EMERGE.csv")
%'
%' # fit2 <- psm(s0 ~ rcs(age.mh,5) + rcs(bmi,5) + gender + race2 + dm.right + coronary +
%' # afib + htn + ath + chf + cereb+any.acute.event + appt.Card.count + stat + warf,
%' # data=antiplt.dat.bl, times=730, x=TRUE, y=TRUE, dist="exponential")
%' # score <- predict(fit2)
%' # pphsm(fit2)
%' # tmp <- NULL
%' # for (i in 1:length(antiplt.dat.bl[,1])){
%' # tmp <- c(tmp, mean(rweibull(10000, shape=1, scale=exp(score[i]))<730))
%' # }
%' # sum((rexp(10000, rate=exp(min(score)))<730))
%' # kappa <- weib.mod$scale
%' # lambda <- exp(-weib.mod$coeff)^kappa
%' @

\clearpage
% \begin{landscape}
\section*{Source Code for this Report}
\tiny
\verbatimtabinput{WCS_KM_Distribution_Generation.Rnw}
% \end{landscape}


\end{document}
```