Survival of C.elegans with three different media

Jonas Gehrlein

12 nov 2018

code for fitting Kaplan-Meier and log-rank test and for displaying survival curves for each type of media

```
data <- read.table(here('data','Worm_4days.csv'), header = TRUE, sep = ";", dec = ",")
head(data)
```

```
##   ï..Time Status Group Replicate
## 1       4      1  FG13         1
## 2       6      1  FG13         1
## 3       6      1  FG13         1
## 4       8      1  FG13         1
## 5       8      1  FG13         1
## 6      10      1  FG13         1
```

here() should show that your position is in the folder 7.semester else select the active project to be 7.semester in the upper right corner.

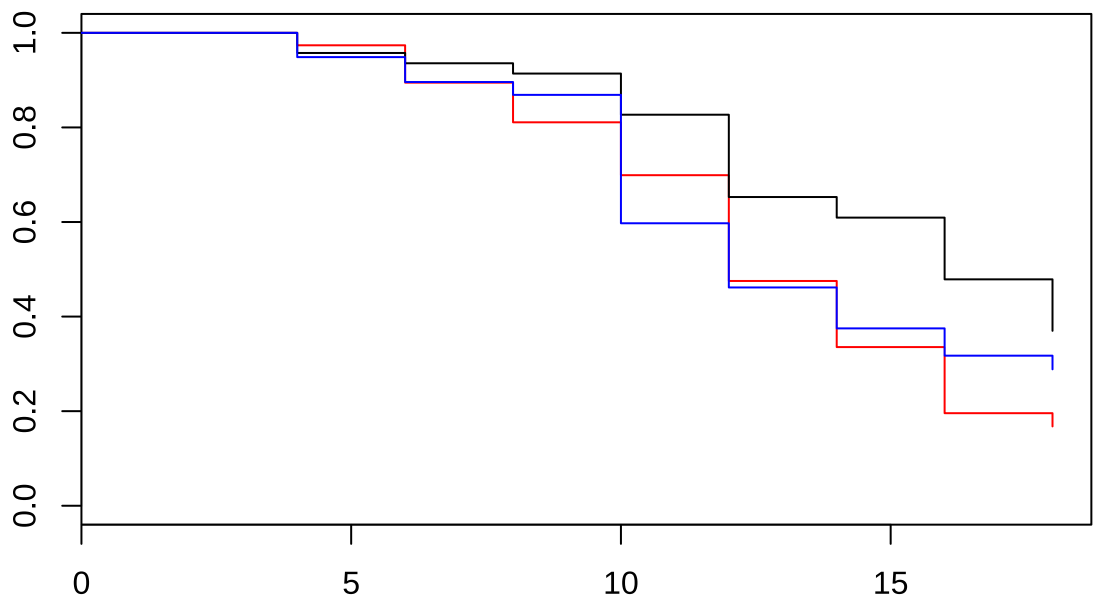
Or make a new .Rproj file in the folder 7.semester. There is a problem with the numbers of the worms so we change that with col.names()

```
colnames(data)<-c("Time","status","media",'Replicate')
head(data)
```

```
##   Time status media Replicate
## 1    4      1  FG13         1
## 2    6      1  FG13         1
## 3    6      1  FG13         1
## 4    8      1  FG13         1
## 5    8      1  FG13         1
## 6   10      1  FG13         1
```

Then we create a survival object with the survival package and make a kaplan-meier curve

```
data$Survobj <- with(data,Surv(data$Time, event = data$status))
km <- survfit(Survobj ~ media, data = data,conf.type = "log-log", error = "greenwood")
s_km <- summary(km)
plot(km, col = c("red","black",'blue'))
```




```r
fit <- coxph(Survobj ~ media, data = data)
summary(fit)
```

```
## Call:
## coxph(formula = Survobj ~ media, data = data)
##
##   n= 124, number of events= 85
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## mediaFG20 -0.6010    0.5482   0.2622 -2.293   0.0219 *
## mediaOP50 -0.1943    0.8234   0.2685 -0.724   0.4693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20    0.5482      1.824    0.3280    0.9165
## mediaOP50    0.8234      1.214    0.4865    1.3936
##
## Concordance= 0.576  (se = 0.031 )
## Rsquare= 0.044   (max possible= 0.997 )
## Likelihood ratio test= 5.57  on 2 df,   p=0.06
## Wald test            = 5.44  on 2 df,   p=0.07
## Score (logrank) test = 5.57  on 2 df,   p=0.06
```

Shows errorbars and lineplot based on one of Anders papers https://onlinelibrary.wiley.com/doi/full/10.

1111/acel.12165 This is done by extracting the surviving proportion, standard errors and time from survfit(). and adding a startpoint where the survival is 100% at 0 hours.

```
df_fly_13 <- data.frame(c(0,s_km$time[1:8]),c(1,s_km$surv[1:8]),
                        c(0,s_km$std.err[1:8]))
colnames(df_fly_13) <- c('Time','Surv','Std.error')
df_fly_20 <- data.frame(c(0,s_km$time[9:16]),c(1,s_km$surv[9:16]),
                        c(0,s_km$std.err[9:16]))
colnames(df_fly_20) <- c('Time','Surv','Std.error')
df_OP50 <- data.frame(c(0,s_km$time[17:24]),c(1,s_km$surv[17:24]),
                        c(0,s_km$std.err[17:24]))
colnames(df_OP50) <- c('Time','Surv','Std.error')

plot(df_fly_13$Time[2:9],df_fly_13$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival heat stress for'~italic(C.elegans)),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_fly_13$Time,df_fly_13$Surv)
arrows(df_fly_13$Time, df_fly_13$Surv-df_fly_13$Std.error, df_fly_13$Time,
       df_fly_13$Surv+df_fly_13$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_13$Time, df_fly_13$Surv - df_fly_13$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

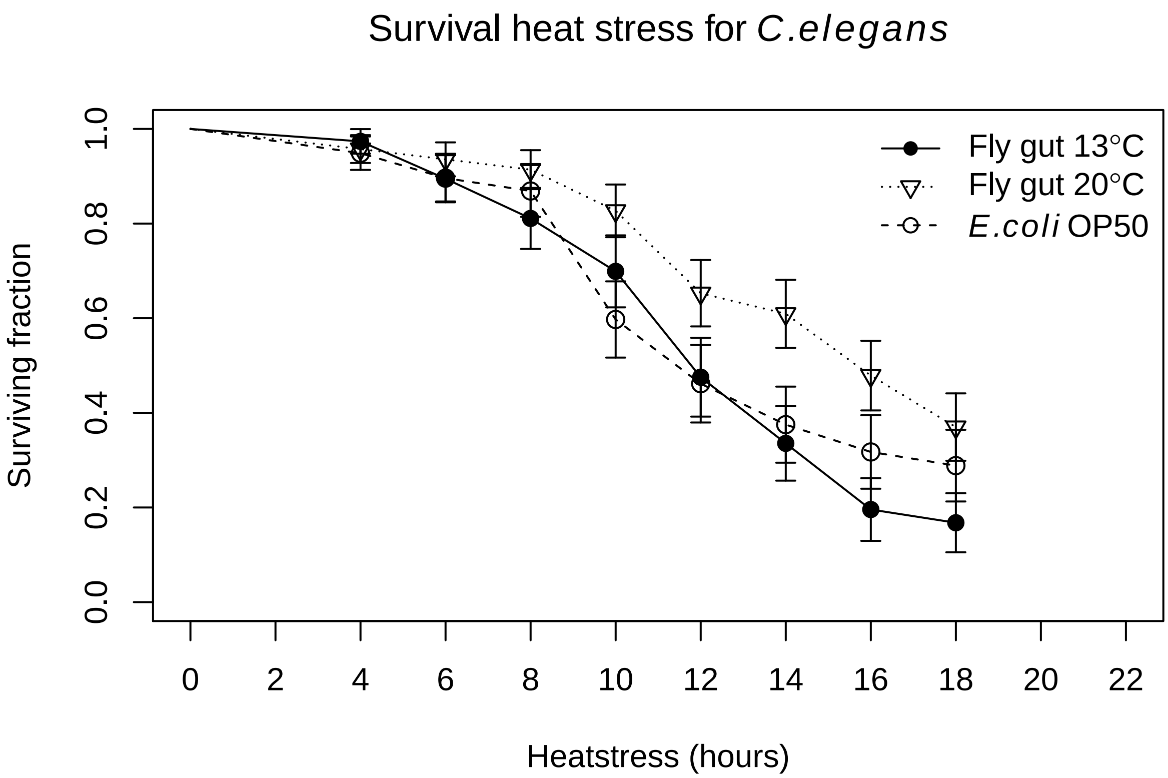
```
points(df_OP50$Time[2:9],df_OP50$Surv[2:9], pch = 1, cex = 1.2)
lines(df_OP50$Time,df_OP50$Surv, lty = 2)
arrows(df_OP50$Time, df_OP50$Surv-df_OP50$Std.error,
      df_OP50$Time, df_OP50$Surv+df_OP50$Std.error,
      length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_OP50$Time, df_OP50$Surv - df_OP50$Std.error,
## df_OP50$Time, : zero-length arrow is of indeterminate angle and so skipped
```

```
points(df_fly_20$Time[2:9],df_fly_20$Surv[2:9], pch = 6)
lines(df_fly_20$Time,df_fly_20$Surv,lty = 3)
arrows(df_fly_20$Time, df_fly_20$Surv-df_fly_20$Std.error, df_fly_20$Time,
       df_fly_20$Surv+df_fly_20$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

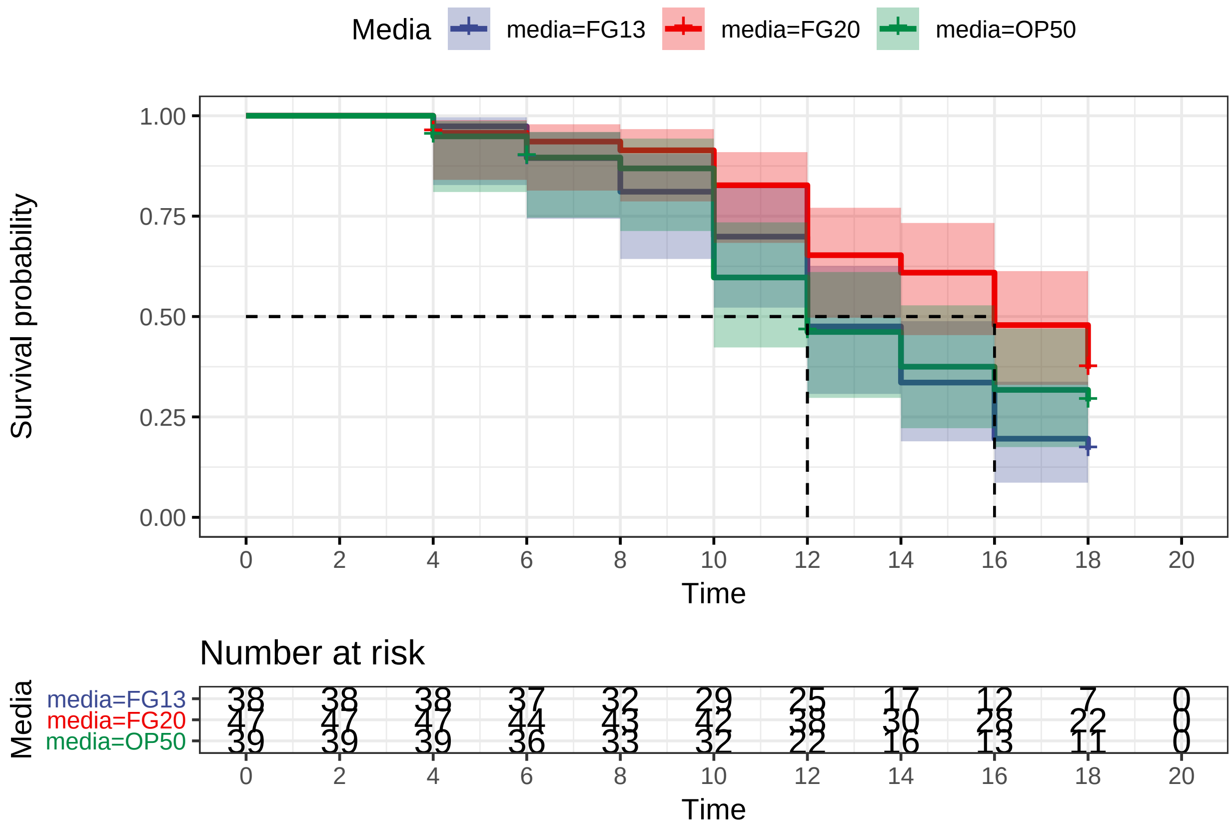
```
## Warning in arrows(df_fly_20$Time, df_fly_20$Surv - df_fly_20$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
legend('topright', pch = c(16,6,1),lty = c(1,3,2),
       legend = c(expression('Fly gut 13'*degree*C),
                  expression('Fly gut 20'*degree*C),
                  expression(italic(E.coli) ~ OP50)), bty = 'n')
```



Then we create a normal Kaplan-Meier curve

```
ggsurvplot(km,data = data, conf.int = TRUE, ggtheme = theme_bw(),
           risk.table = 0.25, palette ='aaas', surv.median.line = 'hv',legend.title = 'Media',break.x.by
```



Now we test for difference between the curves with both log-rank and gehan-wilcoxon and with an cox proportional hazard model

```
survdiff(Survobj ~media, data = data, rho = 0)
```

```
## Call:
## survdiff(formula = Survobj ~ media, data = data, rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 38       30     23.0     2.115     3.498
## media=FG20 47       29     38.4     2.280     5.052
## media=OP50 39       26     23.6     0.238     0.396
##
##  Chisq= 5.6  on 2 degrees of freedom, p= 0.06
```

```
survdiff(Survobj ~media, data = data, rho = 1)
```

```
## Call:
## survdiff(formula = Survobj ~ media, data = data, rho = 1)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 38     21.1     16.7     1.149      2.42
## media=FG20 47     18.6     26.0     2.101      5.80
## media=OP50 39     20.0     17.0     0.531      1.13
##
##  Chisq= 5.9  on 2 degrees of freedom, p= 0.05
```

```r
fit <- coxph(Survobj ~media, data = data)
summary(fit)
```

```
## Call:
## coxph(formula = Survobj ~ media, data = data)
##
##   n= 124, number of events= 85
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## mediaFG20 -0.6010    0.5482   0.2622 -2.293   0.0219 *
## mediaOP50 -0.1943    0.8234   0.2685 -0.724   0.4693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20    0.5482      1.824    0.3280    0.9165
## mediaOP50    0.8234      1.214    0.4865    1.3936
##
## Concordance= 0.576  (se = 0.031 )
## Rsquare= 0.044   (max possible= 0.997 )
## Likelihood ratio test= 5.57  on 2 df,   p=0.06
## Wald test            = 5.44  on 2 df,   p=0.07
## Score (logrank) test = 5.57  on 2 df,   p=0.06
```

Then we compare the different groups individually

```r
d_OP50 <- data[data$media == 'OP50',]
d_13 <- data[data$media == "FG13",]
d_20 <- data[data$media == "FG20",]
d_fly <- rbind(d_13, d_20)
d_2050 <- rbind(d_20, d_OP50)
d_1350 <- rbind(d_13, d_OP50)
```

First tests for each combination first fly media

```r
survdiff(Surv(Time,status) ~ media, data = d_fly, rho = 0)
```

```
## Call:
## survdiff(formula = Surv(Time, status) ~ media, data = d_fly,
##     rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 38       30     21.8      3.11         6
## media=FG20 47       29     37.2      1.82         6
##
##  Chisq= 6  on 1 degrees of freedom, p= 0.01
```

```r
survdiff(Surv(Time,status) ~ media, data = d_fly, rho = 1)
```

```
## Call:
## survdiff(formula = Surv(Time, status) ~ media, data = d_fly,
```

```
##     rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 38     22.0     16.0      2.23      5.56
## media=FG20 47     19.4     25.3      1.41      5.56
##
##  Chisq= 5.6  on 1 degrees of freedom, p= 0.02
```

```
fit_fly <- coxph(Surv(Time,status) ~media, data = d_fly)
summary(fit_fly)
```

```
## Call:
## coxph(formula = Surv(Time, status) ~ media, data = d_fly)
##
##   n= 85, number of events= 59
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## mediaFG20 -0.6352    0.5298   0.2632 -2.413   0.0158 *
## mediaOP50      NA        NA   0.0000     NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20    0.5298      1.887    0.3163    0.8876
## mediaOP50        NA         NA        NA        NA
##
## Concordance= 0.587  (se = 0.035 )
## Rsquare= 0.065   (max possible= 0.996 )
## Likelihood ratio test= 5.75  on 1 df,   p=0.02
## Wald test            = 5.82  on 1 df,   p=0.02
## Score (logrank) test = 6.01  on 1 df,   p=0.01
```

Then 20 degrees and OP50

```
survdiff(Surv(Time,status) ~ media, data = d_2050, rho = 0)
```

```
## Call:
## survdiff(formula = Surv(Time, status) ~ media, data = d_2050,
##     rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG20 47       29     33.9     0.701      2.16
## media=OP50 39       26     21.1     1.123      2.16
##
##  Chisq= 2.2  on 1 degrees of freedom, p= 0.1
```

```
survdiff(Surv(Time,status) ~ media, data = d_2050, rho = 1)
```

```
## Call:
## survdiff(formula = Surv(Time, status) ~ media, data = d_2050,
##     rho = 1)
##
```

```
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG20 47     19.4     24.1      0.90      3.38
## media=OP50 39     20.5     15.8      1.37      3.38
##
##  Chisq= 3.4  on 1 degrees of freedom, p= 0.07
```

```r
fit_2050 <- coxph(Surv(Time,status) ~media, data = d_2050)
summary(fit_2050)
```

```
## Call:
## coxph(formula = Surv(Time, status) ~ media, data = d_2050)
##
##   n= 86, number of events= 55
##
##               coef exp(coef) se(coef)     z Pr(>|z|)
## mediaFG20 -0.3933    0.6748   0.2712 -1.45    0.147
## mediaOP50      NA        NA   0.0000    NA       NA
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20    0.6748      1.482    0.3966     1.148
## mediaOP50        NA         NA        NA        NA
##
## Concordance= 0.569  (se = 0.037 )
## Rsquare= 0.024   (max possible= 0.994 )
## Likelihood ratio test= 2.08  on 1 df,   p=0.1
## Wald test            = 2.1  on 1 df,   p=0.1
## Score (logrank) test = 2.13  on 1 df,   p=0.1
```

Then for 13 degrees and OP50

```r
survdiff(Surv(Time,status) ~ media, data = d_1350, rho = 0)
```

```
## Call:
## survdiff(formula = Surv(Time, status) ~ media, data = d_1350,
##     rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 38       30     27.8     0.172     0.423
## media=OP50 39       26     28.2     0.170     0.423
##
##  Chisq= 0.4  on 1 degrees of freedom, p= 0.5
```

```r
survdiff(Surv(Time,status) ~ media, data = d_1350, rho = 1)
```

```
## Call:
## survdiff(formula = Surv(Time, status) ~ media, data = d_1350,
##     rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 38     19.5     19.1   0.00887    0.0279
## media=OP50 39     18.8     19.2   0.00879    0.0279
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

```
fit_1350 <- coxph(Surv(Time,status) ~media, data = d_1350)
summary(fit_1350)
```

```
## Call:
## coxph(formula = Surv(Time, status) ~ media, data = d_1350)
##
##   n= 77, number of events= 56
##
##                coef exp(coef) se(coef)      z Pr(>|z|)
## mediaFG20        NA        NA   0.0000     NA       NA
## mediaOP50 -0.1731    0.8410   0.2685 -0.645    0.519
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20        NA         NA        NA        NA
## mediaOP50     0.841      1.189    0.4969     1.424
##
## Concordance= 0.506  (se = 0.04 )
## Rsquare= 0.005   (max possible= 0.996 )
## Likelihood ratio test= 0.42  on 1 df,   p=0.5
## Wald test            = 0.42  on 1 df,   p=0.5
## Score (logrank) test = 0.42  on 1 df,   p=0.5
```

Where only 20 degrees and 13 degrees are significantly different from each other. But there is also some difference between 20 degrees and OP50. Where there isn´t a big difference between 13 degrees and OP50.

Then dose response curves with first the two fly media.

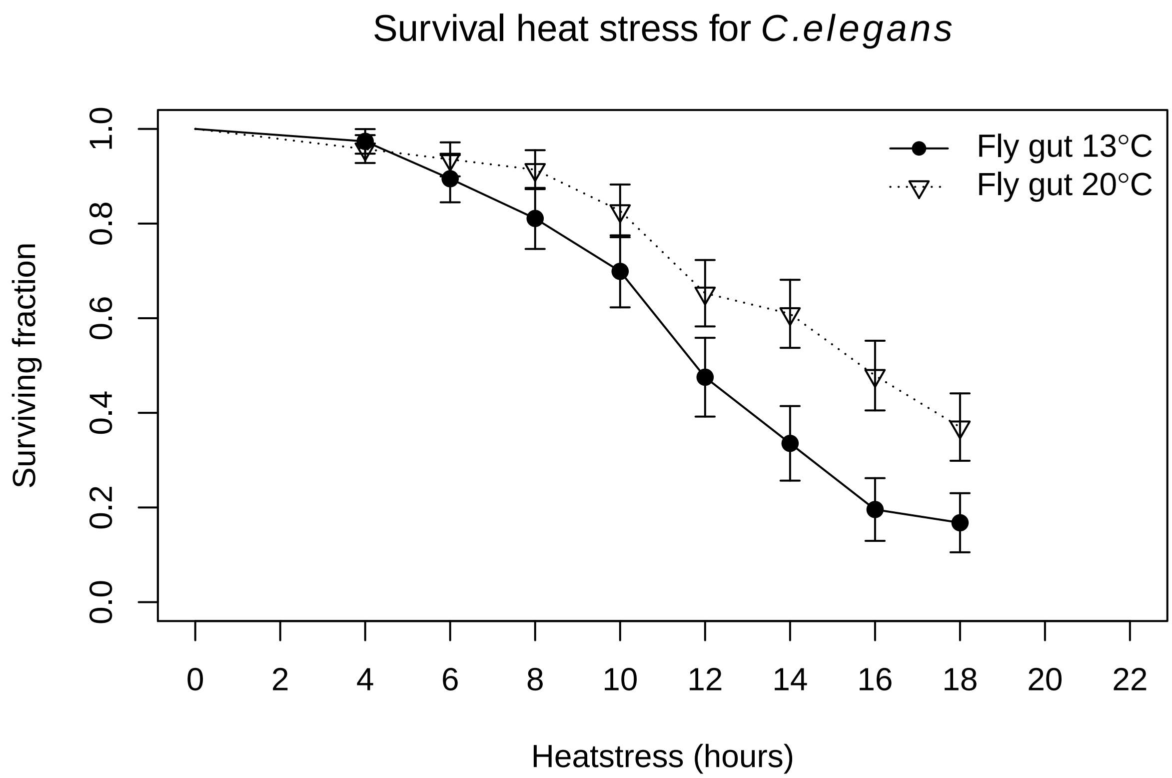
```
plot(df_fly_13$Time[2:9],df_fly_13$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival heat stress for'~italic(C.elegans)),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_fly_13$Time,df_fly_13$Surv)
arrows(df_fly_13$Time, df_fly_13$Surv-df_fly_13$Std.error, df_fly_13$Time,
       df_fly_13$Surv+df_fly_13$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_13$Time, df_fly_13$Surv - df_fly_13$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
points(df_fly_20$Time[2:9],df_fly_20$Surv[2:9], pch = 6)
lines(df_fly_20$Time,df_fly_20$Surv,lty = 3)
arrows(df_fly_20$Time, df_fly_20$Surv-df_fly_20$Std.error, df_fly_20$Time,
       df_fly_20$Surv+df_fly_20$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_20$Time, df_fly_20$Surv - df_fly_20$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
legend('topright', pch = c(16,6),lty = c(1,3),
       legend = c(expression('Fly gut 13'*degree*C),
                  expression('Fly gut 20'*degree*C)), bty = 'n')
```



Than 20 degrees and OP50

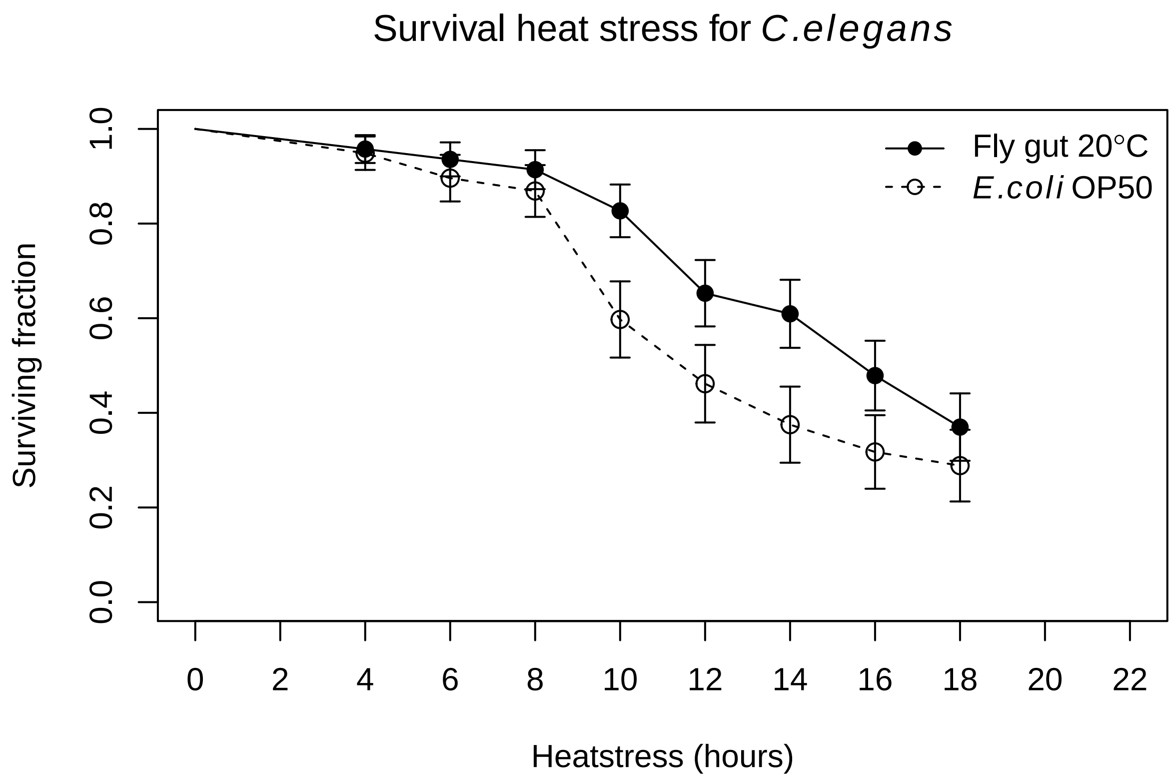
```
plot(df_fly_20$Time[2:9],df_fly_20$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival heat stress for'~italic(C.elegans)),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
points(df_OP50$Time[2:9],df_OP50$Surv[2:9], pch = 1, cex = 1.2)
lines(df_OP50$Time,df_OP50$Surv, lty = 2)
arrows(df_OP50$Time, df_OP50$Surv-df_OP50$Std.error,
      df_OP50$Time, df_OP50$Surv+df_OP50$Std.error,
      length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_OP50$Time, df_OP50$Surv - df_OP50$Std.error,
## df_OP50$Time, : zero-length arrow is of indeterminate angle and so skipped
```

```
lines(df_fly_20$Time,df_fly_20$Surv,lty = 1)
arrows(df_fly_20$Time, df_fly_20$Surv-df_fly_20$Std.error, df_fly_20$Time,
       df_fly_20$Surv+df_fly_20$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_20$Time, df_fly_20$Surv - df_fly_20$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
legend('topright', pch = c(16,1),lty = c(1,2),
       legend = c(expression('Fly gut 20'*degree*C),
                  expression(italic(E.coli) ~ OP50)), bty = 'n')
```



Then 13 degrees and OP50

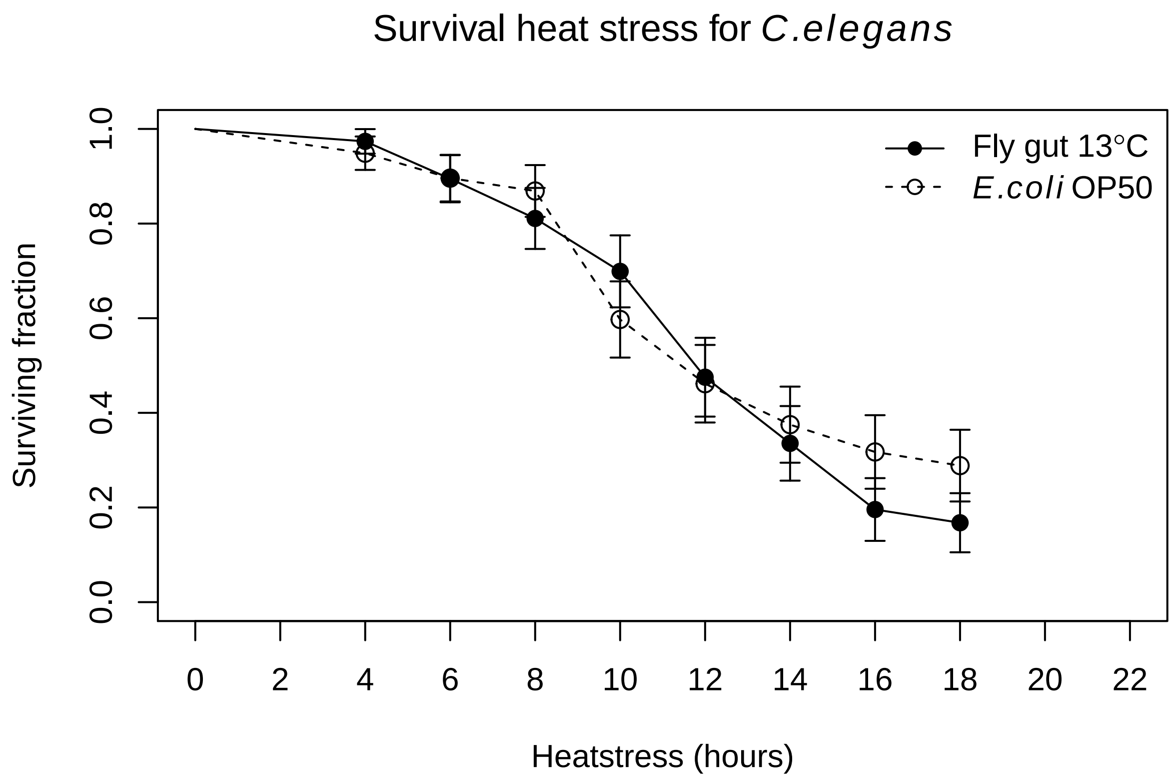
```
plot(df_fly_13$Time[2:9],df_fly_13$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival heat stress for'~italic(C.elegans)),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_fly_13$Time,df_fly_13$Surv)
arrows(df_fly_13$Time, df_fly_13$Surv-df_fly_13$Std.error, df_fly_13$Time,
       df_fly_13$Surv+df_fly_13$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_13$Time, df_fly_13$Surv - df_fly_13$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
points(df_OP50$Time[2:9],df_OP50$Surv[2:9], pch = 1, cex = 1.2)
lines(df_OP50$Time,df_OP50$Surv, lty = 2)
arrows(df_OP50$Time, df_OP50$Surv-df_OP50$Std.error,
      df_OP50$Time, df_OP50$Surv+df_OP50$Std.error,
      length=0.05, angle=90, code=3, col = 'black')
```

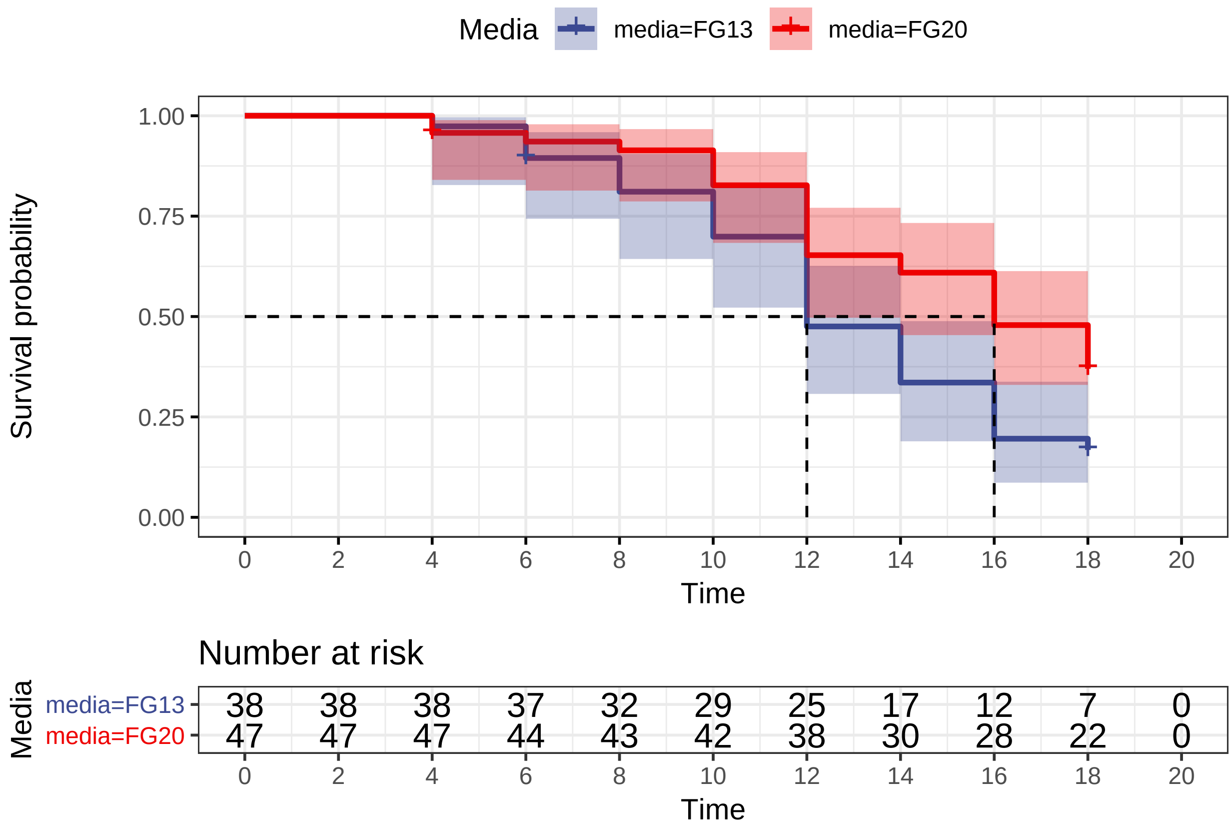
```
## Warning in arrows(df_OP50$Time, df_OP50$Surv - df_OP50$Std.error,
## df_OP50$Time, : zero-length arrow is of indeterminate angle and so skipped
```

```
legend('topright', pch = c(16,1),lty = c(1,2),
       legend = c(expression('Fly gut 13'*degree*C),
                  expression(italic(E.coli) ~ OP50)), bty = 'n')
```

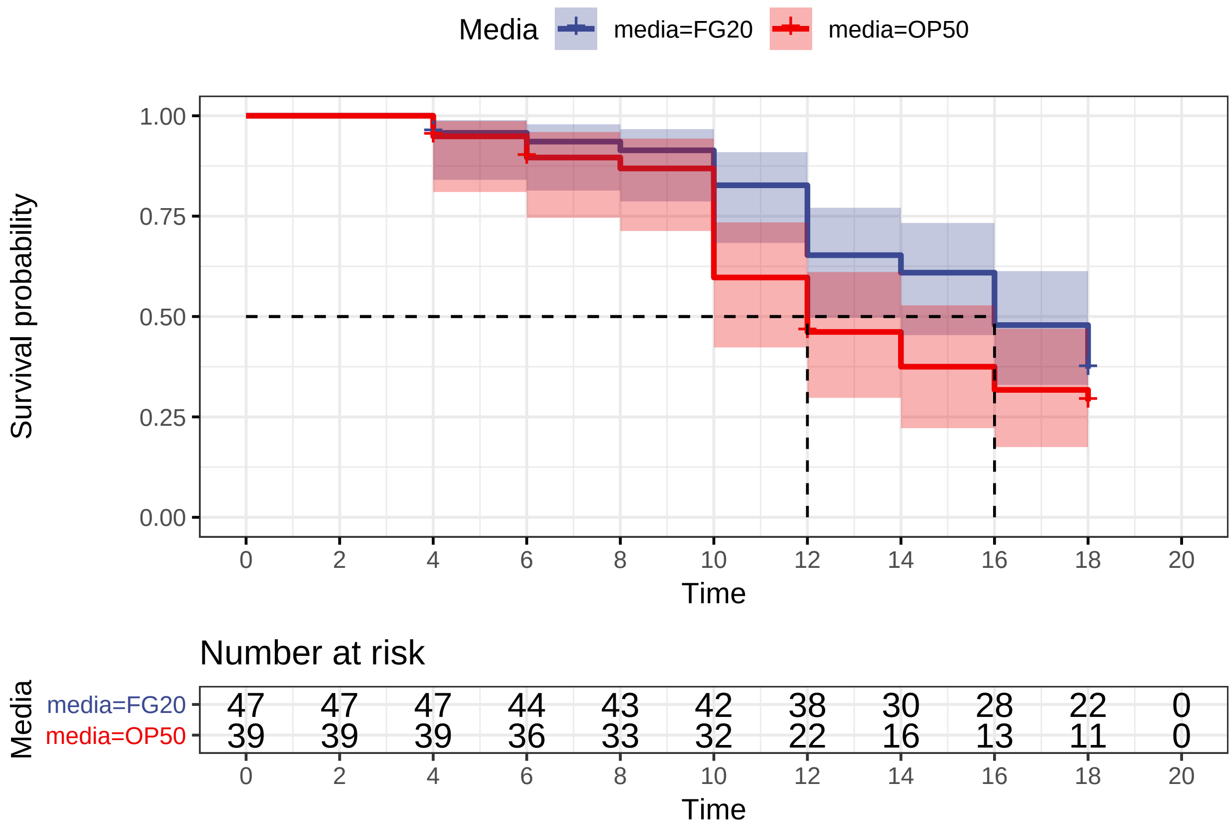


Then Kaplan meier curves

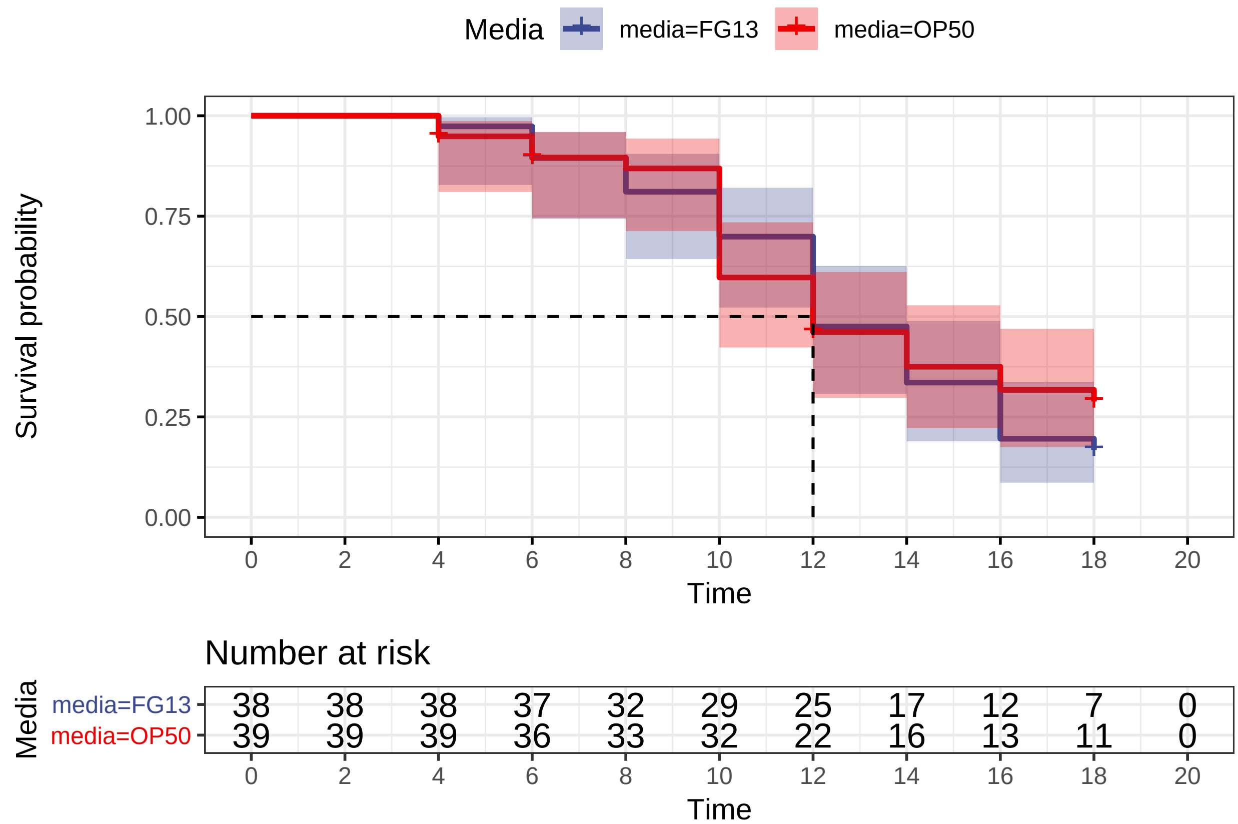
```
km_fly <- survfit(Surv(Time,status) ~ media, data = d_fly,conf.type = "log-log", error = "greenwood")
km_2050 <- survfit(Surv(Time,status) ~ media, data = d_2050,conf.type = "log-log", error = "greenwood")
km_1350 <- survfit(Surv(Time,status) ~ media, data = d_1350,conf.type = "log-log", error = "greenwood")
ggsurvplot(km_fly,data = d_fly, conf.int = TRUE,
           ggtheme = theme_bw(),risk.table = 0.25,
           palette ='aaas', surv.median.line = 'hv',
           legend.title = 'Media',break.x.by = 2)
```




```
ggsurvplot(km_2050,data = d_2050, conf.int = TRUE,
           ggtheme = theme_bw(), risk.table = 0.25,
           palette ='aaas', surv.median.line = 'hv',
           legend.title = 'Media',break.x.by = 2)
```




```
ggsurvplot(km_1350,data = d_1350, conf.int = TRUE,
           ggtheme = theme_bw(),risk.table = 0.25,
           palette ='aaas', surv.median.line = 'hv',
           legend.title = 'Media', break.x.by = 2)
```



We then try to split up in replicates to see if there are any differences between scorers

```
d1 <- data[data$Replicate == 1,]
d2 <- data[data$Replicate == 2,]
d1$Survobj <- with(d1,Surv(d1$Time, event = d1$status))
d2$Survobj <- with(d2,Surv(d2$Time, event = d2$status))
```

Then we make kaplan-meier curves and dose response curves

```
km1 <- survfit(Survobj ~ media, data = d1,conf.type = "log-log")
s_km1 <- summary(km1)
km2 <- survfit(Survobj ~ media, data = d2,conf.type = "log-log")
s_km2 <- summary(km2)
df_fly_13 <- data.frame(c(0,s_km1$time[1:7]),c(1,s_km1$surv[1:7]),
                        c(0,s_km1$std.err[1:7]))
colnames(df_fly_13) <- c('Time','Surv','Std.error')
df_fly_20 <- data.frame(c(0,s_km1$time[8:15]),c(1,s_km1$surv[8:15]),
                        c(0,s_km1$std.err[8:15]))
colnames(df_fly_20) <- c('Time','Surv','Std.error')
df_OP50 <- data.frame(c(0,s_km1$time[16:24]),c(1,s_km1$surv[16:24]),
                        c(0,s_km1$std.err[16:24]))
colnames(df_OP50) <- c('Time','Surv','Std.error')

plot(df_fly_13$Time[2:9],df_fly_13$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival heat stress for'~italic(C.elegans)),
```

```r
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_fly_13$Time,df_fly_13$Surv)
arrows(df_fly_13$Time, df_fly_13$Surv-df_fly_13$Std.error, df_fly_13$Time,
       df_fly_13$Surv+df_fly_13$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_13$Time, df_fly_13$Surv - df_fly_13$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

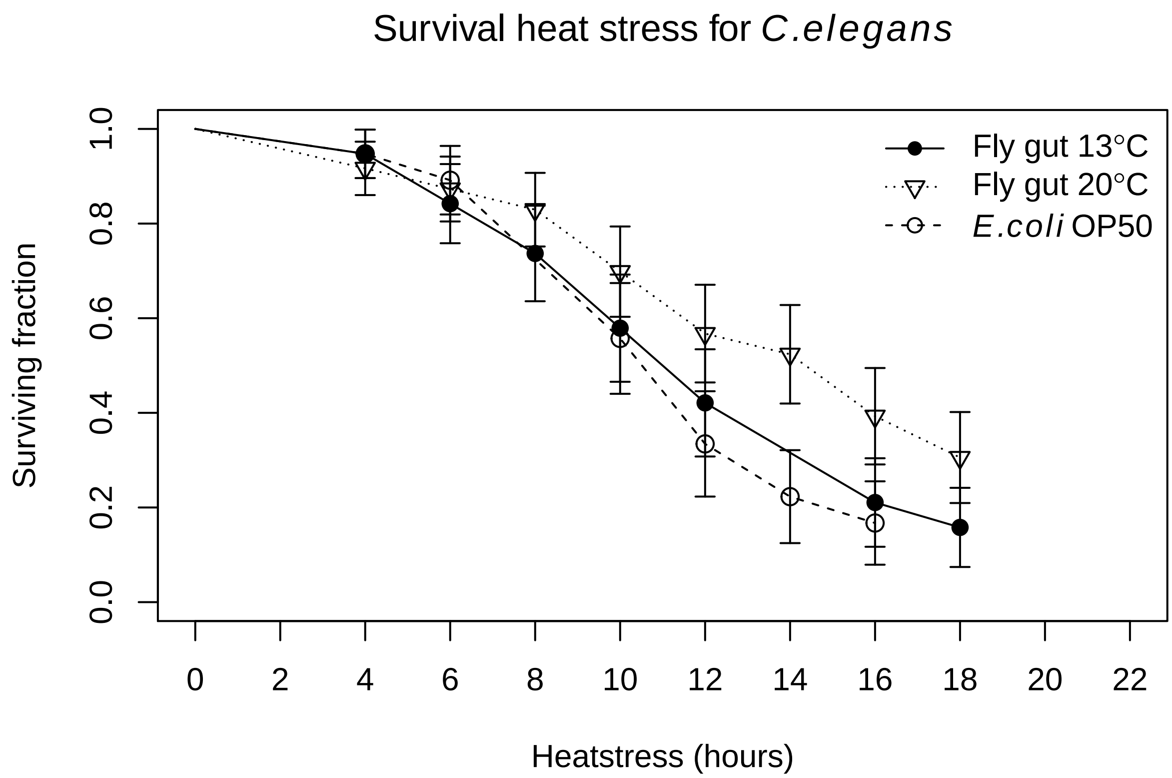
```r
points(df_OP50$Time[2:9],df_OP50$Surv[2:9], pch = 1, cex = 1.2)
lines(df_OP50$Time,df_OP50$Surv, lty = 2)
arrows(df_OP50$Time, df_OP50$Surv-df_OP50$Std.error,
      df_OP50$Time, df_OP50$Surv+df_OP50$Std.error,
      length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_OP50$Time, df_OP50$Surv - df_OP50$Std.error,
## df_OP50$Time, : zero-length arrow is of indeterminate angle and so skipped
```

```r
points(df_fly_20$Time[2:9],df_fly_20$Surv[2:9], pch = 6)
lines(df_fly_20$Time,df_fly_20$Surv,lty = 3)
arrows(df_fly_20$Time, df_fly_20$Surv-df_fly_20$Std.error, df_fly_20$Time,
       df_fly_20$Surv+df_fly_20$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_20$Time, df_fly_20$Surv - df_fly_20$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```r
legend('topright', pch = c(16,6,1),lty = c(1,3,2),
       legend = c(expression('Fly gut 13'*degree*C),
                  expression('Fly gut 20'*degree*C),
                  expression(italic(E.coli) ~ OP50)), bty = 'n')
```




```
df_fly_13 <- data.frame(c(0,s_km2$time[1:6],18),
                        c(1,s_km2$surv[1:6],s_km2$surv[6]),
                        c(0,s_km2$std.err[1:6],s_km2$std.err[6]))
colnames(df_fly_13) <- c('Time','Surv','Std.error')
df_fly_20 <- data.frame(c(0,s_km2$time[7:11]),c(1,s_km2$surv[7:11]),
                        c(0,s_km2$std.err[7:11]))
colnames(df_fly_20) <- c('Time','Surv','Std.error')
df_OP50 <- data.frame(c(0,s_km2$time[12:19]),c(1,s_km2$surv[12:19]),
                        c(0,s_km2$std.err[12:19]))
colnames(df_OP50) <- c('Time','Surv','Std.error')

plot(df_fly_13$Time[2:8],df_fly_13$Surv[2:8],
     pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival heat stress for'~italic(C.elegans)),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_fly_13$Time,df_fly_13$Surv)
arrows(df_fly_13$Time, df_fly_13$Surv-df_fly_13$Std.error, df_fly_13$Time,
       df_fly_13$Surv+df_fly_13$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_13$Time, df_fly_13$Surv - df_fly_13$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
points(df_OP50$Time[2:9],df_OP50$Surv[2:9], pch = 1, cex = 1.2)
lines(df_OP50$Time,df_OP50$Surv, lty = 2)
```

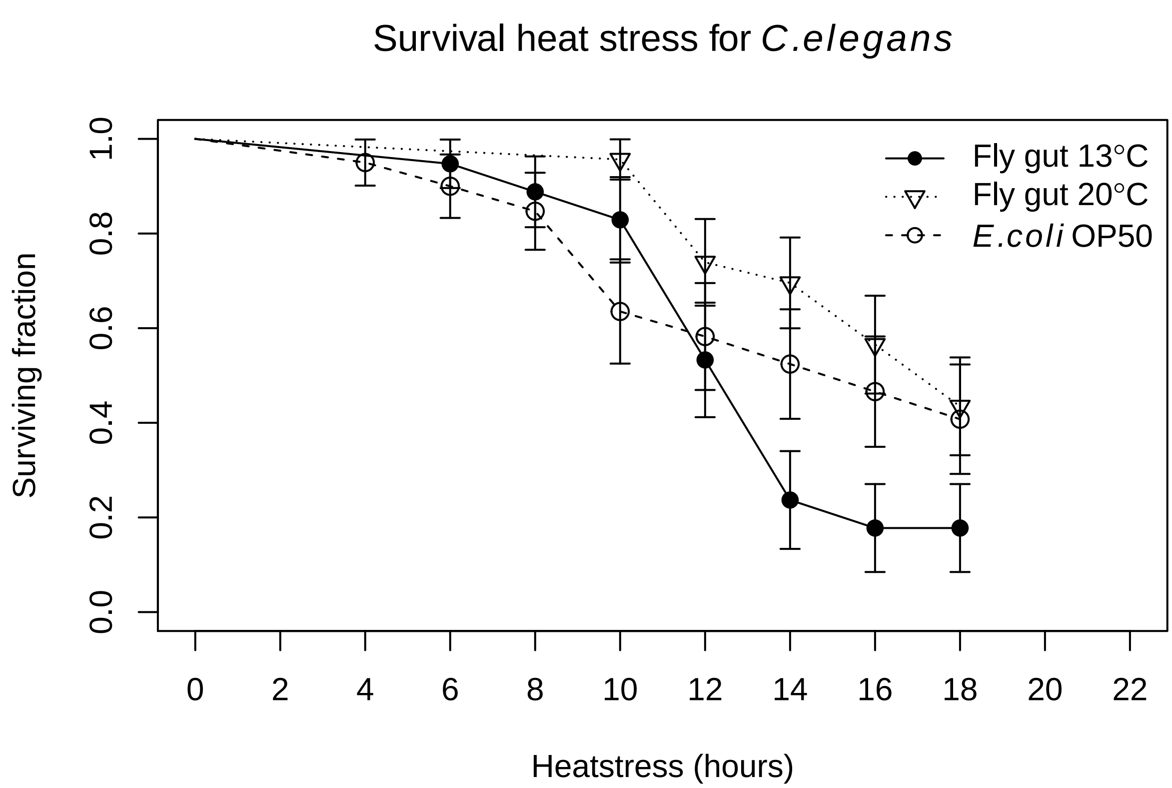
```
arrows(df_OP50$Time, df_OP50$Surv-df_OP50$Std.error,
      df_OP50$Time, df_OP50$Surv+df_OP50$Std.error,
      length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_OP50$Time, df_OP50$Surv - df_OP50$Std.error,
## df_OP50$Time, : zero-length arrow is of indeterminate angle and so skipped
```

```
points(df_fly_20$Time[2:9],df_fly_20$Surv[2:9], pch = 6)
lines(df_fly_20$Time,df_fly_20$Surv,lty = 3)
arrows(df_fly_20$Time, df_fly_20$Surv-df_fly_20$Std.error, df_fly_20$Time,
       df_fly_20$Surv+df_fly_20$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

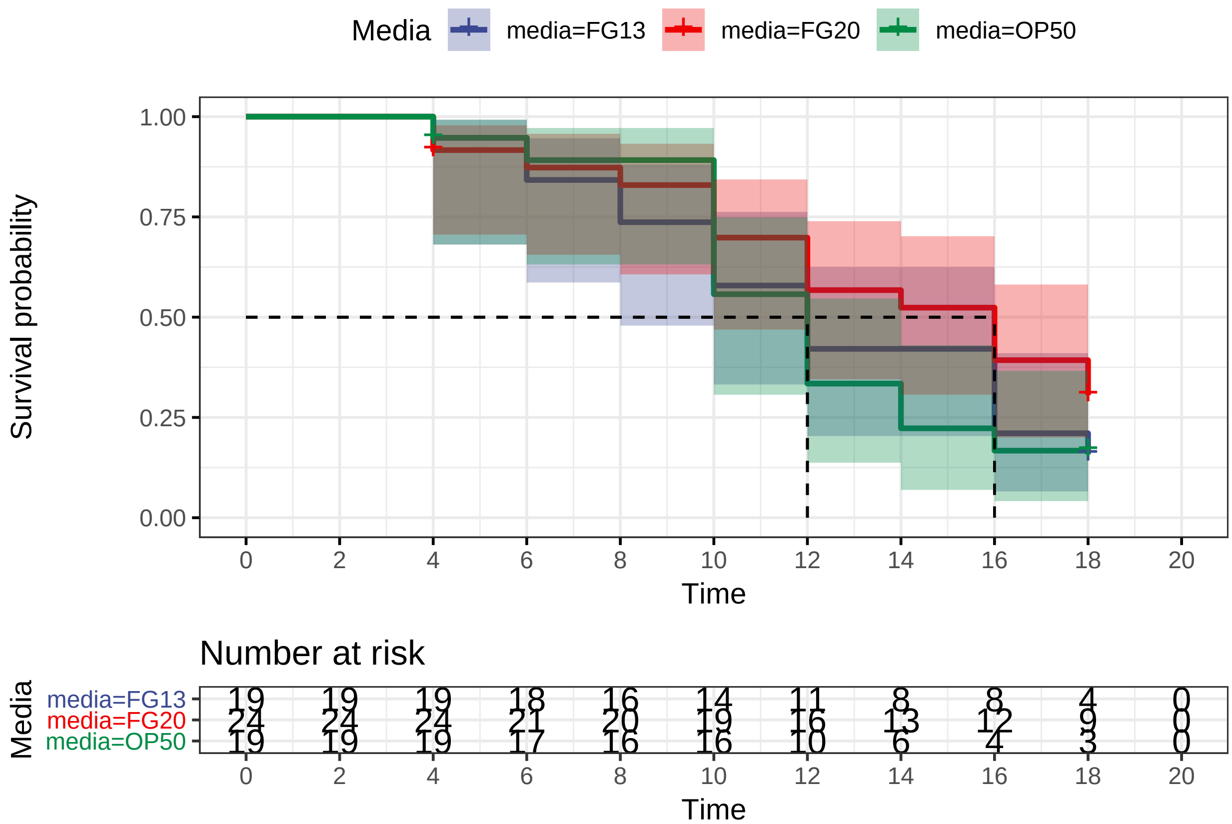
```
## Warning in arrows(df_fly_20$Time, df_fly_20$Surv - df_fly_20$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
legend('topright', pch = c(16,6,1),lty = c(1,3,2),
       legend = c(expression('Fly gut 13'*degree*C),
                  expression('Fly gut 20'*degree*C),
                  expression(italic(E.coli) ~ OP50)), bty = 'n')
```

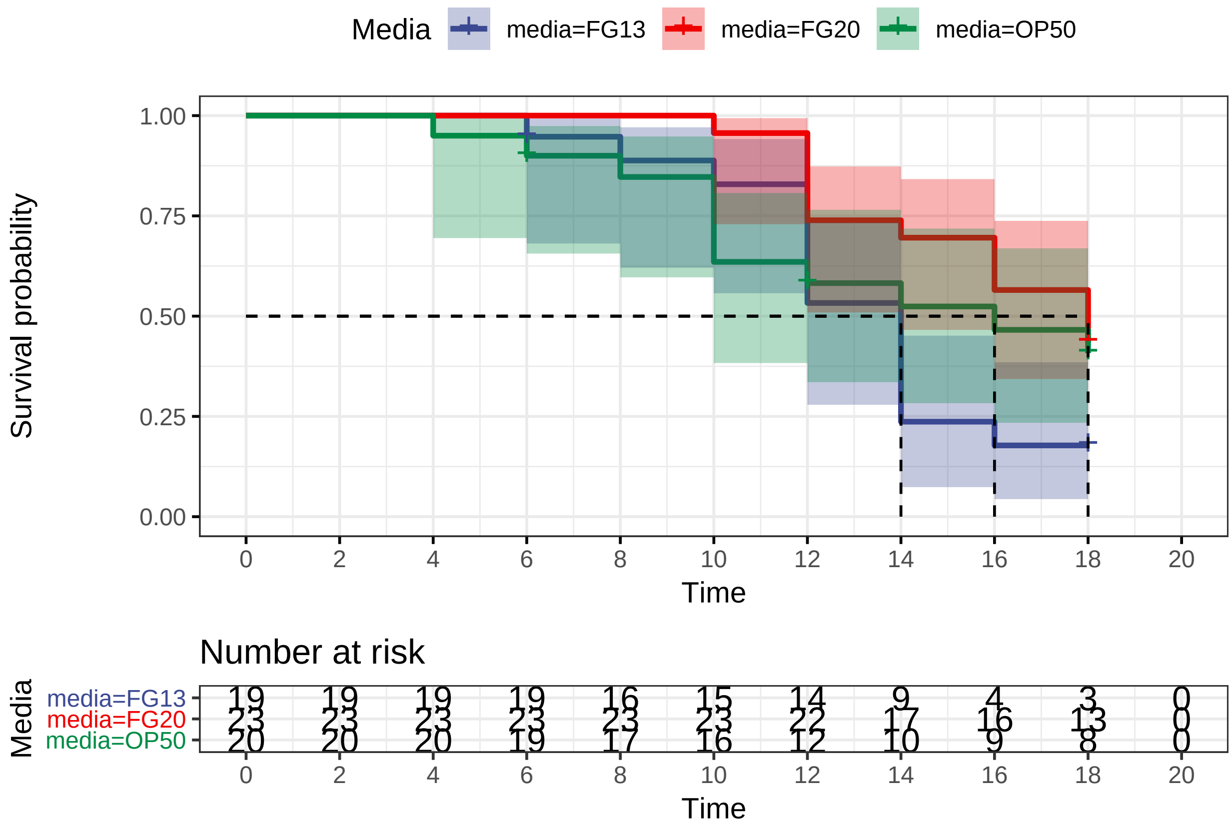


Then kaplan meier

```
ggsurvplot(km1,data = data, conf.int = TRUE, ggtheme = theme_bw(),
           risk.table = 0.25, palette ='aaas', surv.median.line = 'hv',
           legend.title = 'Media',break.x.by = 2)
```




```
ggsurvplot(km2,data = data, conf.int = TRUE, ggtheme = theme_bw(),
           risk.table = 0.25, palette ='aaas', surv.median.line = 'hv',
           legend.title = 'Media',break.x.by = 2)
```



and at last some tests

```
survdiff(Survobj ~media, data = d1, rho = 0)
```

```
## Call:
## survdiff(formula = Survobj ~ media, data = d1, rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 19       16     14.1     0.244     0.431
## media=FG20 24       16     20.3     0.915     2.014
## media=OP50 19       15     12.5     0.480     0.815
##
##  Chisq= 2.1  on 2 degrees of freedom, p= 0.4
```

```
survdiff(Survobj ~media, data = d1, rho = 1)
```

```
## Call:
## survdiff(formula = Survobj ~ media, data = d1, rho = 1)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 19     10.8     9.54     0.162     0.368
## media=FG20 24     10.3    13.14     0.613     1.687
## media=OP50 19     10.6     9.03     0.282     0.613
##
##  Chisq= 1.7  on 2 degrees of freedom, p= 0.4
```

```r
fit_d1 <- coxph(Survobj ~media, data = d1)
summary(fit_d1)
```

```
## Call:
## coxph(formula = Survobj ~ media, data = d1)
##
##   n= 62, number of events= 47
##
##                coef exp(coef) se(coef)      z Pr(>|z|)
## mediaFG20 -0.41066   0.66321  0.35443 -1.159    0.247
## mediaOP50  0.07651   1.07951  0.36076  0.212    0.832
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20    0.6632     1.5078    0.3311     1.328
## mediaOP50    1.0795     0.9263    0.5323     2.189
##
## Concordance= 0.557  (se = 0.046 )
## Rsquare= 0.035   (max possible= 0.995 )
## Likelihood ratio test= 2.2  on 2 df,   p=0.3
## Wald test            = 2.13  on 2 df,   p=0.3
## Score (logrank) test = 2.16  on 2 df,   p=0.3
```

```r
survdiff(Survobj ~media, data = d2, rho = 0)
```

```
## Call:
## survdiff(formula = Survobj ~ media, data = d2, rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 19       14     9.32  2.355719   3.72463
## media=FG20 23       13    17.77  1.282531   2.84312
## media=OP50 20       11    10.91  0.000743   0.00122
##
##  Chisq= 4.3  on 2 degrees of freedom, p= 0.1
```

```r
survdiff(Survobj ~media, data = d2, rho = 1)
```

```
## Call:
## survdiff(formula = Survobj ~ media, data = d2, rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## media=FG13 19    10.58     7.27     1.510      2.95
## media=FG20 23     8.40    12.74     1.475      4.04
## media=OP50 20     8.98     7.96     0.131      0.27
##
##  Chisq= 4.6  on 2 degrees of freedom, p= 0.1
```

```r
fit_d2 <- coxph(Survobj ~media, data = d2)
summary(fit_d2)
```

```
## Call:
## coxph(formula = Survobj ~ media, data = d2)
```

```
##
##   n= 62, number of events= 38
##
##                coef exp(coef) se(coef)      z Pr(>|z|)
## mediaFG20 -0.8083    0.4456   0.3905 -2.070   0.0385 *
## mediaOP50 -0.4944    0.6099   0.4073 -1.214   0.2248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20    0.4456      2.244    0.2073     0.958
## mediaOP50    0.6099      1.639    0.2746     1.355
##
## Concordance= 0.603  (se = 0.039 )
## Rsquare= 0.066   (max possible= 0.989 )
## Likelihood ratio test= 4.23  on 2 df,   p=0.1
## Wald test            = 4.36  on 2 df,   p=0.1
## Score (logrank) test = 4.53  on 2 df,   p=0.1
```

Now we use the replicate dataset

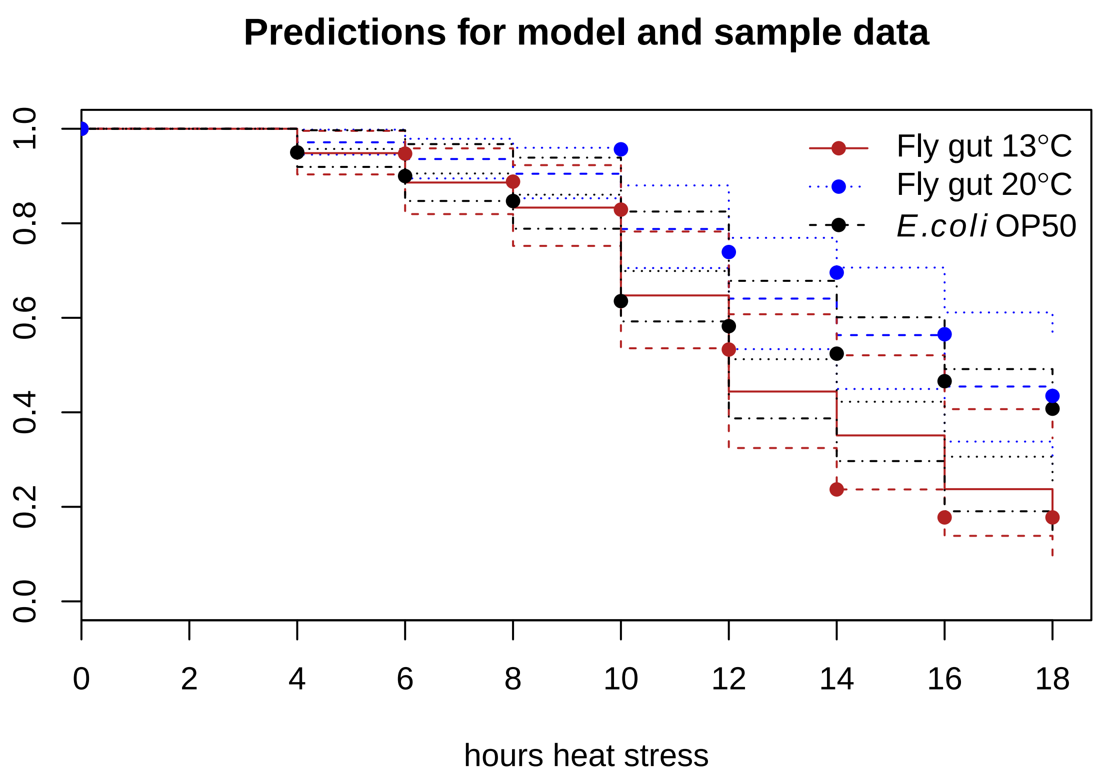
```
data_2 <- read.table(here('data', 'worm_second.csv'), header = TRUE,
                     sep = ';')
head(data_2)
```

```
##   ï..Time Group Metabolite Status
## 1       4  FG20          0      1
## 2       4  FG20          0      1
## 3       8  FG20          0      1
## 4      12  FG20          0      1
## 5       4  FG20          1      1
## 6       4  FG20          1      1
```

```
colnames(data_2) <- c('Time','media','Metabolite','Status')
data_2_rep <- subset(data_2, Metabolite == 0)
```

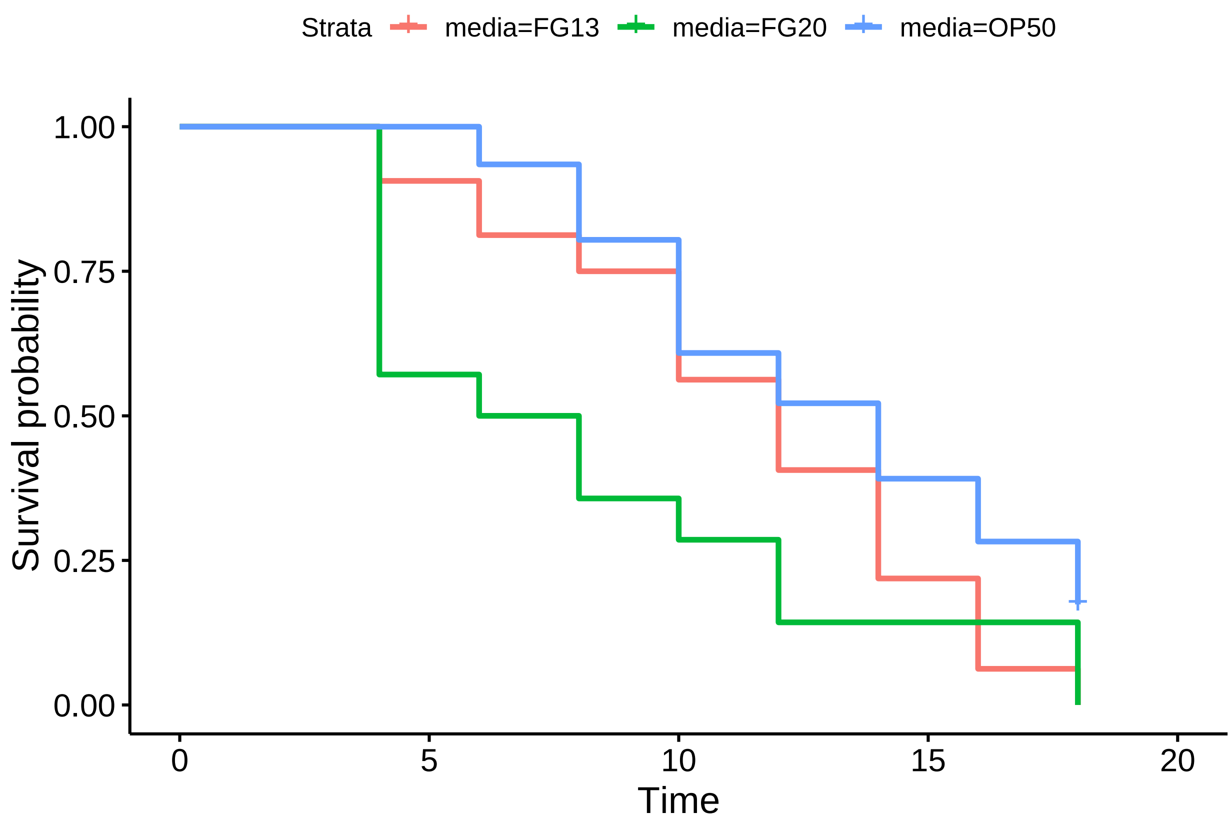
First we try to see if we can extend the cox model from the earlier dataset to the new one to see hoow well we can extrapolate from one dataset to the other

```
plot(survfit(fit, newdata = data.frame(media = 'FG20')), lty = 2, xaxp = c(0,18,9), col = 'blue',
     main = 'Predictions for model and sample data',
     xlab = 'hours heat stress')
lines(survfit(fit, newdata = data.frame(media = 'FG13')), col = 'firebrick')
lines(survfit(fit, newdata = data.frame(media = 'OP50')), lty =3)
points(df_OP50$Time, df_OP50$Surv, col = 'black', pch = 16)
points(df_fly_13$Time,df_fly_13$Surv, col = 'firebrick', pch = 16)
points(df_fly_20$Time,df_fly_20$Surv, col = 'blue', pch = 16)
legend('topright', pch = 16,lty = c(1,3,2),
       col = c('firebrick','blue','black'),
       legend = c(expression('Fly gut 13'*degree*C),
                  expression('Fly gut 20'*degree*C),
                  expression(italic(E.coli) ~ OP50)), bty = 'n')
```



Here the data fits well if we do the same for the new data set.

```
data_2_rep$Survobj <- with(data_2_rep,
                           Surv(data_2_rep$Time, event = data_2_rep$Status))
km_2 <- survfit(Survobj ~ media, data = data_2_rep,conf.type = "log-log",
              error = "greenwood")
s_km_2 <- summary(km_2)
ggsurvplot(km_2)
```




```
fit_2 <- coxph(Survobj ~ media, data = data_2_rep)
summary(fit_2)
```

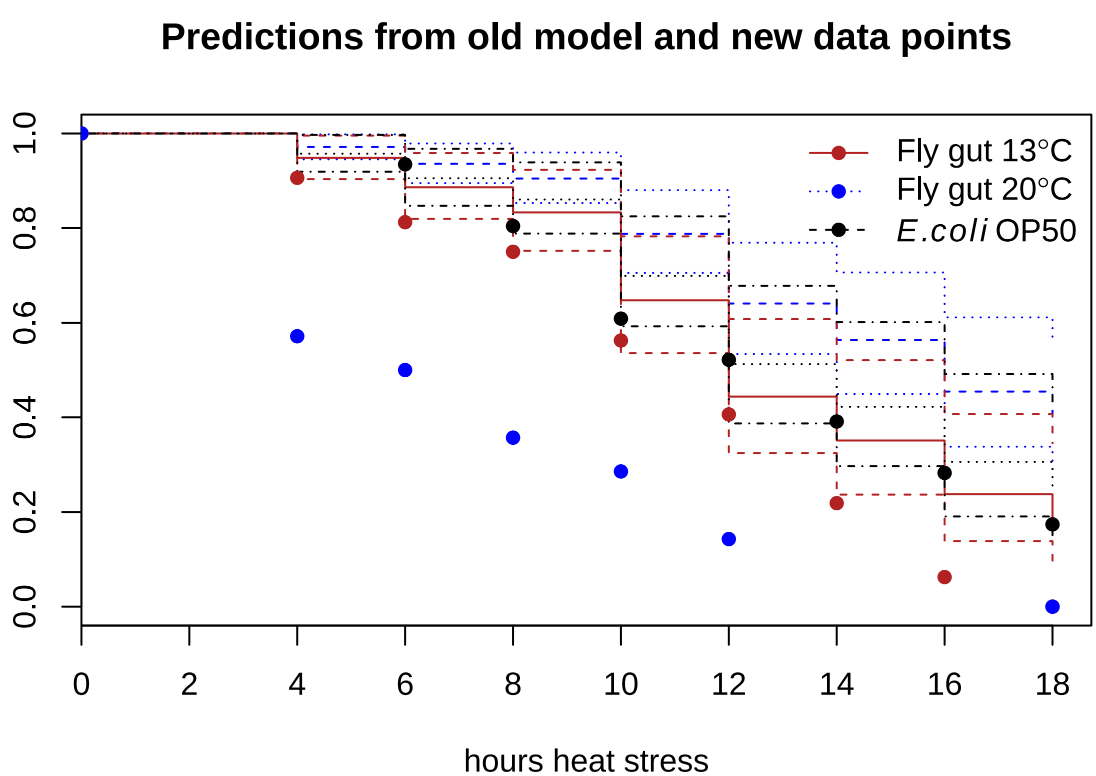
```
## Call:
## coxph(formula = Survobj ~ media, data = data_2_rep)
##
##   n= 92, number of events= 84
##
##                coef exp(coef) se(coef)      z Pr(>|z|)
## mediaFG20  0.4415    1.5551   0.3237  1.364   0.1726
## mediaOP50 -0.5814    0.5591   0.2467 -2.357   0.0184 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mediaFG20    1.5551     0.6431    0.8246    2.9328
## mediaOP50    0.5591     1.7886    0.3447    0.9068
##
## Concordance= 0.618  (se = 0.036 )
## Rsquare= 0.118   (max possible= 0.999 )
## Likelihood ratio test= 11.54  on 2 df,   p=0.003
## Wald test            = 12.18  on 2 df,   p=0.002
## Score (logrank) test = 12.87  on 2 df,   p=0.002
```

We see that Fly gut 20 has extremely bad survival and our former predictions do not hold for the new dataset.

If we try to plot the new data for old predictions

```
df_fly_13_2 <- data.frame(c(0,s_km_2$time[1:8]),c(1,s_km_2$surv[1:8]),
                        c(0,s_km_2$std.err[1:8]))
colnames(df_fly_13_2) <- c('Time','Surv','Std.error')
df_fly_20_2 <- data.frame(c(0,s_km_2$time[9:14]),c(1,s_km_2$surv[9:14]),
                        c(0,s_km_2$std.err[9:14]))
colnames(df_fly_20_2) <- c('Time','Surv','Std.error')
df_OP50_2 <- data.frame(c(0,s_km_2$time[15:21]),c(1,s_km_2$surv[15:21]),
                        c(0,s_km_2$std.err[15:21]))
colnames(df_OP50_2) <- c('Time','Surv','Std.error')
```

```
plot(survfit(fit, newdata = data.frame(media = 'FG20')), lty = 2, xaxp = c(0,18,9), col = 'blue',
     main = 'Predictions from old model and new data points',
     xlab = 'hours heat stress')
lines(survfit(fit, newdata = data.frame(media = 'FG13')), col = 'firebrick')
lines(survfit(fit, newdata = data.frame(media = 'OP50')), lty =3)
points(df_OP50_2$Time, df_OP50_2$Surv, col = 'black', pch = 16)
points(df_fly_13_2$Time,df_fly_13_2$Surv, col = 'firebrick', pch = 16)
points(df_fly_20_2$Time,df_fly_20_2$Surv, col = 'blue', pch = 16)
legend('topright', pch = 16,lty = c(1,3,2),
       col = c('firebrick','blue','black'),
       legend = c(expression('Fly gut 13'*degree*C),
                  expression('Fly gut 20'*degree*C),
                  expression(italic(E.coli) ~ OP50)), bty = 'n')
```



Both FG13 and OP50 is slightly lower than before but the survival in F20 has cratered and this may have happened because of something which liquefied the worms on the plates.

Then we compare data with standard errors with one group in each plot

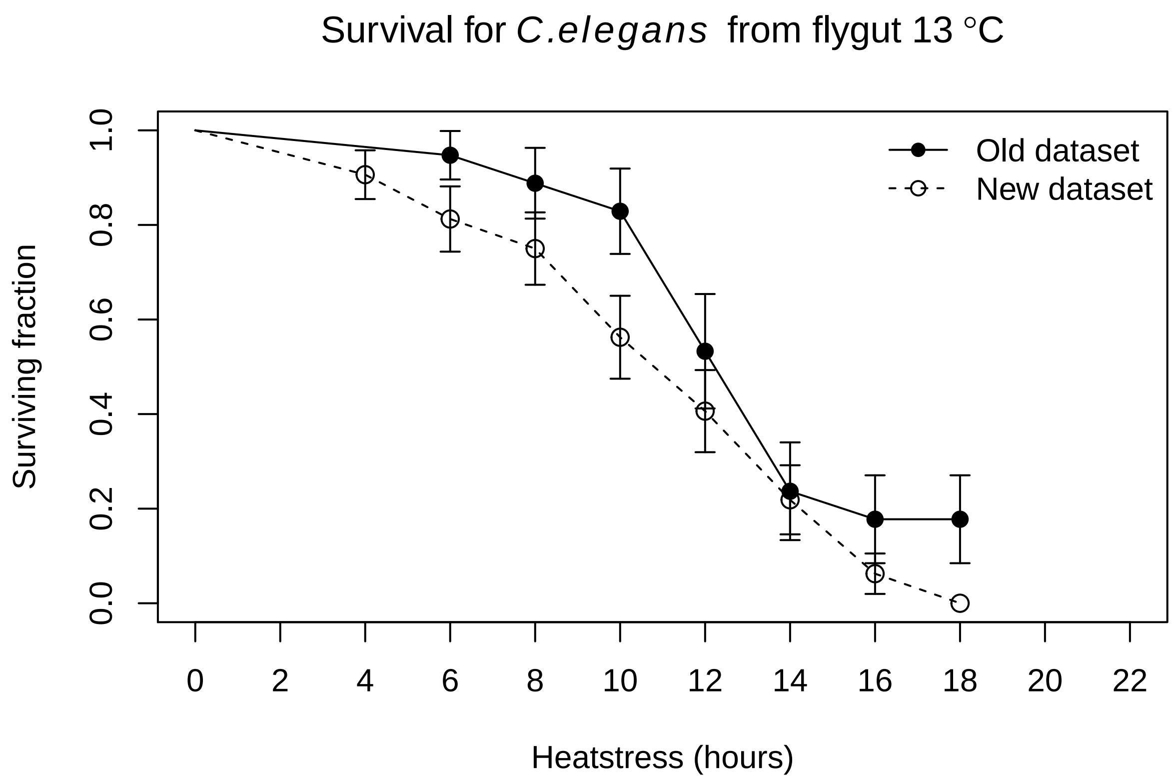
```
plot(df_fly_13$Time[2:9],df_fly_13$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival for' ~ italic(C.elegans) ~' from flygut 13' ~degree*C),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_fly_13$Time,df_fly_13$Surv)
arrows(df_fly_13$Time, df_fly_13$Surv-df_fly_13$Std.error, df_fly_13$Time,
       df_fly_13$Surv+df_fly_13$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_13$Time, df_fly_13$Surv - df_fly_13$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
points(df_fly_13_2$Time[2:9],df_fly_13_2$Surv[2:9], pch = 1, cex = 1.2)
lines(df_fly_13_2$Time,df_fly_13_2$Surv, lty = 2)
arrows(df_fly_13_2$Time, df_fly_13_2$Surv-df_fly_13_2$Std.error, df_fly_13_2$Time, df_fly_13_2$Surv+df_
```

```
## Warning in arrows(df_fly_13_2$Time, df_fly_13_2$Surv -
## df_fly_13_2$Std.error, : zero-length arrow is of indeterminate angle and so
## skipped
```

```
legend('topright', pch = c(16,1),lty = c(1,2),
       legend = c('Old dataset', 'New dataset'), bty = 'n')
```



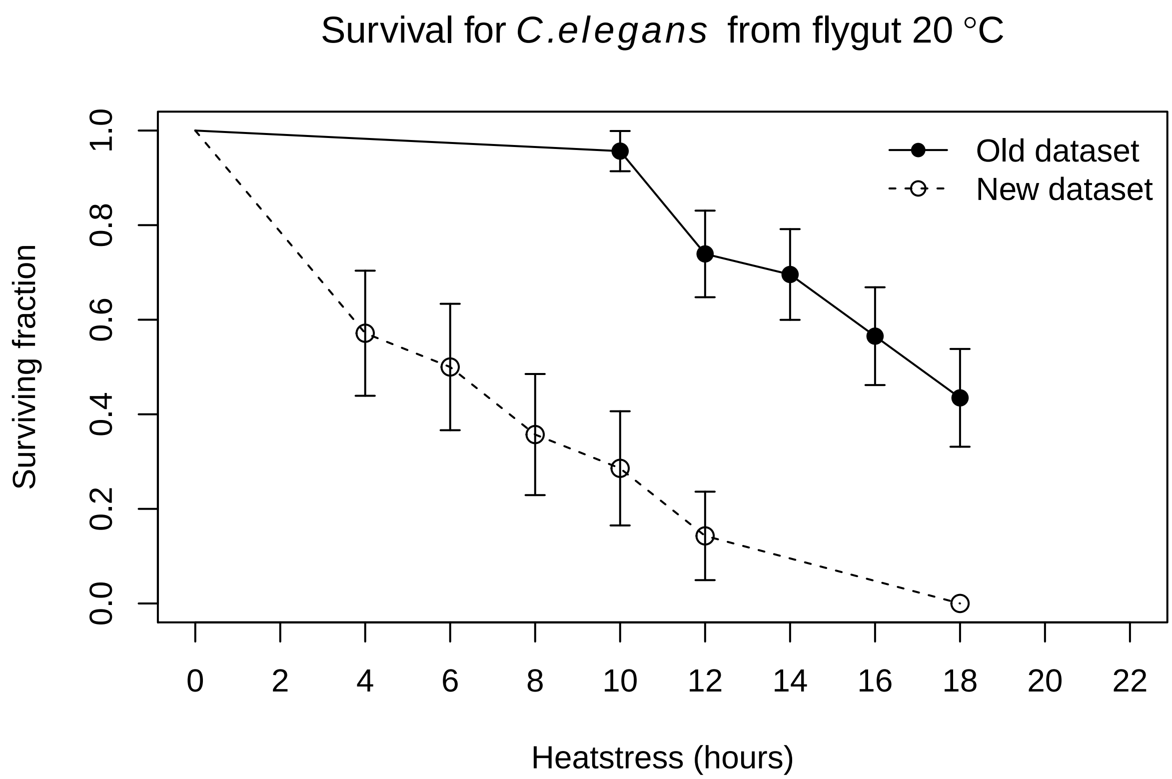
```r
plot(df_fly_20$Time[2:9],df_fly_20$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival for' ~ italic(C.elegans) ~' from flygut 20' ~degree*C),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_fly_20$Time,df_fly_20$Surv)
arrows(df_fly_20$Time, df_fly_20$Surv-df_fly_20$Std.error, df_fly_20$Time,
       df_fly_20$Surv+df_fly_20$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_fly_20$Time, df_fly_20$Surv - df_fly_20$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```r
points(df_fly_20_2$Time[2:9],df_fly_20_2$Surv[2:9], pch = 1, cex = 1.2)
lines(df_fly_20_2$Time,df_fly_20_2$Surv, lty = 2)
arrows(df_fly_20_2$Time, df_fly_20_2$Surv-df_fly_20_2$Std.error, df_fly_20_2$Time, df_fly_20_2$Surv+df_
```

```
## Warning in arrows(df_fly_20_2$Time, df_fly_20_2$Surv -
## df_fly_20_2$Std.error, : zero-length arrow is of indeterminate angle and so
## skipped
```

```r
legend('topright', pch = c(16,1),lty = c(1,2),
       legend = c('Old dataset', 'New dataset'), bty = 'n')
```



```r
plot(df_OP50$Time[2:9],df_OP50$Surv[2:9], pch = 16, cex = 1.2, xlab = 'Heatstress (hours)',
     ylab = 'Surviving fraction',
     main = expression('Survival for' ~ italic(C.elegans) ~ 'from' ~ italic(E.coli) ~ OP50),
     xlim = c(0,22), ylim = c(0,1), xaxp = c(0,22,11))
lines(df_OP50$Time,df_OP50$Surv)
arrows(df_OP50$Time, df_OP50$Surv-df_OP50$Std.error, df_OP50$Time,
       df_OP50$Surv+df_OP50$Std.error, length=0.05, angle=90, code=3, col = 'black')
```

```
## Warning in arrows(df_OP50$Time, df_OP50$Surv - df_OP50$Std.error,
## df_OP50$Time, : zero-length arrow is of indeterminate angle and so skipped
```

```r
points(df_OP50_2$Time[2:9],df_OP50_2$Surv[2:9], pch = 1, cex = 1.2)
lines(df_OP50_2$Time,df_OP50_2$Surv, lty = 2)
arrows(df_OP50_2$Time, df_OP50_2$Surv-df_OP50_2$Std.error, df_OP50_2$Time, df_OP50_2$Surv+df_OP50_2$Std
```

```
## Warning in arrows(df_OP50_2$Time, df_OP50_2$Surv - df_OP50_2$Std.error, :
## zero-length arrow is of indeterminate angle and so skipped
```

```r
legend('topright', pch = c(16,1),lty = c(1,2),
       legend = c('Old dataset', 'New dataset'), bty = 'n')
```