2261 HW3

Orly Olbum

3/9/2021

Problem 1

What proportion of the patients have censored survival times?

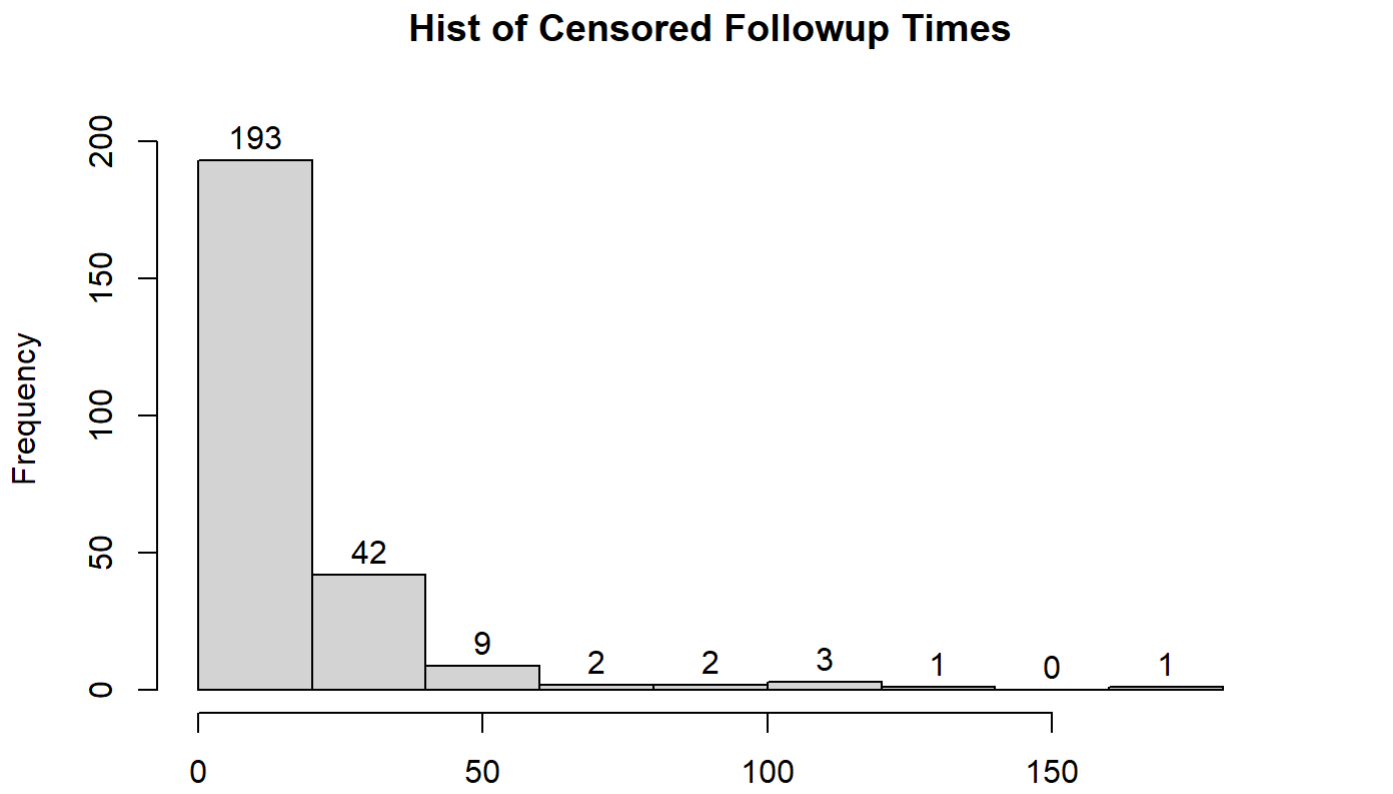
```
support$censored[support$slos == support$d.time] = 1
support$censored[support$slos != support$d.time] = 0
sum(support$censored)/length(support$censored)
```

```
## [1] 0.253
```

There are 253 censored survival times out of 1000 observations, so the proportion of censored observations is .253.

Make a histogram or empirical distribution function estimate of the censored followup times.

```
hist(support$d.time[support$censored == 1], main = "Hist of Censored Followup Times", xlab = " "
, labels = TRUE, ylim = c(0, 205))
```



What is the typical follow-up duration for a patient in the study who has survived so far?

Find the KM estimate for follow-up days with the newly defined “censor” as the indicator for censorship, then calculate average survival time.

```
fit1 = survfit(Surv(slos, censored) ~ 1, data = support, type = "kaplan-meier")
print(fit1, print.rmean = TRUE)
```

```
## Call: survfit(formula = Surv(slos, censored) ~ 1, data = support, type = "kaplan-meier")
##
##          n     events     *rmean *se(rmean)     median    0.95LCL    0.95UCL
##    1000.00     253.00      90.20       9.03      52.00      37.00      87.00
##     * restricted mean with upper limit =  241
```

The mean survival time for patients who have survived so far is 90.20 days.

What is the typical survival time for patients who have died?

```
fit2 = survfit(Surv(slos, death) ~ 1, data = support, type = "kaplan-meier")
print(fit2, print.rmean = TRUE)
```

```
## Call: survfit(formula = Surv(slos, death) ~ 1, data = support, type = "kaplan-meier")
##
##          n     events     *rmean *se(rmean)     median    0.95LCL    0.95UCL
##    1000.00     668.00      32.30       2.39      15.00      14.00      17.00
##     * restricted mean with upper limit =  241
```

Using death as an indicator rather than censoring, the mean survival time is 32.30 days.

Taking censoring into account, what is the median survival time from the Kaplan-Meier estimate of the overall survival function?

```
fit1
```

```
## Call: survfit(formula = Surv(slos, censored) ~ 1, data = support, type = "kaplan-meier")
##
##       n  events  median 0.95LCL 0.95UCL
##    1000     253      52      37      87
```

Using censoring as an indicator (overall data), the median survival time is 52 days.

Problem 2a

Using the data reported in Section 1.3, find the quantities specified below for the AML low risk group. Note that for this question, we should use the “t1” (Time To Death Or On Study Time) and “dind” (death indicator).

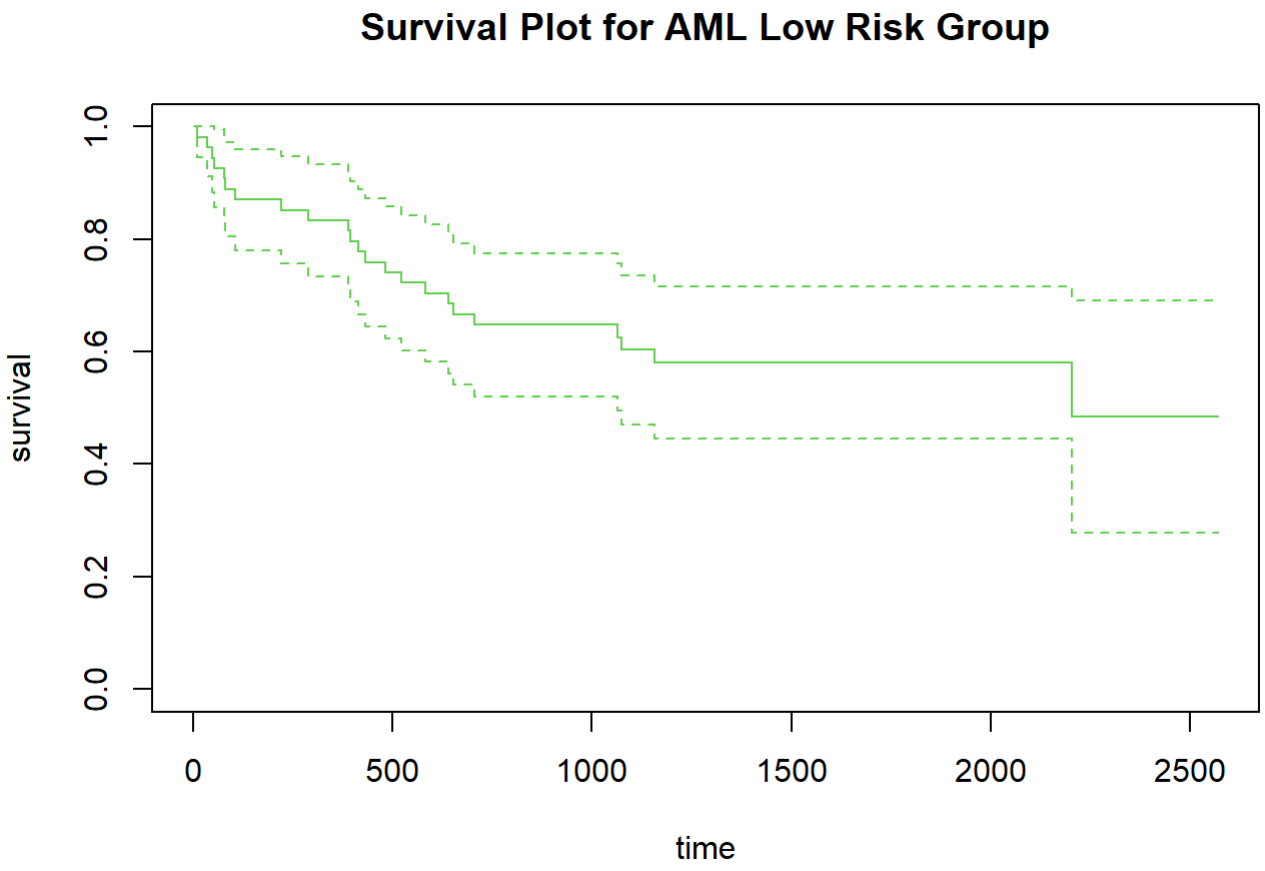
(a)

Estimate the survival functions and their standard errors (Greenwood’s formula) for the AML low risk group.

```
model1 = survfit(Surv(t1, dind) ~ 1, data = aml_low, type = "kaplan-meier", error = "greenwood",
                 conf.int = 0.95, conf.type = "plain")
summary(model1)
```

```
## Call: survfit(formula = Surv(t1, dind) ~ 1, data = aml_low, error = "greenwood",
##     type = "kaplan-meier", conf.int = 0.95, conf.type = "plain")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    10     54       1    0.981  0.0183        0.946        1.000
##    35     53       1    0.963  0.0257        0.913        1.000
##    48     52       1    0.944  0.0312        0.883        1.000
##    53     51       1    0.926  0.0356        0.856        0.996
##    79     50       1    0.907  0.0394        0.830        0.985
##    80     49       1    0.889  0.0428        0.805        0.973
##   105     48       1    0.870  0.0457        0.781        0.960
##   222     47       1    0.852  0.0483        0.757        0.947
##   288     46       1    0.833  0.0507        0.734        0.933
##   390     45       1    0.815  0.0529        0.711        0.918
##   393     44       1    0.796  0.0548        0.689        0.904
##   414     43       1    0.778  0.0566        0.667        0.889
##   431     42       1    0.759  0.0582        0.645        0.873
##   481     41       1    0.741  0.0596        0.624        0.858
##   522     40       1    0.722  0.0610        0.603        0.842
##   583     39       1    0.704  0.0621        0.582        0.825
##   641     38       1    0.685  0.0632        0.561        0.809
##   653     37       1    0.667  0.0642        0.541        0.792
##   704     36       1    0.648  0.0650        0.521        0.776
##  1063     29       1    0.626  0.0665        0.496        0.756
##  1074     28       1    0.603  0.0678        0.471        0.736
##  1156     27       1    0.581  0.0688        0.446        0.716
##  2204      6       1    0.484  0.1054        0.278        0.691
```

```
plot(model1, main = "Survival Plot for AML Low Risk Group", xlab = "time", ylab = "survival", co
l = 3)
```



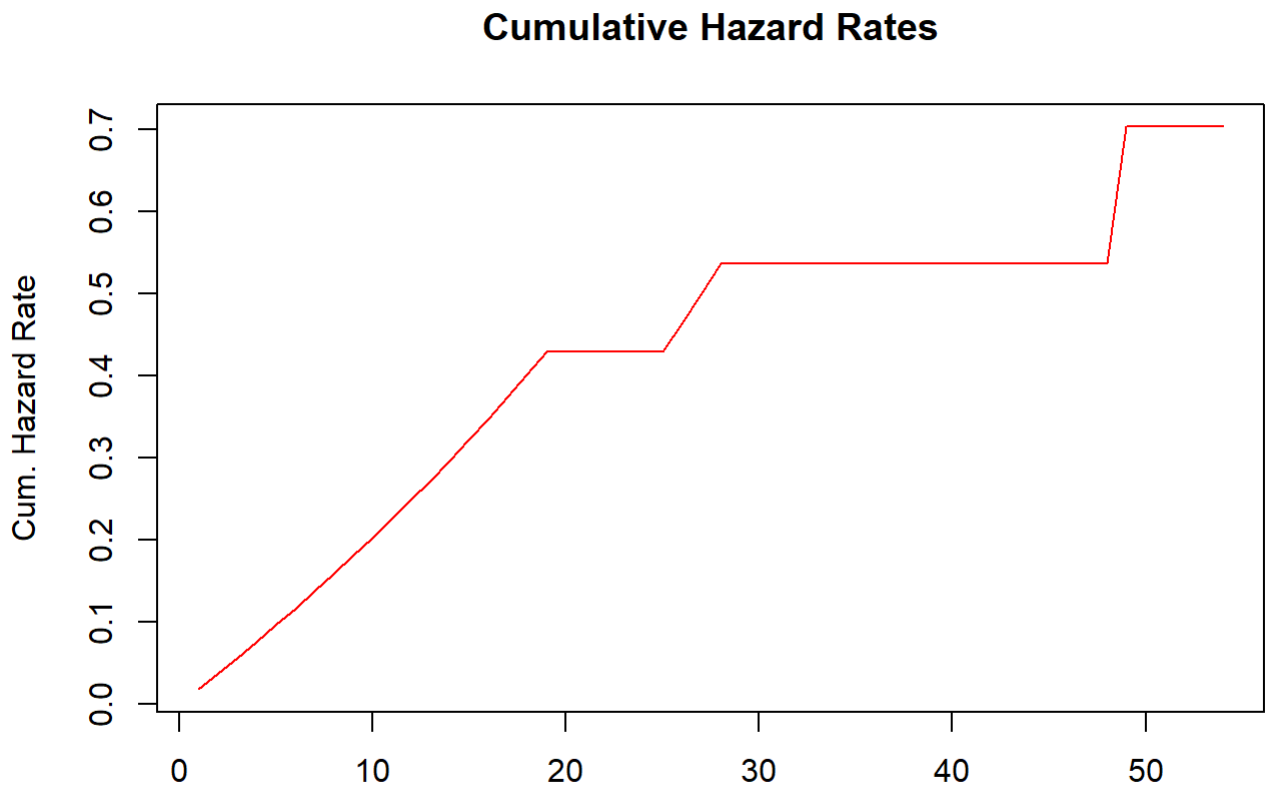
(b)

Estimate the cumulative hazard rates and their standard errors (Aalen's formula) for the AML low risk group.

```
# H_est1 = -log(fit1$surv) # KM cum. hazard rate
h = model1$n.event / model1$n.risk
cum.haz = cumsum(h)
cum.haz
```

```
##  [1] 0.01851852 0.03738644 0.05661721 0.07622506 0.09622506 0.11663322
##  [7] 0.13746655 0.15874315 0.18048228 0.20270450 0.22543177 0.24868759
## [13] 0.27249711 0.29688735 0.32188735 0.34752838 0.37384417 0.40087120
## [19] 0.42864897 0.42864897 0.42864897 0.42864897 0.42864897 0.42864897
## [25] 0.42864897 0.46313173 0.49884602 0.53588306 0.53588306 0.53588306
## [31] 0.53588306 0.53588306 0.53588306 0.53588306 0.53588306 0.53588306
## [37] 0.53588306 0.53588306 0.53588306 0.53588306 0.53588306 0.53588306
## [43] 0.53588306 0.53588306 0.53588306 0.53588306 0.53588306 0.53588306
## [49] 0.70254972 0.70254972 0.70254972 0.70254972 0.70254972 0.70254972
```

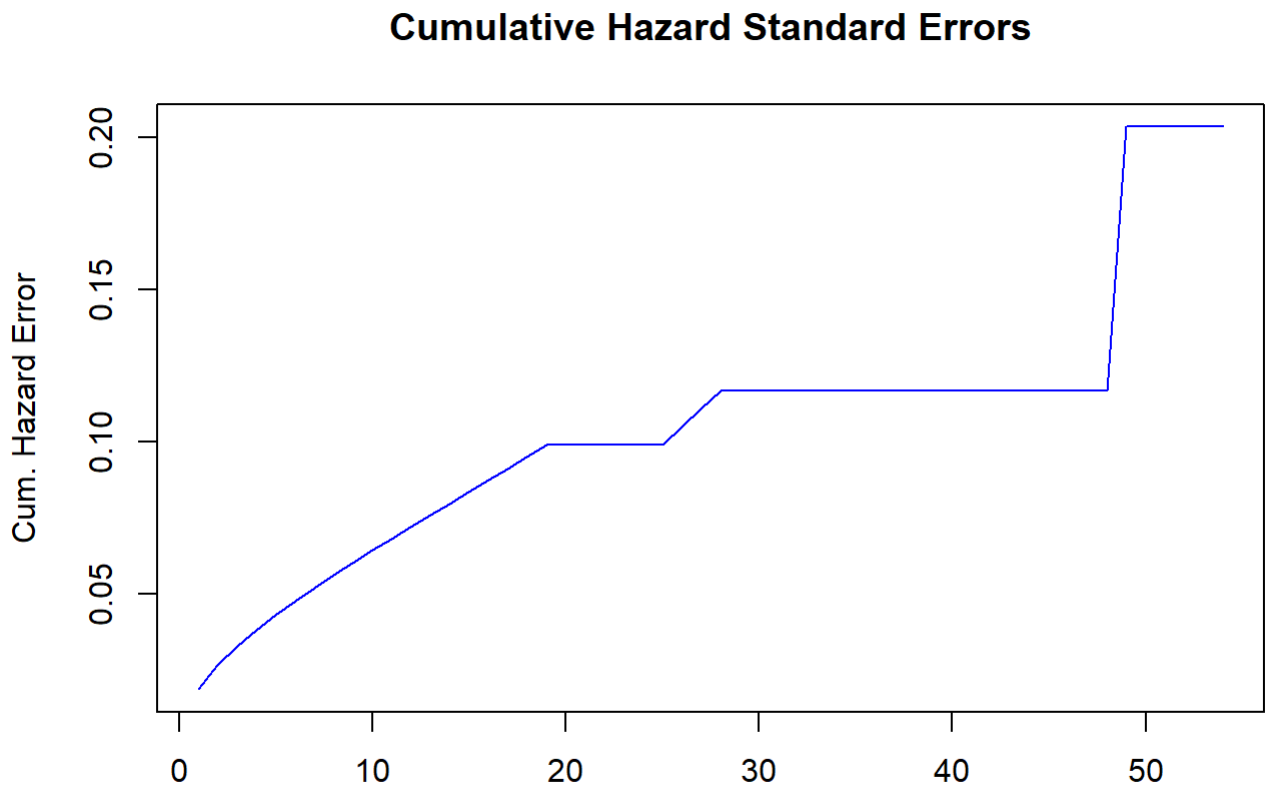
```
plot(cum.haz, main = "Cumulative Hazard Rates", xlab = "", ylab = "Cum. Hazard Rate", type = "l"
, col = "red")
```




```
# NA standard errors
h.er = model1$n.event / model1$n.risk^2
cum.h.er = sqrt(cumsum(h.er)) # formula is for variances, take square root for standard errors
cum.h.er
```

```
##  [1] 0.01851852 0.02643736 0.03269184 0.03812118 0.04304909 0.04764155
##  [7] 0.05199755 0.05618219 0.06024142 0.06420947 0.06811303 0.07197373
## [13] 0.07580970 0.07963664 0.08346853 0.08731814 0.09119747 0.09511802
## [19] 0.09909108 0.09909108 0.09909108 0.09909108 0.09909108 0.09909108
## [25] 0.09909108 0.10491951 0.11083147 0.11685613 0.11685613 0.11685613
## [31] 0.11685613 0.11685613 0.11685613 0.11685613 0.11685613 0.11685613
## [37] 0.11685613 0.11685613 0.11685613 0.11685613 0.11685613 0.11685613
## [43] 0.11685613 0.11685613 0.11685613 0.11685613 0.11685613 0.11685613
## [49] 0.20355130 0.20355130 0.20355130 0.20355130 0.20355130 0.20355130
```

```
plot(cum.h.er, main = "Cumulative Hazard Standard Errors", xlab = "", ylab = "Cum. Hazard Error"
, type = "l", col = "blue")
```




```
# another way to do the exact same thing
# ch = nelsonaalen(aml_low, t1, dind)
# plot(aml_low$t1, ch, main = "Cum. Hazard", xlab = "t1", ylab = "hazard rate", pch = 20, col =
 2)
```

(c)

Estimate the mean time to death and find 95% linear confidence interval for the mean survival time for the AML low risk group.

```
model.mean = unname(summary(model1)$table[5])
model.se = unname(summary(model1)$table[6])
lo = model.mean - 1.96*model.se
hi = model.mean + 1.96*model.se
ci = c(lo, hi)
model.mean; ci
```

```
## [1] 1644.645
```

```
## [1] 1356.353 1932.938
```

```
# round(summary(model1)$table, 2)
```

Mean time to death is 1644.645 days, and a 95% confidence interval for the mean is (1356.353, 1932.938).

(d)

Work out estimates of the median time to death and find 95% log-transformed confidence interval for the median survival time for the AML low risk group.

```
model1
```

```
## Call: survfit(formula = Surv(t1, dind) ~ 1, data = aml_low, error = "greenwood",
##     type = "kaplan-meier", conf.int = 0.95, conf.type = "plain")
##
##       n  events  median 0.95LCL 0.95UCL
##      54      23    2204    1063      NA
```

The median time to death is 2204 days, with a lower bound of 1063 days and no upper bound.

Problem 3

Using the data reported in Section 1.3 for the ALL group, find the following quantities for the two competing risks of relapse and death in remission:

(a)

The estimated cumulative incidence in 500 days.

```
cum.inc = cuminc(ftime = all$t1, fstatus = all$dind, group = all$rind)
cum.inc
```

```
## Tests:
##       stat          pv df
## 1 8.107552 0.004408121  1
## Estimates and Variances:
## $est
##           500      1000      1500      2000
## 0 1 0.4331984 0.4736842 0.4736842 0.4736842
## 1 1 0.7500000 0.9166667        NA        NA
##
## $var
##            500       1000       1500       2000
## 0 1 0.01029957 0.01051986 0.01051986 0.01051986
## 1 1 0.01921658 0.01081573         NA         NA
```

The cumulative incidence in 500 days is 0.433 and 0.75 for the relapse and death events, respectively.

(b)

The estimated errors of the two estimators in part (a).

```
sqrt(0.01029957)
```

```
## [1] 0.1014868
```

```
sqrt(0.01921658)
```

```
## [1] 0.1386239
```

The estimated errors from (a) are 0.101 and 0.139, respectively.

Problem 4

Two hundred and fifty items of an electronic device were used to test their failure times under excessive heat conditions, where 212 items failed within a week and the remaining 44 items were censored at the end of the week. The Nelson-Aalen estimate of the cumulative hazard function is given in Figure 1 and of the parabolic shape. Based on the results from one-sample log-rank tests given in Table 1, draw your conclusion about the distribution of the failure time for this device.

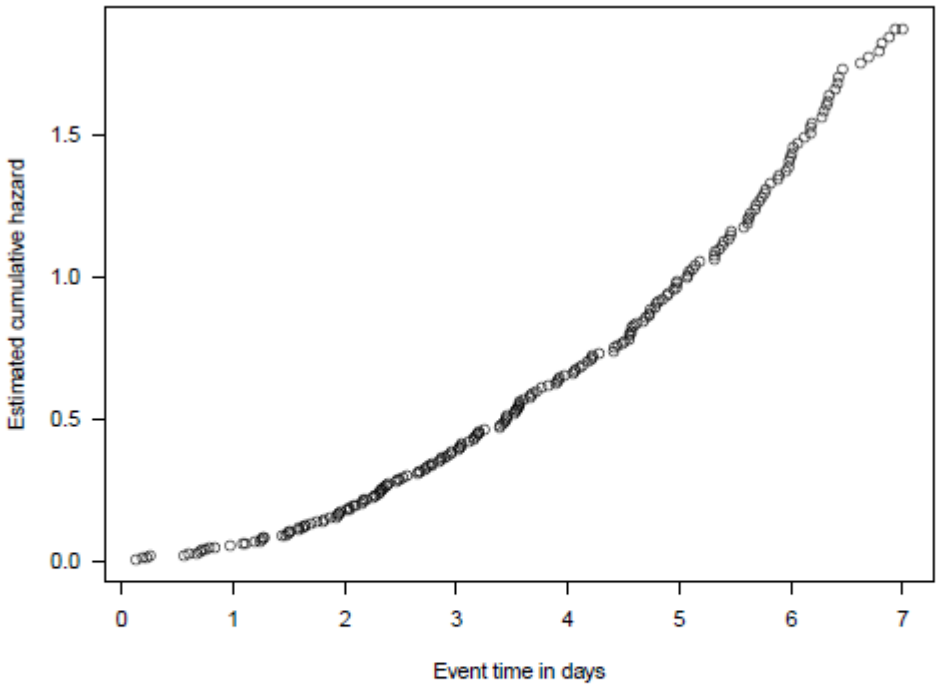


Figure 1: *The Nelson-Aalen estimate of the cumulative hazard function*

Table 1: Test results based on standardized $Z(\tau)$ defined in eq(7.2.1)

H_0	$T_1 = \frac{Z(\tau)}{\sqrt{V[Z(\tau)]}}$	p-value
$H_0(t) = 0.1^2t^2$	21.6	< 0.0001
$H_0(t) = 0.2^2t^2$	0.16	0.873
$H_0(t) = 0.3^2t^2$	-12.3	< 0.0001

With a parabolic Cumulative Hazard function plot and two significant results from the test shown, the failure time distribution for this device is likely exponential distribution.