In an analysis, we first explore the data. Hereby we are also interested in estimating the survival function S from right censored data.

We consider again the previous example.

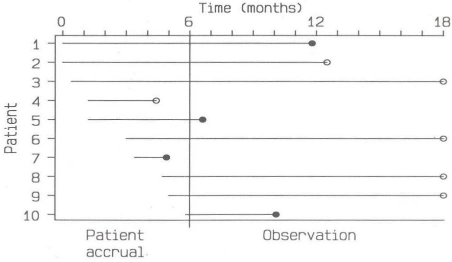


Figure 13.1 Diagram showing patients entering a study at different times and the observation of known (●) and censored (○) survival times.

We extract the following data from the graph.

Patient	Random. time (months)	Last obs. (months)	Died?	lifetime (months)	status (1 = yes)
1	0.0	11.8	yes	11.8	1
2	0.0	12.5	no	12.5	0
3	0.4	18.0	no	17.6	0
4	1.2	4.4	no	3.2	0
5	1.2	6.6	yes	5.4	1
6	3.0	18.0	no	15.0	0
7	3.4	4.9	yes	1.5	1
8	4.7	18.0	no	13.3	0
9	5.0	18.0	no	13.0	0
10	5.8	10.1	yes	4.3	1

Example: We want to know the probability that a patient has died at 6 months?

If we ignore censoring, we use the empirical distribution function and estimate

$$\hat{p}(6) = \frac{1}{10} \sum_{i=1}^{10} I(\text{lifetime}_i \leq 6) = \frac{4}{10}.$$

However, one patient (no. 4) has survived 3.2 months (censored observation).

How to deal with this patient?

- Assume no. 4 died before 6 months $\rightarrow$ overestimation ($\frac{4}{10}$)
- Assume no. 4 survived 6 months $\rightarrow$ underestimation ($\frac{3}{10}$)
- Ignore no. 4 $\rightarrow$ loss of information ($\frac{3}{9}$)

Hence, censoring is causing problems!

To derive an estimator for the survival distribution of the lifetime, we introduce the following notation:

- The j-th time interval is $[a_{j-1}, a_j[$, $j = 1, \ldots, K+1$. $a_0 = 0$ and $a_{K+1} = +\infty$.
- d_j is # failures in the j-th interval.
- c_j is # censored observations in the j-th interval.
- n_j is # individuals entering the j-th interval.

We decompose the survival function $S(a_j)$ as

$$
\begin{aligned}
S(a_j) &= P(T > a_j | T > a_{j-1}) S(a_{j-1}) \\
&= P(T > a_j | T > a_{j-1}) \dots P(T > a_1 | T > a_0) S(a_0)
\end{aligned}
$$

For each interval $[a_{j-1}, a_j[$, we estimate the conditional survival probability by

$$
\begin{aligned}
P(T > a_j | T > a_{j-1}) &= P(\text{Survive interval } [a_{j-1}, a_j[| T > a_{j-1}) \\
&= 1 - \frac{\text{Number failures}_j}{\text{Number at risk}_j}.
\end{aligned}
$$

Some questions:

1. How do we define the number at risk n'_j in the j-th interval?
2. What should we do with censored people?

Answer:

We assume that the censoring times are uniformly distributed over the interval, and hence occur on average half-way through the interval. For each interval, we define the number at risk n'_j as

$$n'_j = n_j - \frac{c_j}{2}.$$

We call the obtained estimator, the Actuarial estimator.

Until now, we assumed that the individual lifetimes were observed. Sometimes they are grouped in fixed time intervals.

Some examples:

- Cohort life table: presents the actual mortality experience from birth to death for a group of people born at about the same time.
- Current life table: From a cross-sectional study, mostly census information, the number of individuals alive at each age is recorded, together with statistics on number of deaths in each age group.
- Clinical life table: applies to grouped survival data from studies in patients with a specific disease.

An example: Time to weaning of breast-fed newborns

- In the National Labor Survey of Youth (NLSY) data set, youths between age 14 and 21 were yearly interviewed.
- Females were asked about pregnancies and breast-feeding.
- Data on 927 breast-fed first-born children.
- Duration of breast-feeding was recorded in weeks. An indicator whether breast-feeding was completed.

Weeks	n_j	c_j	d_j
$[0, 2[$	927	2	77
$[2, 3[$	848	3	71
$[3, 5[$	774	6	119
$[5, 7[$	649	9	75
$[7, 11[$	565	7	109
$[11, 17[$	449	5	148
$[17, 25[$	296	3	107
$[25, 37[$	186	0	74
$[37, 53[$	112	0	85
$[53, \infty[$	27	0	27

Using SAS software

```
data wean;
input Weeks Status Freq;
cards;
1 1 77
2.5 1 71
4 1 119
6 1 75
8 1 109
15 1 148
20 1 107
30 1 74
40 1 85
60 1 27
1 0 2
2.5 0 3
4 0 6
6 0 9
8 0 7
15 0 5
20 0 3
30 0 0
40 0 0
60 0 0 ;
```

```
proc lifetest data=wean method=lt intervals=(0 2 3 5 7 11 17 25 37 53);
time Weeks*Status(0);
freq Freq;
run;
```

The LIFETEST Procedure

Life Table Survival Estimates

Interval [Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure
0	2	77	2	926.0	0.0832	0.00907	1.0000	0
2	3	71	3	846.5	0.0839	0.00953	0.9168	0.0832
3	5	119	6	771.0	0.1543	0.0130	0.8399	0.1601
5	7	75	9	644.5	0.1164	0.0126	0.7103	0.2897
7	11	109	7	561.5	0.1941	0.0167	0.6276	0.3724
11	17	148	5	446.5	0.3315	0.0223	0.5058	0.4942
17	25	107	3	294.5	0.3633	0.0280	0.3381	0.6619
25	37	74	0	186.0	0.3978	0.0359	0.2153	0.7847
37	53	85	0	112.0	0.7589	0.0404	0.1296	0.8704
53	.	27	0	27.0	1.0000	0	0.0313	0.9687

Evaluated at the Midpoint of the Interval

Interval [Lower,	Upper)	Survival Standard Error	Median Residual Lifetime	Median Standard Error	PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	2	0	11.2078	0.5880	0.0416	0.00454	0.04338	0.004939
2	3	0.00907	10.6957	0.5639	0.0769	0.00877	0.087546	0.01038
3	5	0.0121	11.0717	0.5413	0.0648	0.00554	0.083626	0.007639
5	7	0.0149	11.3915	0.5006	0.0413	0.00457	0.061779	0.00712
7	11	0.0160	11.5839	0.8624	0.0305	0.00273	0.053748	0.005118
11	17	0.0166	11.5508	0.7793	0.0279	0.00209	0.066219	0.005335
17	25	0.0158	14.4748	1.3803	0.0154	0.00139	0.055498	0.005231
25	37	0.0138	15.5765	1.2836	0.00714	0.000790	0.041387	0.00466
37	53	0.0114	10.5412	0.9960	0.00615	0.000630	0.076439	0.00656
53	.	0.00591	.	.	.	.	.	.

Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	Percent Censored
927	892	35	3.78

NOTE: There were 3 observations with missing values, negative time values or frequency values less than 1.

The main quantity of interest is the probability that an event will not occur by time t:

$$S(t) = P(T > t).$$

Kaplan and Meier (1958) develop an estimator for the survival function

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)^{\delta_i} = \prod_{t_i \leq t} \left(\frac{n_i - d_i}{n_i}\right)^{\delta_i}.$$

where

- d_i = number of patients died at t_i
- n_i = number of patients at risk before t_i

This estimator is called Kaplan-Meier or Product-limit estimator.

Let $t_1 < t_2 < \ldots < t_k$ be the ordered lifetimes.

The main idea: conditional probability

To survive until time t_{j+1}, you need to first survive until time t_j units, and then until t_{j+1}.

Symbolically:

$$\begin{aligned} S(t_{j+1}) &= P(T > t_{j+1} | T > t_j) S(t_j) \\ &= P(\text{Survive interval }]t_j, t_{j+1}] | T > t_j) S(t_j). \end{aligned}$$

There are three possibilities for each interval:

- There is a censoring $\rightarrow$ We assume that they survive until the end of the interval. The conditional probability is 1.

- There is a death, but no censoring $\rightarrow$ conditional probability of surviving the interval is $1 - \frac{d}{r}$ where d is the number of deaths within the interval and r the number at risk at the beginning of the interval.

- There are tied deaths and censoring $\rightarrow$ We assume that censoring occurs at the end of the interval such that the conditional probability is $1 - \frac{d}{r}$.

Lifetime (Months)	Status	Number Events (d_i)	Number at risk (n_i)	$\hat{S}(t)$
0				1
1.5	1	1	10	$S(0)\,[1-1/10] = 0.9000$
3.2	0	0	9	$S(1.5) \times 1 = 0.9000$
4.3	1	1	8	$S(3.2)\,[1-1/8] = 0.7875$
5.4	1	1	7	$S(4.3)\,[1-1/7] = 0.6750$
11.8	1	1	6	$S(5.4)\,[1-1/6] = 0.5625$
12.5	0	0	5	$S(11.8) \times 1 = 0.5625$
13.0	0	0	4	$S(12.5) \times 1 = 0.5625$
13.3	0	0	3	$S(13.0) \times 1 = 0.5625$
15.0	0	0	2	$S(13.3) \times 1 = 0.5625$
17.6	0	0	1	$S(15.0) \times 1 = 0.5625$

Returning to the introduction, we note that the probability of surviving more than 6 months is

$$\hat{S}(6) = 67.5\%.$$

Comparing with some "naive" estimators.

- Assume patient with $t = 3.2$ died before 6 months $\rightarrow$ 60%.
- Assume patient with $t = 3.2$ survived 6 months $\rightarrow$ 70%.
- Ignore patient with $t = 3.2 \rightarrow 66.6\%$

Some properties for the KM-estimator

- The KM-estimator is a step-function which only jumps at uncensored observations. The different jumps are random, dependent on censored observations.

$$W_j = \hat{S}(t_{j-1}) - \hat{S}(t_j) = \frac{d_j}{n_j} \prod_{t_i \leq t_{j-1}} \left(1 - \frac{d_i}{n_i}\right)^{\delta_i}.$$

- When the largest observation t_k is censored, the KM-estimator does not converge to zero at infinity and is often taken as undefined.
- The KM-estimator is the nonparametric MLE of

$$L = \prod_{j=0}^{k} \left\{ \left[S(t_j^-) - S(t_j)\right]^{d_j} \prod_{l=1}^{m_j} S(t_{jl}) \right\}.$$

If there is no censoring, the KM-estimator reduces to the empirical survival function.

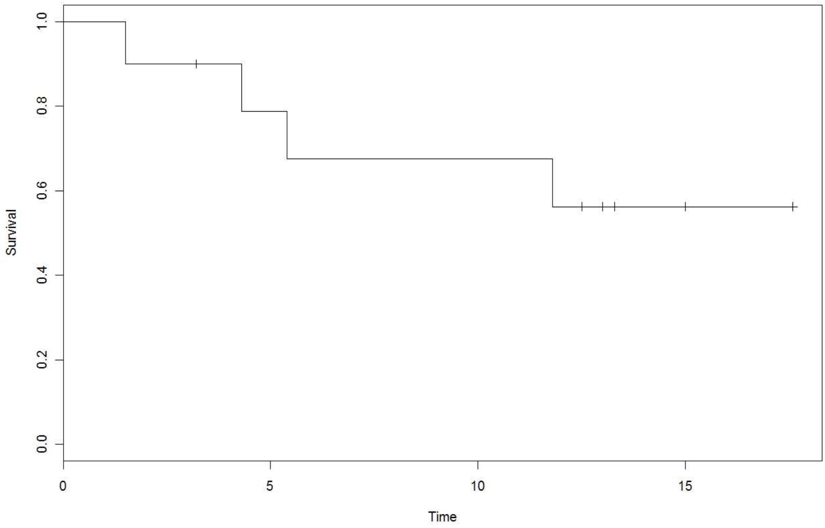
If $t_1 < t_2 < \ldots < t_k$ then $n_i = n - \sum_{j=1}^{i-1} d_j$.

Hence

$$
\begin{aligned}
\hat{S}(t) &= \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \\
&= \left(\frac{n - d_1}{n}\right)\left(\frac{n - d_1 - d_2}{n - d_1}\right)\ldots\left(\frac{n - d_1 - \ldots - d_i}{n - d_1 - \ldots - d_{i-1}}\right) \\
&= \frac{n - \sum_{j=1}^{i} d_j}{n} = \frac{\#\text{ observations } > t}{n}.
\end{aligned}
$$

Using R software

```
> library(survival)

> Time<-c(1.5,3.2,4.3,5.4,11.8,12.5,13.0,13.3,15.0,17.6)
> Status<-c(1,0,1,1,1,0,0,0,0,0)

> Surv(Time,Status)
 [1]  1.5   3.2+  4.3   5.4  11.8  12.5+ 13.0+ 13.3+ 15.0+ 17.6+

> survfit(Surv(Time,Status)~1)
Call: survfit(formula = Surv(Time, Status) ~ 1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
   10.0    10.0    10.0     4.0      NA     5.4      NA

> summary(survfit(Surv(Time,Status)~1))
Call: survfit(formula = Surv(Time, Status) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1.5     10       1    0.900  0.0949        0.732            1
  4.3      8       1    0.787  0.1340        0.564            1
  5.4      7       1    0.675  0.1551        0.430            1
 11.8      6       1    0.562  0.1651        0.316            1

> plot(survfit(Surv(Time,Status)~1,conf.type="none"),xlab="Time",ylab="Survival")
```



Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
Time

Using SAS software

```
data clin;
input Time Status;
cards;
1.5 1
3.2 0
...
17.6 0
;
run;

proc lifetest data=clin;
time Time*Status(0);
run;
```

The LIFETEST Procedure

Product-Limit Survival Estimates

Time	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	10
1.5000	0.9000	0.1000	0.0949	1	9
3.2000*	.	.	.	1	8
4.3000	0.7875	0.2125	0.1340	2	7
5.4000	0.6750	0.3250	0.1551	3	6
11.8000	0.5625	0.4375	0.1651	4	5
12.5000*	.	.	.	4	4
13.0000*	.	.	.	4	3
13.3000*	.	.	.	4	2
15.0000*	.	.	.	4	1
17.6000*	.	.	.	4	0

NOTE: The marked survival times are censored observations.

```
Summary Statistics for Time Variable Time

             Quartile Estimates

             Point     95% Confidence Interval
Percent    Estimate      [Lower      Upper)

     75        .         11.8000       .
     50        .          5.4000       .
     25      5.4000       1.5000       .


    Mean    Standard Error

  9.2063            1.4535

NOTE: The mean survival time and its standard error were underestimated because the largest
      observation was censored and the estimation was restricted to the largest event time.


Summary of the Number of Censored and Uncensored Values

                                  Percent
   Total  Failed    Censored    Censored

      10       4           6       60.00
```

An alternative estimator for the Kaplan-Meier estimator is the Nelson-Aalen estimator.

Hereby we used the link between the survival function and the cumulative hazard function,

$$S(t) = \exp(-\Lambda(t)).$$

Estimating first the cumulative hazard function, we then get another estimator for the survival function,

$$\hat{\Lambda}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i} \Rightarrow \hat{S}(t) = \exp\left(-\sum_{t_i \leq t} \frac{d_i}{n_i}\right).$$

```
proc lifetest data=clin nelson method=breslow;
time Time*Status(0);
run;

The LIFETEST Procedure

                    Survival Function and Cumulative Hazard Rate

                    Breslow                      Nelson-Aalen
                                 Survival                    Cum Haz
                                 Standard    Cumulative      Standard     Number     Number
Time          Survival  Failure   Error       Hazard          Error       Failed      Left

 0.0000       1.0000          0          0          0              .          0         10
 1.5000       0.9048     0.0952     0.0954     0.1000         0.1000          1          9
 3.2000*           .          .          .          .              .          1          8
 4.3000       0.7985     0.2015     0.1359     0.2250         0.1601          2          7
 5.4000       0.6922     0.3078     0.1590     0.3679         0.2146          3          6
11.8000       0.5859     0.4141     0.1719     0.5345         0.2717          4          5
12.5000*           .          .          .          .              .          4          4
13.0000*           .          .          .          .              .          4          3
13.3000*           .          .          .          .              .          4          2
15.0000*           .          .          .          .              .          4          1
17.6000*      0.5859     0.4141          .          .              .          4          0
```

We note that

$$\hat{S}(5.4) = \exp(-0.3679) = 0.6922.$$

Using R software

```
> library(survival)

> Time<-c(1.5,3.2,4.3,5.4,11.8,12.5,13.0,13.3,15.0,17.6)
> Status<-c(1,0,1,1,1,0,0,0,0,0)

> Surv(Time,Status)
[1]  1.5   3.2+  4.3   5.4  11.8  12.5+ 13.0+ 13.3+ 15.0+ 17.6+

> survfit(Surv(Time,Status)~1)
Call: survfit(formula = Surv(Time, Status) ~ 1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
10.0    10.0    10.0     4.0      NA     5.4      NA

> summary(survfit(Surv(Time,Status)~1,type="fleming-harrington"))
Call: survfit(formula = Surv(Time, Status) ~ 1, type = "fleming-harrington")

time n.risk n.event survival std.err lower 95% CI upper 95% CI
 1.5     10       1    0.905  0.0954        0.736            1
 4.3      8       1    0.799  0.1359        0.572            1
 5.4      7       1    0.692  0.1590        0.441            1
11.8      6       1    0.586  0.1719        0.330            1
```

Next to an estimate $\hat{S}(t)$ for the survival function $S(t)$ at t, we want to have an idea about the variability of this estimate.

This is given by Greenwood's formula

$$\widehat{\mathrm{Var}}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}$$

where $t_1, \ldots, t_k$ are the uncensored observed lifetimes.

Looking at the survival function, we rewrite as

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \hat{\lambda}_j\right)$$

where $\hat{\lambda}_j = \frac{d_j}{n_j}$.

The number of events d_j has a binomial distribution, which is approximated by a normal distribution. Hence,

$$\widehat{\mathrm{Var}}(\hat{\lambda}_j) = \frac{\hat{\lambda}_j \left(1 - \hat{\lambda}_j\right)}{n_j}.$$

In large samples, the $\hat{\lambda}_j$ are independent.

Instead of $\hat{S}(t)$, we look at

$$\log\left[\hat{S}(t)\right] = \sum_{t_j \le t} \log\left(1 - \hat{\lambda}_j\right).$$

Using the delta-method, we get that

$$\begin{aligned}
\widehat{\text{Var}}\left[\log\left(\hat{S}(t)\right)\right] &= \sum_{t_j \le t} \widehat{\text{Var}}\left[\log\left(1 - \hat{\lambda}_j\right)\right] \\
&= \sum_{t_j \le t} \left(\frac{1}{1 - \hat{\lambda}_j}\right)^2 \frac{\hat{\lambda}_j(1 - \hat{\lambda}_j)}{n_j} \\
&= \sum_{t_j \le t} \frac{d_j}{n_j(n_j - d_j)}.
\end{aligned}$$

Since $\hat{S}(t) = \exp\left[\log(\hat{S}(t))\right]$, we get the result from the delta-method,

$$\widehat{\text{Var}}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}.$$

<u>**Delta-method:**</u>

If Y is normally distributed with mean μ and variance σ^2, then $g(Y)$ is approximately normal with mean $g(\mu)$ and variance $g'(\mu)^2\sigma^2$.

With no censoring, $\hat{S}(t)$ is the empirical survival function and

$$\hat{S}(t) \approx N\left(S(t), \frac{S(t)(1-S(t))}{n}\right).$$

A pointwise $(1-\alpha)\%$ confidence interval for $S(t)$ is given by

$$\hat{S}(t) \pm z_{1-\frac{\alpha}{2}} \text{s.e.}(\hat{S}(t)).$$

With censoring,

- $\hat{S}(t)$ still approximately normal.
- the mean of $\hat{S}(t)$ is still the true $S(t)$.
- variance $\rightarrow$ Greenwood's formula.

However, the bounds of the C.I can be < 0 or > 1 !

Hence, better C.I's by using transformations of $S(t)$.

- Log-function: $\log[\hat{S}(t)] \pm z_{1-\frac{\alpha}{2}} \mathrm{s.e.} \log[\hat{S}(t)]$

$$\Rightarrow \left[\hat{S}(t) \exp\left(z_{1-\frac{\alpha}{2}} \hat{\tau}(t)\right), \hat{S}(t) \exp\left(-z_{1-\frac{\alpha}{2}} \hat{\tau}(t)\right)\right]$$

with $\hat{\tau}^2(t) = \frac{\widehat{\mathrm{Var}}(\hat{S}(t))}{\hat{S}(t)^2}$ (default: R).

- Log-log-function: $\log(-\log[\hat{S}(t)]) \pm z_{1-\frac{\alpha}{2}} \mathrm{s.e.} \log(-\log[\hat{S}(t)])$

$$\Rightarrow \left[\hat{S}(t)^{\exp\left(z_{1-\frac{\alpha}{2}} \hat{\tau}(t)\right)}, \hat{S}(t)^{\exp\left(-z_{1-\frac{\alpha}{2}} \hat{\tau}(t)\right)}\right]$$

with $\hat{\tau}^2(t) = \frac{\widehat{\mathrm{Var}}(\hat{S}(t))}{(\hat{S}(t) \log(\hat{S}(t)))^2}$ (default: SAS).

A second example: Sea sickness

Prediction of sea sickness, Burns, Aviat Space Environ Med (1984)

- 21 persons are subjected to 2-hr "rocking" with 0.167 Hz frequency and 0.111 G acceleration.
- Time until event: time when vomited for the first time.
- Two persons requested the stop of the experiment.

Time (minutes)	Vomit (1=yes)
30	1
50	1
50	0
51	1
66	0
82	1
92	1
120	0
...	...
120	0

```
proc lifetest data=vomit;
time time*vomit(0);
survival out=out1 conftype=log;
run;
proc print data=out1;
run;
```

```
The LIFETEST Procedure

                   Product-Limit Survival Estimates

                                               Survival
                                               Standard      Number       Number
    time      Survival     Failure              Error        Failed        Left

   0.000        1.0000           0                  0            0          21
  30.000        0.9524      0.0476             0.0465            1          20
  50.000        0.9048      0.0952             0.0641            2          19
  50.000*            .           .                  .            2          18
  51.000        0.8545      0.1455             0.0778            3          17
  66.000*            .           .                  .            3          16
  82.000        0.8011      0.1989             0.0894            4          15
  98.000        0.7477      0.2523             0.0981            5          14
 120.000*            .           .                  .            5          13
 120.000*            .           .                  .            5          12
 120.000*            .           .                  .            5          11
 120.000*            .           .                  .            5          10
```

120.000*	.	.	.	5	9
120.000*	.	.	.	5	8
120.000*	.	.	.	5	7
120.000*	.	.	.	5	6
120.000*	.	.	.	5	5
120.000*	.	.	.	5	4
120.000*	.	.	.	5	3
120.000*	.	.	.	5	2
120.000*	.	.	.	5	1
120.000*	.	.	.	5	0

NOTE: The marked survival times are censored observations.

Summary Statistics for Time Variable time

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	.	.	.
50	.	.	.
25	98.000	51.000	.

Mean	Standard Error
89.259	4.789

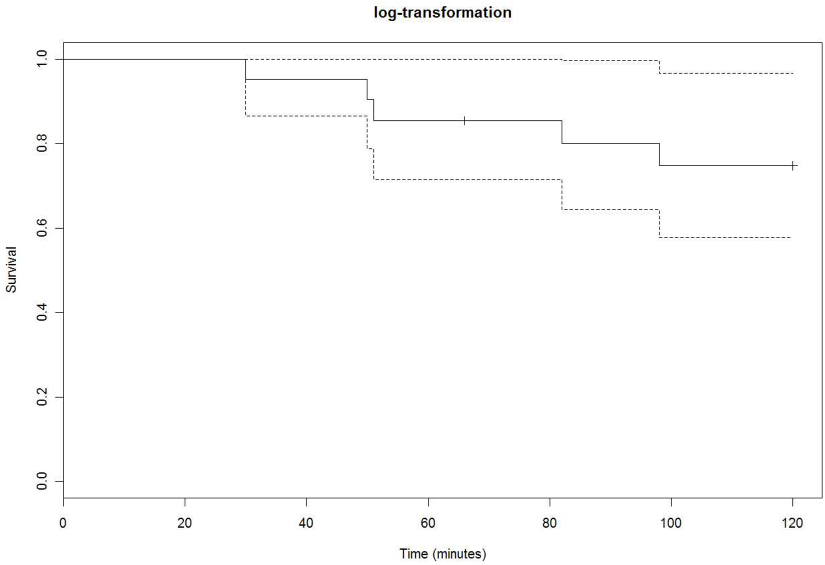
NOTE: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.

Summary of the Number of Censored and Uncensored Values

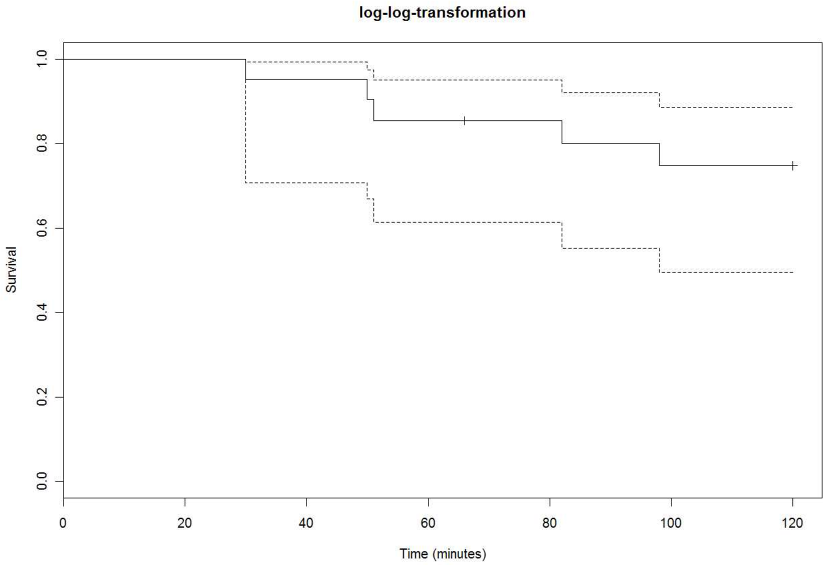
Total	Failed	Censored	Percent Censored
21	5	16	76.19

Obs	time	_CENSOR_	SURVIVAL	CONFTYPE	SDF_LCL	SDF_UCL
1	0	.	1.00000		1.00000	1.00000
2	30	0	0.95238	LOG	0.86552	1.00000
3	50	0	0.90476	LOG	0.78754	1.00000
4	50	1	0.90476		.	.
5	51	0	0.85450	LOG	0.71491	1.00000
6	66	1	0.85450		.	.
7	82	0	0.80109	LOG	0.64375	0.99689
8	98	0	0.74769	LOG	0.57817	0.96690
9	120	1	.		.	.
10	120	1	.		.	.
11	120	1	.		.	.
12	120	1	.		.	.
13	120	1	.		.	.
14	120	1	.		.	.
15	120	1	.		.	.
16	120	1	.		.	.
17	120	1	.		.	.
18	120	1	.		.	.
19	120	1	.		.	.
20	120	1	.		.	.
21	120	1	.		.	.
22	120	1	.		.	.

```
> survfit(Surv(time,vomit)~1)
Call: survfit(formula = Surv(time, vomit) ~ 1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
     21      21      21       5      NA      NA      NA

> summary(survfit(Surv(time,vomit)~1))
Call: survfit(formula = Surv(time, vomit) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   30     21       1    0.952  0.0465        0.866        1.000
   50     20       1    0.905  0.0641        0.788        1.000
   51     18       1    0.854  0.0778        0.715        1.000
   82     16       1    0.801  0.0894        0.644        0.997
   98     15       1    0.748  0.0981        0.578        0.967

> summary(survfit(Surv(time,vomit)~1,conf.type="log-log"))
Call: survfit(formula = Surv(time, vomit) ~ 1, conf.type = "log-log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   30     21       1    0.952  0.0465        0.707        0.993
   50     20       1    0.905  0.0641        0.670        0.975
   51     18       1    0.854  0.0778        0.613        0.951
   82     16       1    0.801  0.0894        0.552        0.921
   98     15       1    0.748  0.0981        0.495        0.887

> plot(survfit(Surv(time,vomit)~1),xlab="Time (minutes)",ylab="Survival",main="log-transformation")
> plot(survfit(Surv(time,vomit)~1,conf.type="log-log"),xlab="Time (minutes)",ylab="Survival",
main="log-log-transformation")
```



log-transformation
Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
Time (minutes)


log-log-transformation
Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
Time (minutes)

"Redistribute to the right"- algorithm

Efron developed an alternative method to compute the KM-estimator.

He set up an algorithm which starts by assuming no censoring and, as in the empirical distribution, each observation has the same mass $\frac{1}{n}$.

Afterwards in different steps, we distribute the mass of the censored observations over the observations which are still at risk.

Example: $3, 4, 5+, 6, 6+, 8+, 11, 14, 15, 16+$.

Data	Step 0	Step 1	Step 2	Step 3	$\hat{S}(t)$
3	$\frac{1}{10}$	0.100	0.100	0.100	0.900
4	$\frac{1}{10}$	0.100	0.100	0.100	0.800
5+	$\frac{1}{10}$	0.000	0.000	0.000	0.800
6	$\frac{1}{10}$	$\frac{1}{10} + \frac{1}{7}\frac{1}{10}$ $= 0.114$	0.114	0.114	0.686
6+	$\frac{1}{10}$	0.114	0.000	0.000	0.686
8+	$\frac{1}{10}$	0.114	$0.114 + \frac{1}{5}0.114$ $= 0.137$	0.000	0.686
11	$\frac{1}{10}$	0.114	0.137	$0.137 + \frac{1}{4}0.137$ $= 0.171$	0.515
14	$\frac{1}{10}$	0.114	0.137	0.171	0.343
15	$\frac{1}{10}$	0.114	0.137	0.171	0.171
16+	$\frac{1}{10}$	0.114	0.137	0.171	0.000∗