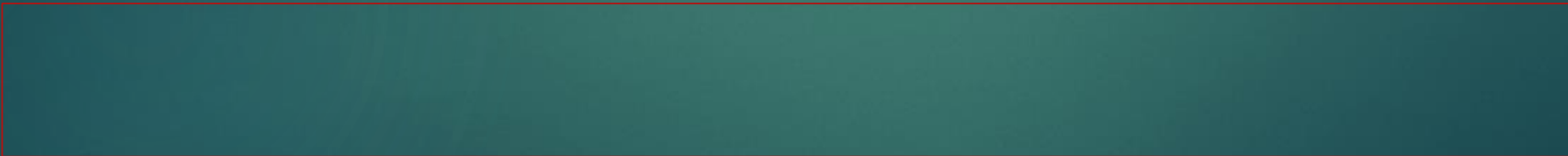


برآورد تابع خطر تجمعی



The Nelson Aalen Estimation

- Using a similar approach to the Kaplan-Meier, we can also estimate the cumulative hazard function $H(t)$
- Remember: $H(t)$ denotes the expected number of events up to and including time point t .

- So a natural estimator of $H(t)$ is

$$\hat{H}_{NA}(t) = \sum_{i: t_i \leq t} \frac{d_i}{r_i}$$

which is called the Nelson-Aalen estimator

The Breslow Estimator

- Remember: the cumulative hazard function is related to the survival function
- Therefore, an estimator for the survival function based on the Nelson-Aalen estimator is

$$\hat{S}_B(t) = \exp\{-\hat{H}_{NA}(t)\} = \prod_{i: t_i \leq t} \exp(-d_i/r_i)$$

which has been suggested by Breslow and therefore it is known as the *Breslow Estimator*

Note: an unnatural and undesirable feature of the Breslow estimator is that, its estimated survival probability will never reach zero even if the last observation is an event.

The Nelson Aalen & Breslow Estimator

- Using the same example

1 5⁺ 6 6 8 8⁺ 9 11⁺

+ denotes a censored time

i	t_i	r_i	d_i	$\frac{r_i - d_i}{r_i}$	d_i / r_i	$\hat{S}_B(t)$
1	1	8	1	7/8	1/8	$\exp(-1/8) = 0.882$
2	6	6	2	4/6	2/6	$\exp\{-(1/8 + 2/6)\} = 0.632$
3	8	4	1	3/4	1/4	$\exp\{-(1/8 + 2/6 + 1/4)\} = 0.492$
4	9	2	1	1/2	1/2	$\exp\{-(1/8 + 2/6 + 1/4 + 1/2)\} = 0.297$

$$\hat{S}_{KM}(t) = 1, \quad 0 \leq t < 1$$

$$= 7/8 = 0.875, \quad 1 \leq t < 6$$

$$= (7/8)(4/6) = 0.583, \quad 6 \leq t < 8$$

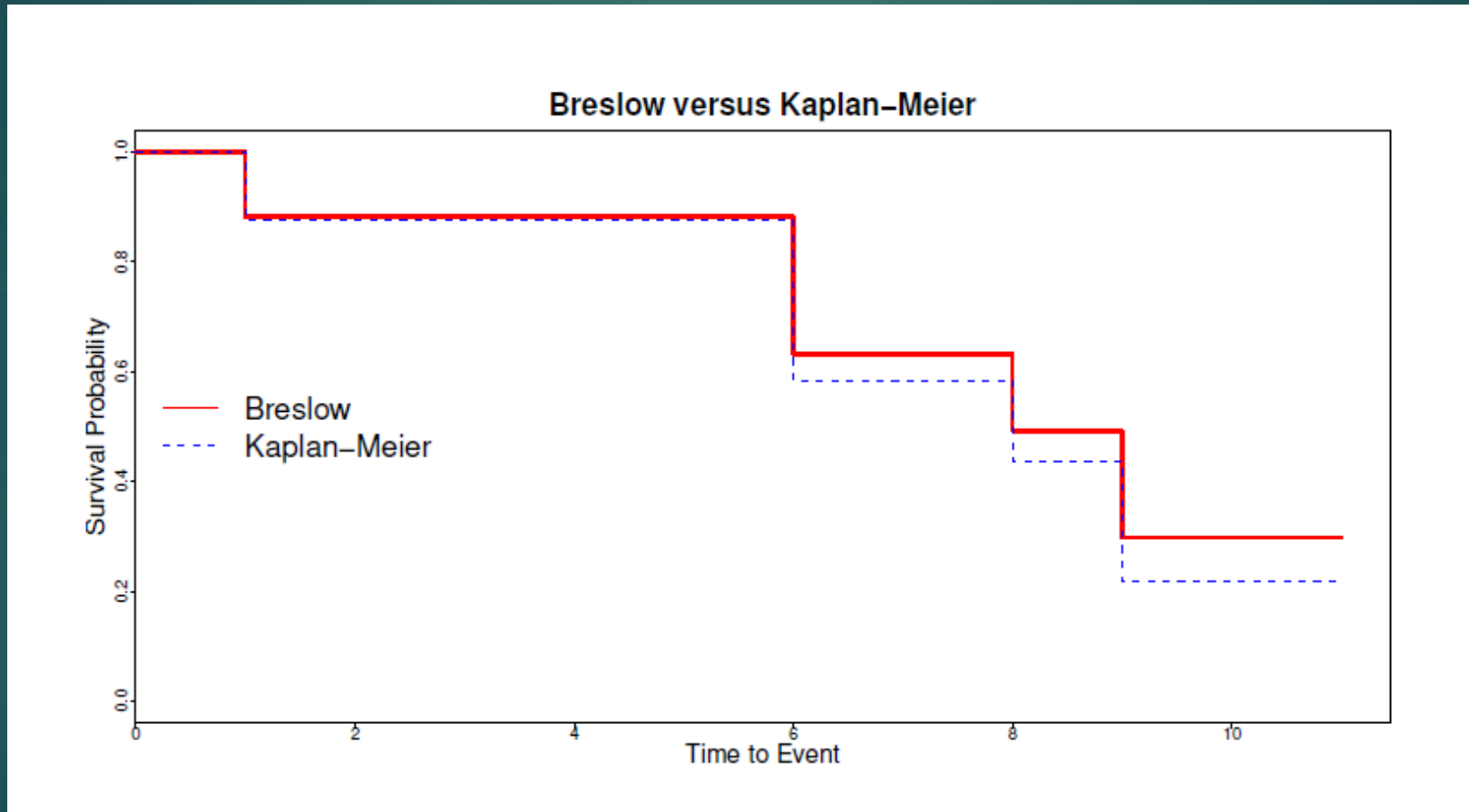
$$= (7/8)(4/6)(3/4) = 0.438, \quad 8 \leq t < 9$$

$$= (7/8)(4/6)(3/4)(1/2) = 0.219, \quad 9 \leq t < 11$$

Note: Spss didn't calculate C.Hazrd. Instead used K-M to calculate C.Hazrd function. $H(t) = -\ln(k-M)$

Also, the survival probability that calculate with Breslow method usually is more than K-M method

Survival Function, Breslow Estimator



The Nelson Aalen & Breslow Estimator

The Nelson–Aalen estimator of the cumulative hazard rate function is derived from [Nelson \(1972\)](#) and [Aalen \(1978\)](#) and is defined up to the largest observed time as

$$\hat{H}(t) = \sum_{j|t_j \leq t} \frac{d_j}{n_j}$$

تر رجم زاده، آمار زیستی

Its variance ([Aalen 1978](#)) may be estimated by

$$\widehat{\text{Var}}\{\hat{H}(t)\} = \sum_{j|t_j \leq t} \frac{d_j}{n_j^2}$$

- Using the same example

1 5⁺ 6 6 8 8⁺ 9 11⁺

+ denotes a censored time

i	t_i	r_i	d_i	$\frac{r_i - d_i}{r_i}$	d_i / r_i	$\hat{S}_B(t)$
1	1	8	1	7/8	1/8	$\exp(-1/8) = 0.882$
2	6	6	2	4/6	2/6	$\exp\{-(1/8 + 2/6)\} = 0.632$
3	8	4	1	3/4	1/4	$\exp\{-(1/8 + 2/6 + 1/4)\} = 0.492$
4	9	2	1	1/2	1/2	$\exp\{-(1/8 + 2/6 + 1/4 + 1/2)\} = 0.297$

ti	ri	di	C.h	error
1	8	1	0.125	0.125
6	6	2	0.4583	0.2668
8	4	1	0.7083	0.3656
9	2	1	1.2083	0.6194

- ▶ $\widehat{\sigma}_H^2(t_i) = \sum_{t_i \leq t} \frac{d_i}{r_i^2}$
- ▶ $\widehat{\sigma}_H^2(t_1 = 1) = \frac{1}{8^2} = 0.0156 \text{ } error = 0.125$
- ▶ $\widehat{\sigma}_H^2(t_2 = 6) = \frac{1}{8^2} + \frac{2}{6^2} = 0.0719 \text{ } error = 0.2668$
- ▶ $\widehat{\sigma}_H^2(t_2 = 8) = \frac{1}{8^2} + \frac{2}{6^2} + \frac{1}{4^2} = 0.1337 \text{ } error = 0.3656$
- ▶ $\widehat{\sigma}_H^2(t_2 = 9) = \frac{1}{8^2} + \frac{2}{6^2} + \frac{1}{4^2} + \frac{1}{2^2} = 0.3837 \text{ } error = 0.6194$
- ▶ **Note:** But didn't used its in the Confidence Interval

Pointwise confidence intervals are calculated using the asymptotic variance of $\ln \hat{H}(t)$,

$$\hat{\phi}^2(t) = \frac{\widehat{\text{Var}}\{\hat{H}(t)\}}{\{\hat{H}(t)\}^2}$$

The confidence bounds are then $\hat{H}(t) \exp\{\pm z_{\alpha/2} \hat{\phi}(t)\}$.

$$\widehat{\phi}^2(t) = \frac{\widehat{\text{Var}}\{\hat{H}(t)\}}{\{\hat{H}(t)\}^2}$$

$$\widehat{\sigma}_H^2(t_1 = 1) = 0.0156 \quad \widehat{\phi}^2(1) = \frac{0.0156}{0.125^2} = 1, (0.125 \times \exp\{\pm 1.96 \times 1\} (0.0176, 0.8874))$$

$$\widehat{\sigma}_H^2(t_2 = 6) = 0.0719 \quad \widehat{\phi}^2(6) = \frac{0.0719}{0.4583^2} = 0.3423 \quad (0.4583 \times \exp\{\pm 1.96 \times 0.5851\} (0.1456, 1.4427))$$

$$\widehat{\sigma}_H^2(t_2 = 8) = 0.1337 \quad \widehat{\phi}^2(8) = \frac{0.1337}{0.7083^2} = 0.2665 \quad (0.7083 \times \exp\{\pm 1.96 \times 0.5162\} (0.2575, 1.9481))$$

$$\widehat{\sigma}_H^2(t_2 = 9) = 0.3837 \quad \widehat{\phi}^2(9) = \frac{0.3837}{1.2083^2} = 0.2628 \quad (1.2083 \times \exp\{\pm 1.96 \times 0.5126\} (0.4424, 3.29996))$$

STATA: sts list, cumhaz

Nelson-Aalen cumulative hazard function

Time	At risk	Fail	Lost	Cumulative hazard	Std. error	[95% conf. int.]	
1	8	1	0	0.1250	0.1250	0.0176	0.8874
5	7	0	1	0.1250	0.1250	0.0176	0.8874
6	6	2	0	0.4583	0.2668	0.1465	1.4344
8	4	1	1	0.7083	0.3656	0.2576	1.9481
9	2	1	0	1.2083	0.6194	0.4424	3.3001
11	1	0	1	1.2083	0.6194	0.4424	3.3001

Example of calculation of Cumulative Hazard

- ▶ McIlmurray and Turkie (2) describe a clinical trial of 25 patients for the treatment and 24 patients for the Control of Dukes' C colorectal cancer.

T
C

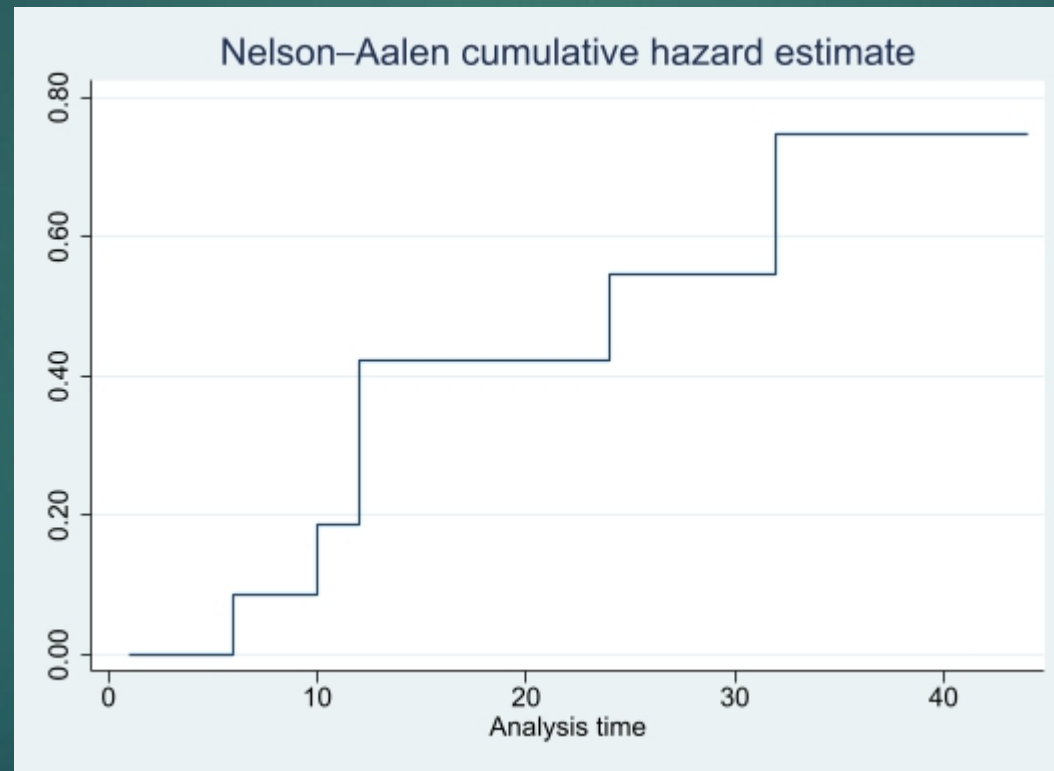
Table 12.1 Survival in 49 patients with Dukes' C colorectal cancer randomly assigned to either γ linoleic acid or control treatment	
Treatment	Survival time (months)
γ linoleic acid (n=25)	1+, 5+, 6, 6, 9+, 10, 10, 10+, 12, 12, 12, 12, 12+, 13+, 15+, 16+, 20+, 24, 24+, 27+, 32, 34+, 36+, 36+, 44+
Control (n=24)	3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 15+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+, 42

STATA: sts list, cumhaz Out Put

Nelson-Aalen cumulative hazard function

Time	At risk	Fail	Lost	Cumulative hazard	Std. error	[95% conf. int.]	
1	25	0	1	0.0000	0.0000	.	.
5	24	0	1	0.0000	0.0000	.	.
6	23	2	0	0.0870	0.0615	0.0217	0.3477
9	21	0	1	0.0870	0.0615	0.0217	0.3477
10	20	2	1	0.1870	0.0937	0.0700	0.4993
12	17	4	1	0.4223	0.1504	0.2101	0.8487
13	12	0	1	0.4223	0.1504	0.2101	0.8487
15	11	0	1	0.4223	0.1504	0.2101	0.8487
16	10	0	1	0.4223	0.1504	0.2101	0.8487
20	9	0	1	0.4223	0.1504	0.2101	0.8487
24	8	1	1	0.5473	0.1956	0.2716	1.1025
27	6	0	1	0.5473	0.1956	0.2716	1.1025
32	5	1	0	0.7473	0.2797	0.3588	1.5563
34	4	0	1	0.7473	0.2797	0.3588	1.5563
36	3	0	2	0.7473	0.2797	0.3588	1.5563
44	1	0	1	0.7473	0.2797	0.3588	1.5563

STATA: sts graph, cumhaz



The Breslow Estimator

- The difference between the Kaplan-Meier and the Breslow estimators is always very small
 - ▷ as $n \rightarrow \infty$ the two estimates are equivalent
- The Breslow estimator is biased upwards, especially close to zero, but it has lower variance
 - ▷ if the largest observed time T in the data set is an event, then $\hat{S}_{KM}(t) = 0$ whereas $\hat{S}_B(t)$ is positive

The Breslow Estimator

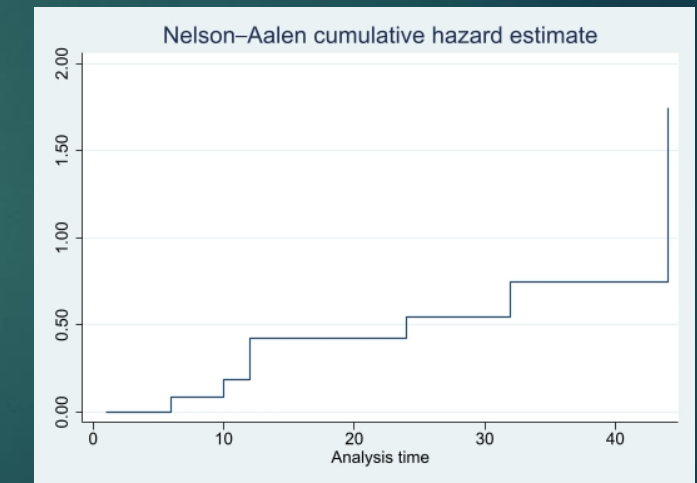
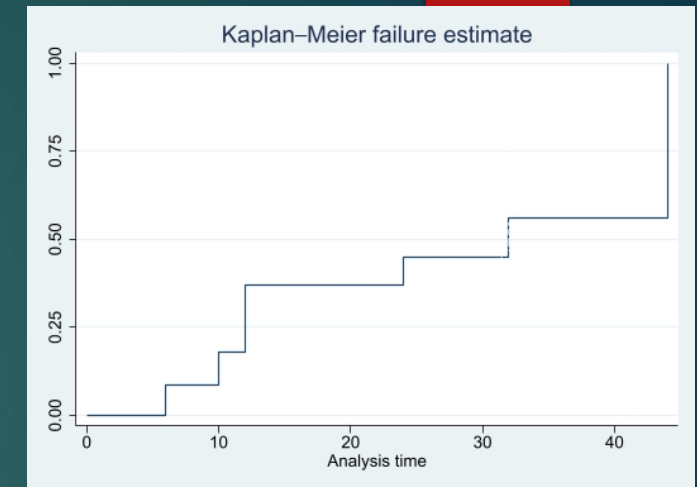
- As before, we observe that the two estimators are indistinguishable
- The Kaplan-Meier is more popular
- However, a lot of theoretical developments in statistics have been based on the Breslow estimator

KM or NA

- ▶ In small samples, the Kaplan-Meier product-limit estimator is superior when estimating the survivor function, and the Nelson-Aalen estimator is superior when estimating the cumulative hazard function.
- ▶ For the survivor function and the cumulative hazard function, both the Kaplan-Meier estimator and the Nelson Aalen estimator are consistent estimates of each, and the statistics are asymptotically equivalent (Klein and Moeschberger 2003, 104).
- ▶ That is, in very large samples, it does not matter how you estimate the survivor function, whether by Kaplan-Meier or by transforming the Nelson-Aalen.

KM or NA Cumulative Hazard

- ▶ The top curve is the cumulative hazard obtained by transforming the Kaplan-Meier survivor function, and the bottom one is the Nelson-Aalen curve. By analogy to the above discussion, the Kaplan-Meier version of the cumulative hazard is always greater than or equal to that from the Nelson-Aalen estimator. See, Lemeshow, and May (2008) for details.
- ▶ The Nelson-Aalen estimator, however, does offer another advantage. When some samples enter the study with a delay, such that by the time they enter the study, all individuals in the study have experienced the event in question. So, the K_M estimator is zero, it remains zero regardless of future activity. The Nelson-Aalen estimator does not suffer from this problem because it will increase with every failure event, even after the Kaplan-Meier estimate would have reached zero. When this is at issue, the Nelson-Aalen estimator is preferred.



What to avoid

- Compare mean survival $\Rightarrow$ Censoring makes this meaningless
- Overinterpret the tail of a survival curve $\Rightarrow$ There are generally few subjects in tails
- Compare proportions surviving at a fixed time $\Rightarrow$ Fine for description, not for hypothesis testing