November 18, 2020

Introduction to Survival Analysis

Thomas Hielscher (t.hielscher@dkfz.de)
Div. of Biostatistics

Kaplan Meier method and log-rank test

Survival endpoint

- Response of interest: time to event (failure time, survival time, event time)
  Examples:
  - Time from to start of therapy to death
  - Time from surgery to tumor recurrence

  State the event, start and end of the period
- Endpoint may not be observed
  Examples:
  - End of study before all patients had relevant outcome (incomplete follow-up)
  - Relocation (lost to follow-up)
  - Death from other cause (different outcome)

  $\rightarrow$ Incomplete responses are (right-)censored
  $\rightarrow$ Different methods for analysis and visualisation required

Censoring

Types of (right-)censoring

- Fixed type I censoring: study ends after a pre-specified follow-up time for each subject (e.g. animal experiments)
- Random type I censoring: study ends at a pre-specified time point (e.g. cut-off date in clinical study)
- Type II censoring: study ends after pre-specified number of events are observed

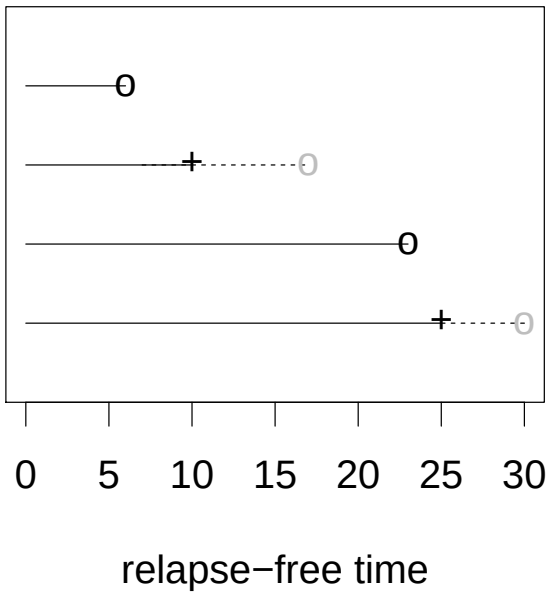
Fundamental Principle: Mechanism of censoring is independent (non-informative) of the mechanism of failure

Examples for dependent censoring:

- death from other cause (competing risk)
- drop-out due to drug-toxicity

Example data

Data from Freireich et al.,"The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia" Blood 21, 699-716, 1963.



patient	time	relapse
A	6	yes
B	10	no
C	23	yes
D	25	no

Survival time is not calender time

Survival data

- T_i denotes survival time of subject i, $i = 1, ..., N$
- C_i denotes censoring time of subject i
- The observed response is $Y_i = min(T_i, C_i)$
- δ_i denotes the event indicator: 1 if $T_i \leq C_i$, 0 if $T_i > C_i$

patient	i	T_i	C_i	Y_i	δ_i
A	1	6		6	1
B	2		10	10	0
⋮					

Risk set

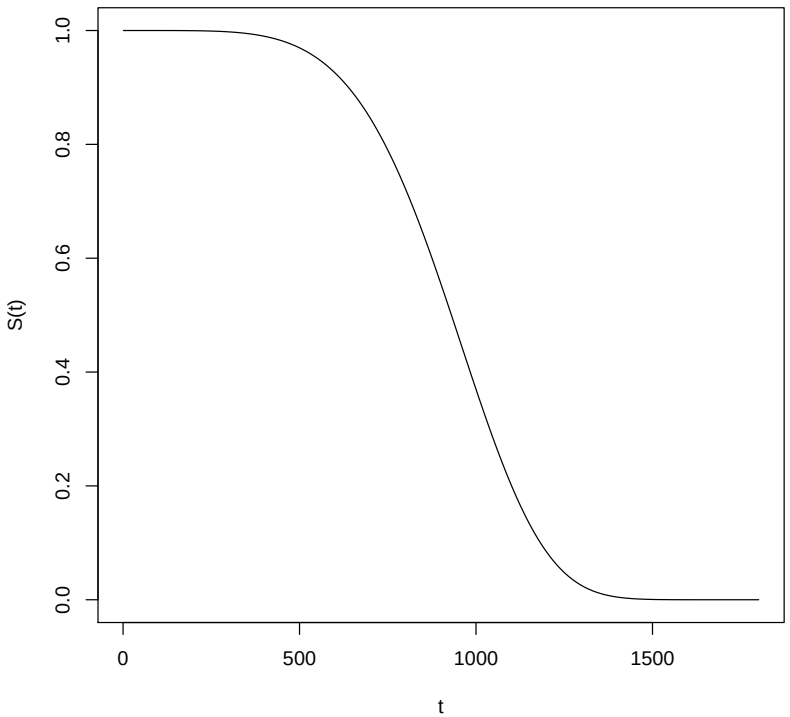
Example data, treatment group only (n=21):

n_t denotes number of patients at risk at time t, i.e. $Y_i \geq t$

time	patients at risk n_t	relapse d	censoring c
6	21	3	1
7	17	1	0
8	16	0	0
9	16	0	1
⋮			

Theoretical survival function

- T denotes the survival time, $T \geq 0$
- $F(t)$: distribution function of the survival times T
- Survival function $S(t) = Pr(T > t) = 1 - F(t)$



- $S(t)$ gives the probability that a subject survives past time t
- $S(t)$ is non-increasing
- $S(0) = 1$
- $S(\infty) = 0$ (exception e.g. cure rate models)
- Only in theory smooth $S(t)$, in practice time scale discrete (days, weeks,..)

Estimation of the survival function

- parametric methods: assumptions on the distribution of T (e.g. Weibull, exponential, log-normal)
  - AFT regression models
- semi-parametric methods: assumption on relative effects over time
  - Cox PH model (see lecture next week)
- non-parametric methods: no assumption on distribution of T
  - Kaplan-Meier method
  - Log-rank test

Kaplan-Meier method (1)

- Probability to survive past t is estimated as proportion of patients being alive at t among patients at risk

- Concept of conditional probabilities: in order to survive timepoint t, you have to survive timepoint $t - 1$

Kaplan-Meier method (2)

Idea of conditional probabilities:

p_1 = probability of surviving the first day ($S(1) = p_1$)

p_2 = probability of surviving the second day, given one has survived day 1 (conditional probability)

$\rightarrow$ (unconditional) probability of surviving day 2 is $S(2) = p_1 \times p_2$

Generalizes to:

$$\begin{aligned} S(t) &= S(t-1) \times p_t \\ &= p_1 \times p_2 \times \ldots \times p_{t-1} \times p_t \end{aligned}$$

Kaplan-Meier method (3)

Estimation of survival propability p_t:

- no failure, no censoring at timepoint t:
  $\hat{p}_t = \frac{n_t - 0}{n_t} = 1$, $n_{t+1} = n_t$
  estimated survival probability $\hat{S}(t) = \hat{S}(t-1)$
- d failures at timepoint t, $d \geq 1$:
  $\hat{p}_t = \frac{n_t - d}{n_t} < 1$, $n_{t+1} = n_t - d$
  $\hat{S}(t)$ drops since $\hat{S}(t) = \hat{S}(t-1) \times \hat{p}_t$, patients at risk decrease
- c censorings at timepoint t, $c \geq 1$:
  $\hat{p}_t = \frac{n_t - 0}{n_t} = 1$, $n_{t+1} = n_t - c$
  $\hat{S}(t)$ remains constant, patients at risk decrease

Kaplan-Meier method (4)

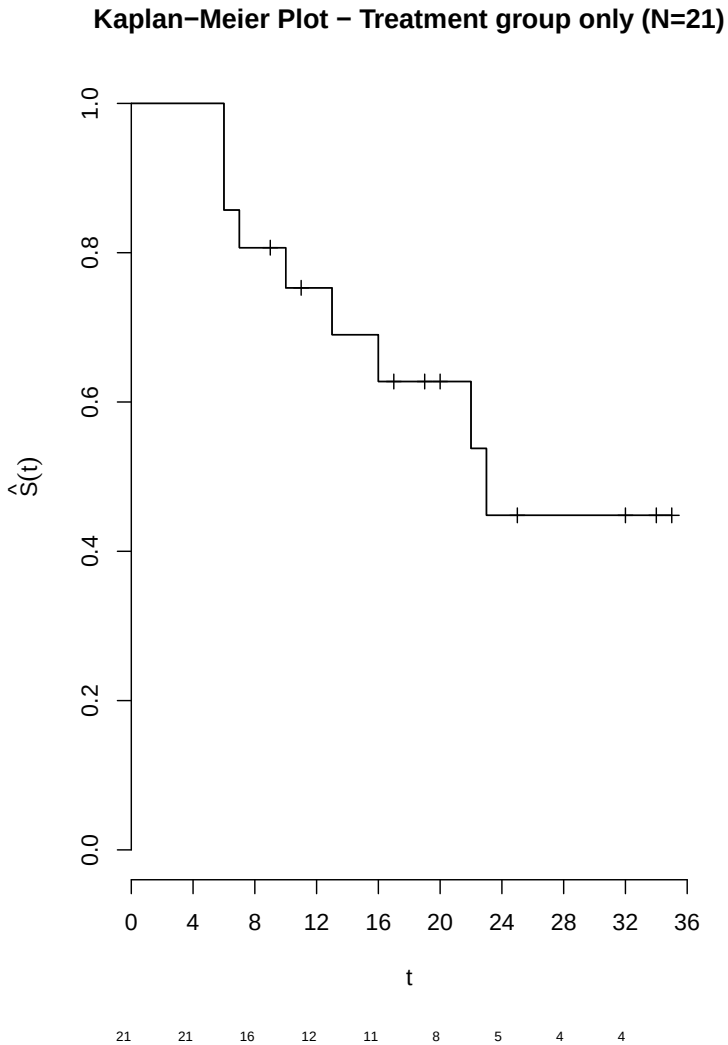
Kaplan Meier estimator:

$$\hat{S}(t) = \prod_t \hat{p}_t = \hat{p}_1 \times \hat{p}_2 \times \ldots \times \hat{p}_{t-1} \times \hat{p}_t = \prod_t \frac{n_t - d}{n_t}$$

- Also called: product limit estimate
- Unbiased under non-informative censoring
- In absence of censoring, KM estimator reduces to binomial estimate
- Variance estimation (Greenwood's formula):

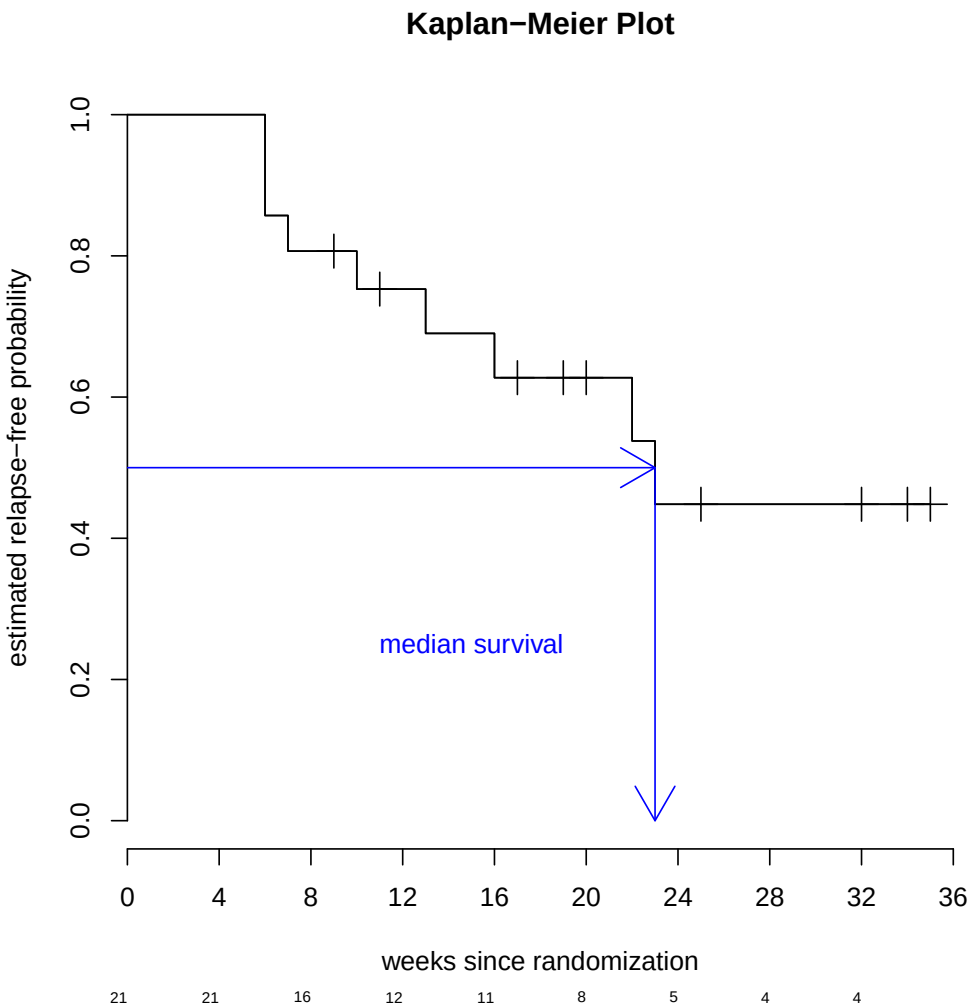
$$\hat{V}[\hat{S}(t)] = \hat{S}(t)^2 \sum_t \frac{d}{n_t(n_t - d)}$$

Example data



t	risk	d	c	$\hat{p}_t$			$\hat{S}(t)$
6	21	3	1	$\frac{21-3}{21}$	=	0.86	0.86
7	17	1	0	$\frac{17-1}{17}$	=	0.94	0.81
9	16	0	1	$\frac{16-0}{16}$	=	1.00	0.81
10	15	1	1	$\frac{15-1}{15}$	=	0.93	0.75
11	13	0	1	$\frac{13-0}{13}$	=	1.00	0.75
13	12	1	0	$\frac{12-1}{12}$	=	0.92	0.69
16	11	1	0	$\frac{11-1}{11}$	=	0.91	0.63
17	10	0	1	$\frac{10-0}{10}$	=	1.00	0.63
19	9	0	1	$\frac{9-0}{9}$	=	1.00	0.63
20	8	0	1	$\frac{8-0}{8}$	=	1.00	0.63
22	7	1	0	$\frac{7-1}{7}$	=	0.86	0.54
23	6	1	0	$\frac{6-1}{6}$	=	0.83	0.45
25	5	0	1	$\frac{5-0}{5}$	=	1.00	0.45
32	4	0	2	$\frac{4-0}{4}$	=	1.00	0.45
34	2	0	1	$\frac{2-0}{2}$	=	1.00	0.45
35	1	0	1	$\frac{1-0}{1}$	=	1.00	0.45

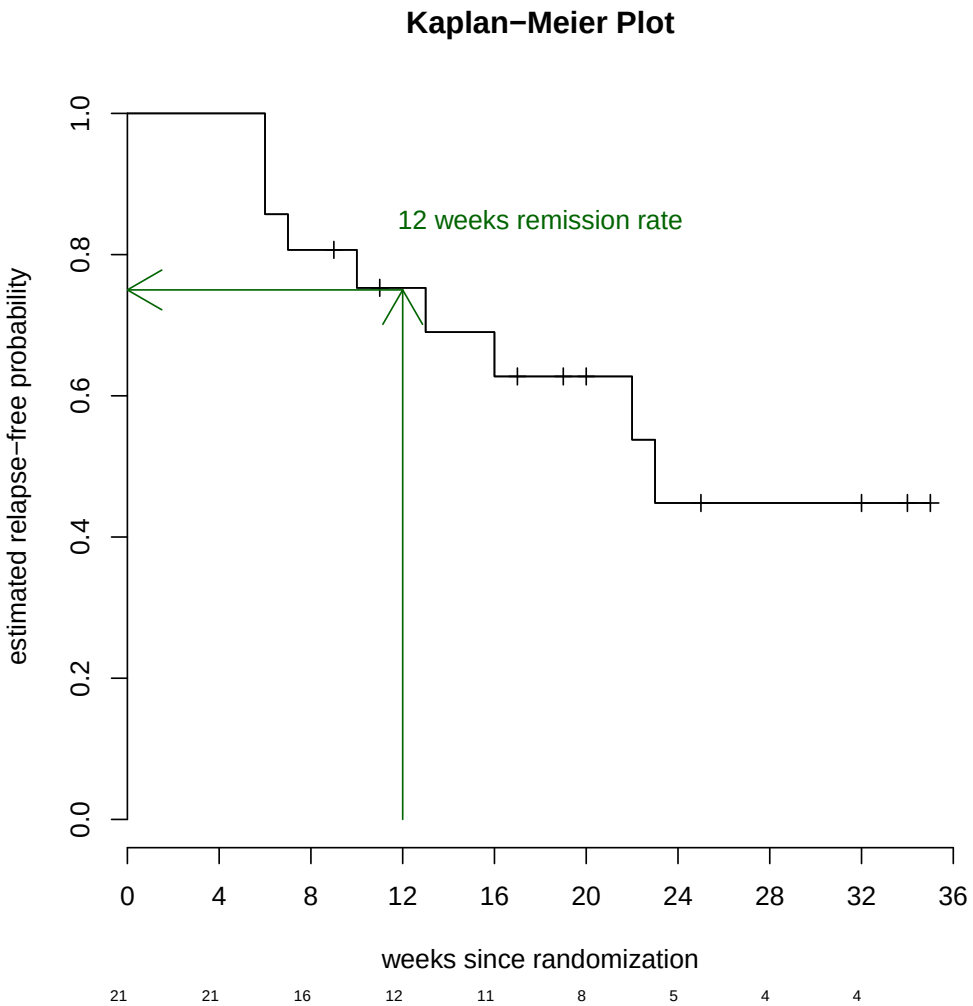
Median survival time



- Timepoint where $\hat{S}(t) = 0.5$
- Not always observed/inaccurate due to few patients at risk
- Confidence interval

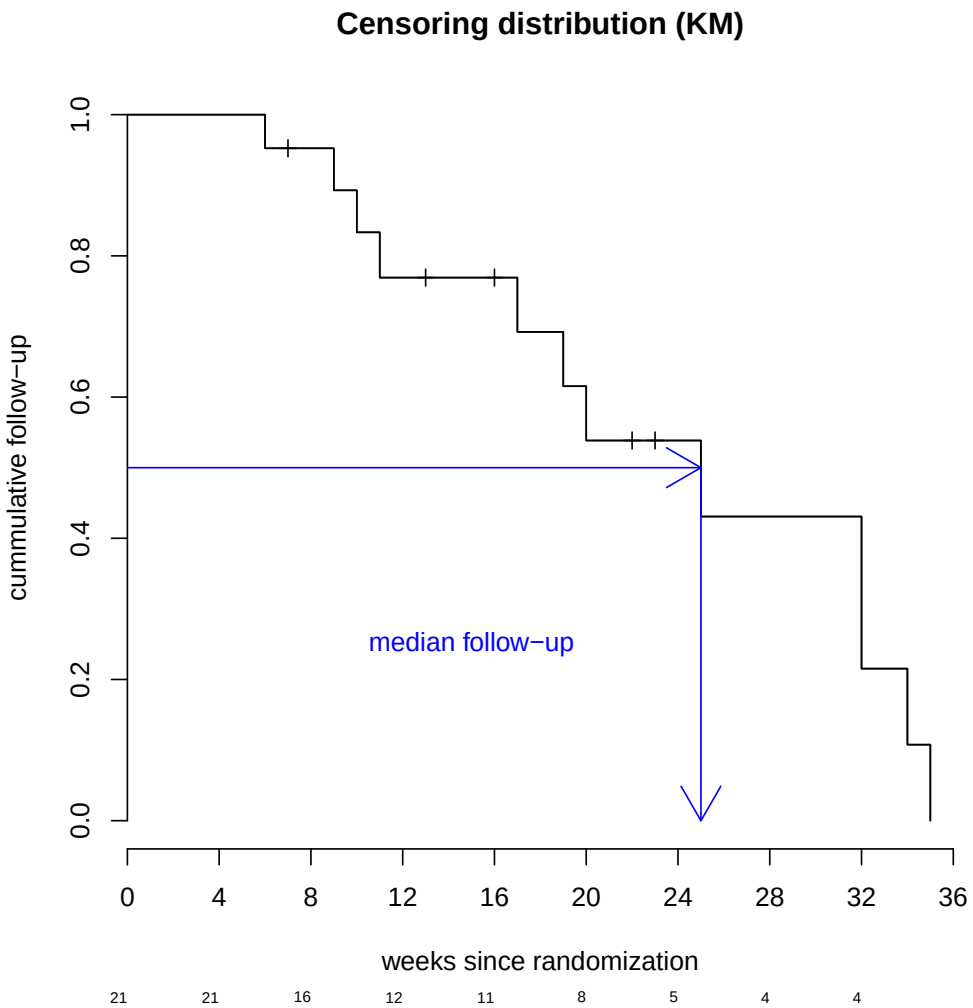
Survival rate

Survival probability at timepoint t:



- Use standard time points (e.g. 1 year, ...)
- Confidence interval

Median follow-up

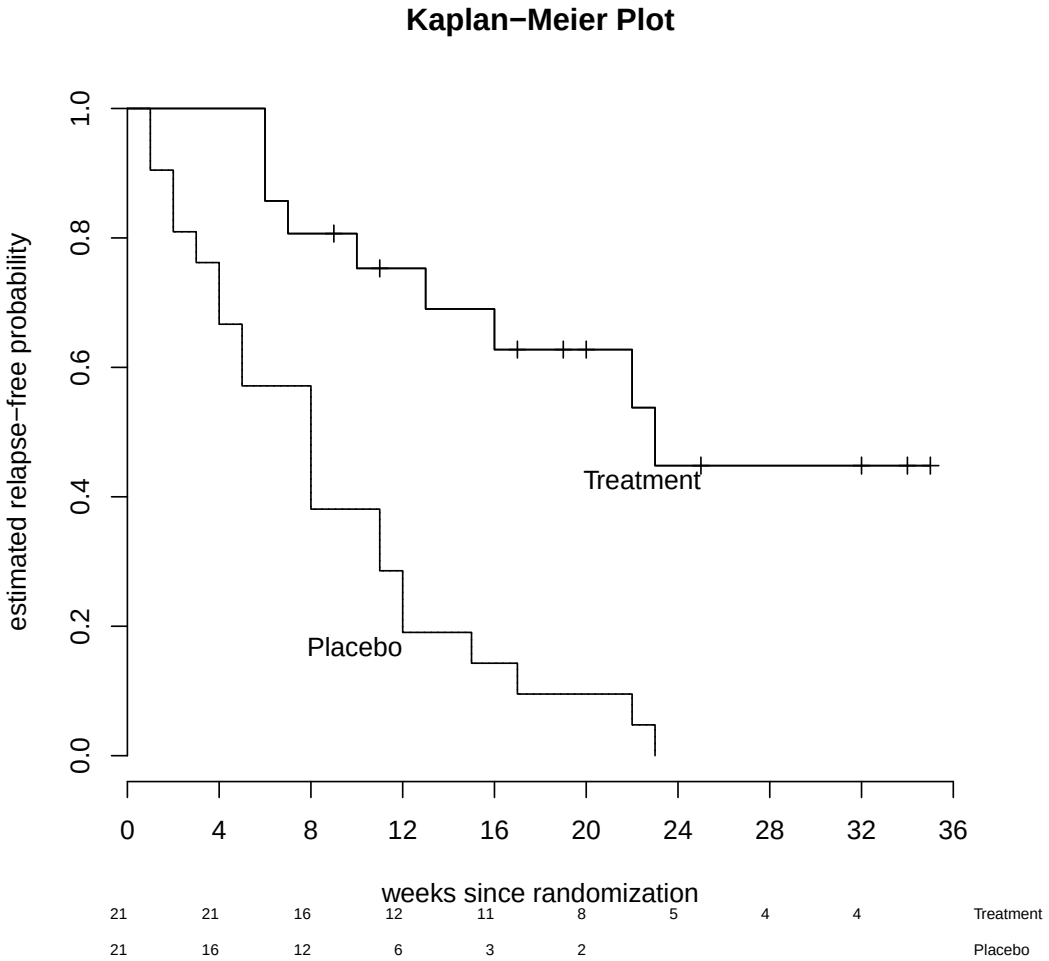


- Switch event indicator: 1-δ_i (Korn, 1986)
- Estimated censoring distribution
- Timepoint where $\hat{S}(t) = 0.5$

Mean survival time

- Generally: do not report simple mean survival time from $\hat{S}(t)$
  - area under the survival curve
  - not defined if last observation is censored, i.e., survival curve does not go to zero
  - groups not comparable if maximum observed time are different
- Restricted mean survival time (RMST):
  - average survival from time 0 to a specified time point t_{max}
  - t_{max} observed in all groups/curves
  - estimated as the area under the survival curve up to that point t_{max}
  - groups can only be compared for the same t_{max}

Example data



log-rank test (1)

- Mantel-Haenszel/Mantel Cox test
- Non-parametric test
- Null hypothesis H_0: there is no difference between the populations in the probability of an event at any time point
- Idea: describe survival process with a series of time-dependent 2×2 contingency tables, one for each event time $t_{(j)}$
- Compare observed vs. expected number of events at each time point

log-rank test (2)

group	events at $t_{(j)}$	no event at $t_{(j)}$	at risk at $t_{(j)}$
A	d_{Aj}	$n_{Aj} - d_{Aj}$	n_{Aj}
B	d_{Bj}	$n_{Bj} - d_{Bj}$	n_{Bj}
Totals	d_j	$n_j - d_j$	n_j

- One table for each event time $t_{(j)}$, $j = 1, .., m$
- Under H_0 of no difference:
  expected number of events at $t_{(j)}$ in group A is $\hat{e}_{Aj} = \frac{d_j n_{Aj}}{n_j}$
- Test statistic: $Q = \frac{\left(\sum d_{Aj} - \sum \hat{e}_{Aj}\right)^2}{\sum \hat{V}(\hat{e}_{Aj})}$
- Variance of $\hat{e}_{Aj}$ is $\hat{V}(\hat{e}_{Aj}) = \frac{n_{Aj} n_{Bj} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$
- Q is asy. χ^2_1 distributed under H_0

Example data

j	$t_{(j)}$	n_{Aj}	d_{Aj}	n_{Bj}	d_{Bj}	$\hat{e}_{Aj}$	$\hat{V}(\hat{e}_{Aj})$
1	1	21	0	21	2	1.00	0.49
2	2	21	0	19	2	1.05	0.49
3	3	21	0	17	1	0.55	0.25
4	4	21	0	16	2	1.14	0.48
5	5	21	0	14	2	1.20	0.47
6	6	21	3	12	0	1.91	0.65
7	7	17	1	12	0	0.59	0.24
8	8	16	0	12	4	2.29	0.87
9	10	15	1	8	0	0.65	0.23
10	11	13	0	8	2	1.24	0.45
11	12	12	0	6	2	1.33	0.42
12	13	12	1	4	0	0.75	0.19
13	15	11	0	4	1	0.73	0.20
14	16	11	1	3	0	0.79	0.17
15	17	10	0	3	1	0.77	0.18
16	22	7	1	2	1	1.56	0.30
17	23	6	1	1	1	1.71	0.20
Total			9		21	19.25	6.26

- $t(j) = 1$

group	events	no event	at risk
A	0	21	21
B	2	19	21
Totals	2	40	42

- $Q = (9 - 19.25)^2 / 6.26 = 16.79$
- p-value:
  $1 - P(Q \geq 16.79 | H_0) < 0.001$

Properties/assumptions of log-rank test

- Independence of censoring required
- Can be generalized to more than two groups:
  global null hypothesis vs. trend test
- Not appropriated for crossing survival curves (see numerater of Q)
- Weights within Q can be defined to derive different tests
- Gives equal weights to all time points
- Optimal power in case of proportional hazards $\rightarrow$ next week
- Provides no effect size $\rightarrow$ Regression model

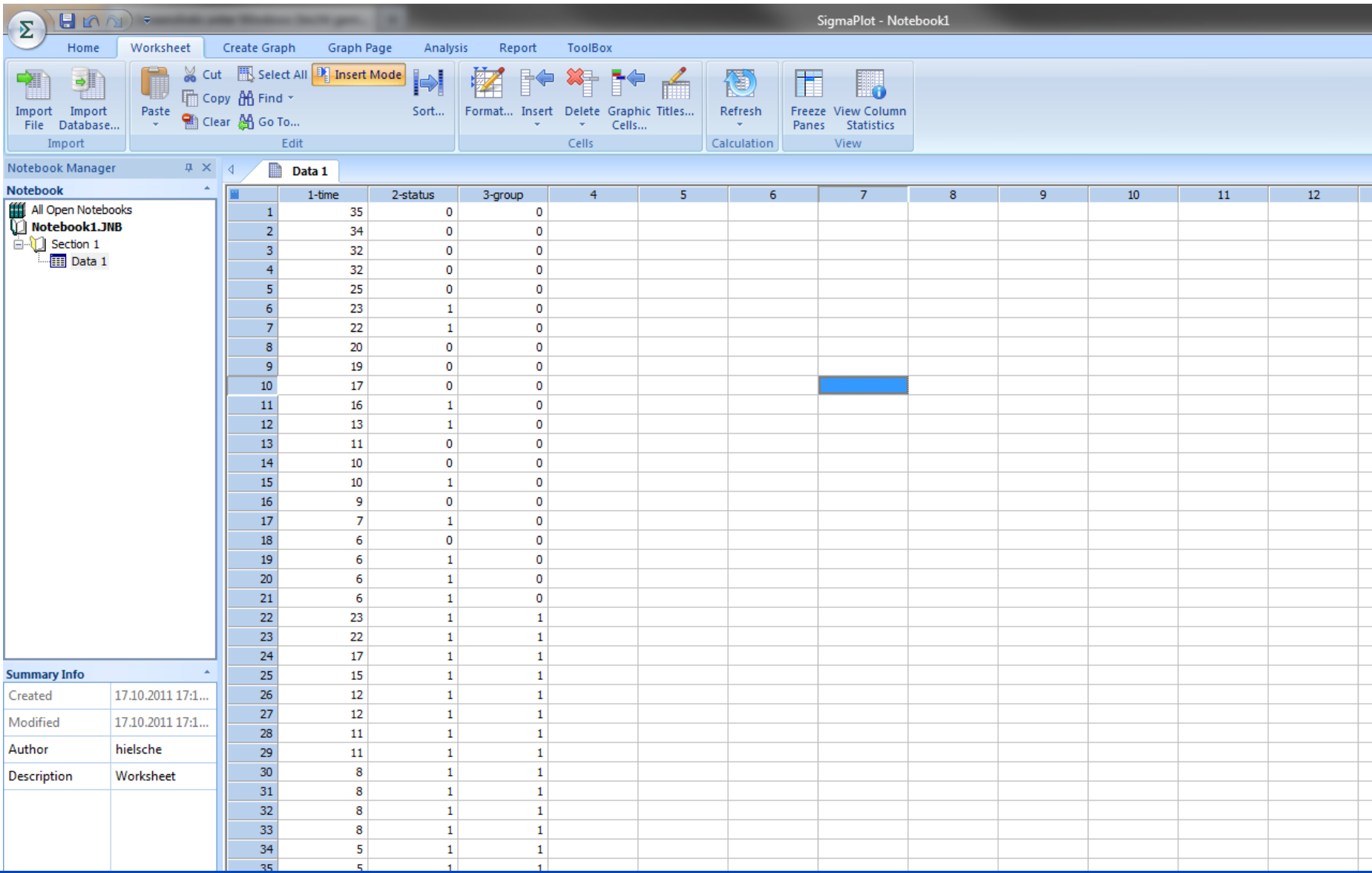
Gehan-Wilcoxon test

- Gehan-Breslow (SigmaPlot), generalized Wilcoxon
- Same class of tests as log-rank test
- Time points are weighted based on number at risk
- More weight on earlier events, detects differences early in time
- (Sometimes) more powerful if proportional hazard assumptions does not hold $\rightarrow$ next week
- Equivalent to Wilcoxon rank sum test in case of no censoring

Stratified log-rank test

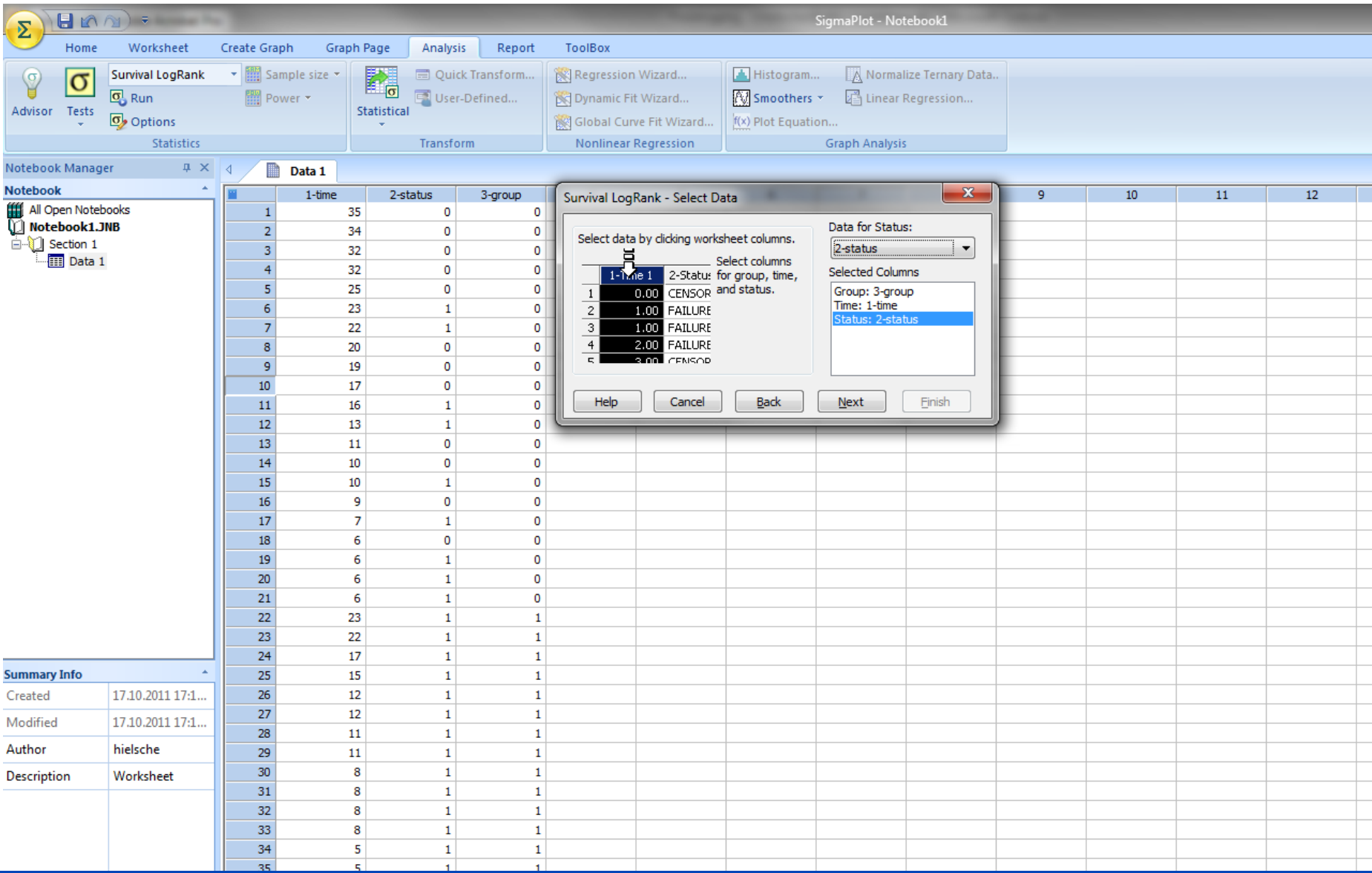
- Control for effect of additional categorical variable/confounder
- Split data into subgroups (strata)
- Calculate numerator/denominator of Q in each subgroup and average across subgroups
- Too small strata/subgroups affect power
- No estimation of effect of stratified variable (Regression model)
- Not implemented in SigmaPlot/GraphPad Prism

Log-rank test using SigmaPlot? (1)

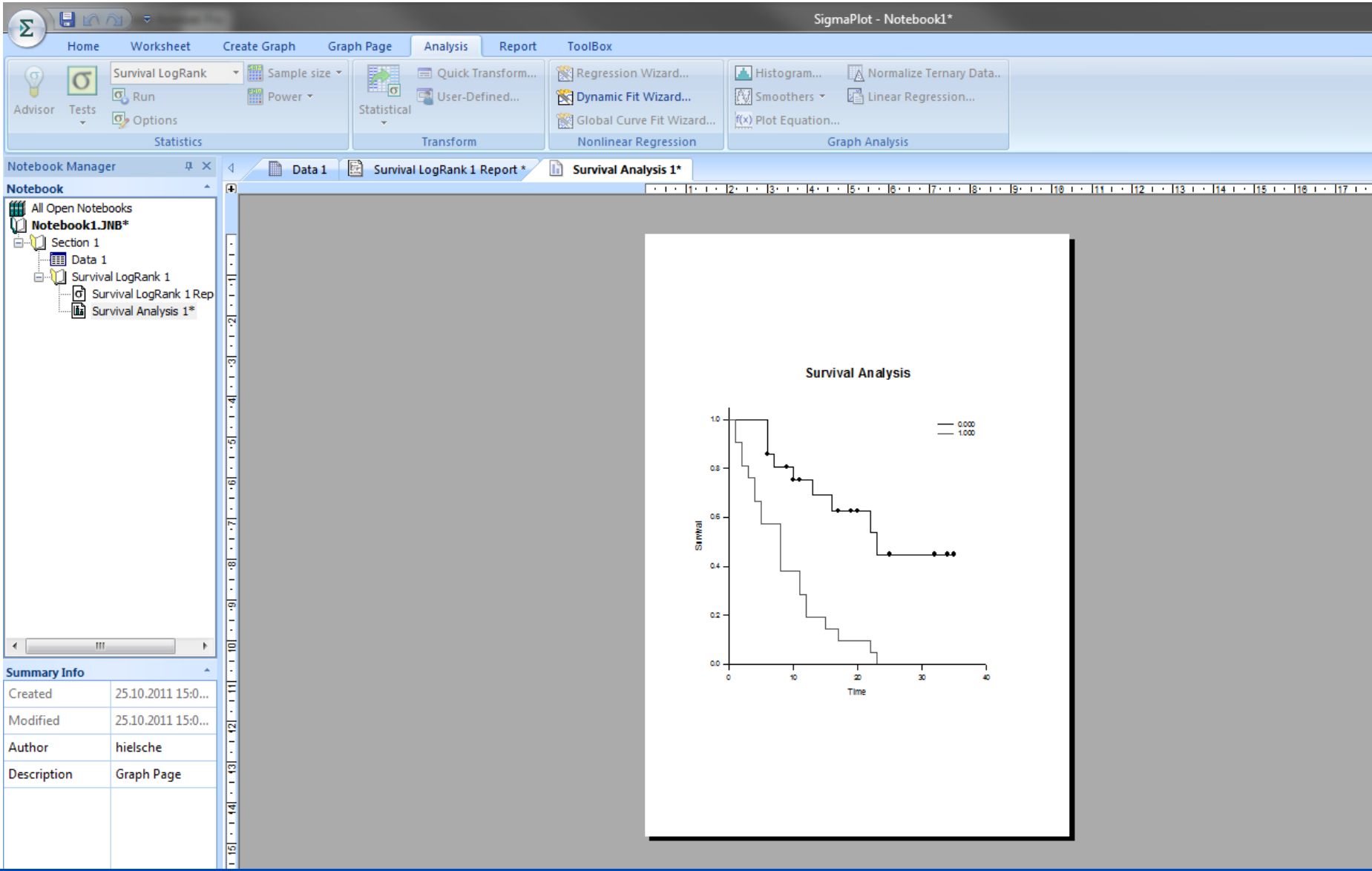


	1-time	2-status	3-group	4	5	6	7	8	9	10	11	12
1	35	0	0									
2	34	0	0									
3	32	0	0									
4	32	0	0									
5	25	0	0									
6	23	1	0									
7	22	1	0									
8	20	0	0									
9	19	0	0									
10	17	0	0									
11	16	1	0									
12	13	1	0									
13	11	0	0									
14	10	0	0									
15	10	1	0									
16	9	0	0									
17	7	1	0									
18	6	0	0									
19	6	1	0									
20	6	1	0									
21	6	1	0									
22	23	1	1									
23	22	1	1									
24	17	1	1									
25	15	1	1									
26	12	1	1									
27	12	1	1									
28	11	1	1									
29	11	1	1									
30	8	1	1									
31	8	1	1									
32	8	1	1									
33	8	1	1									
34	5	1	1									
35	5	1	1									

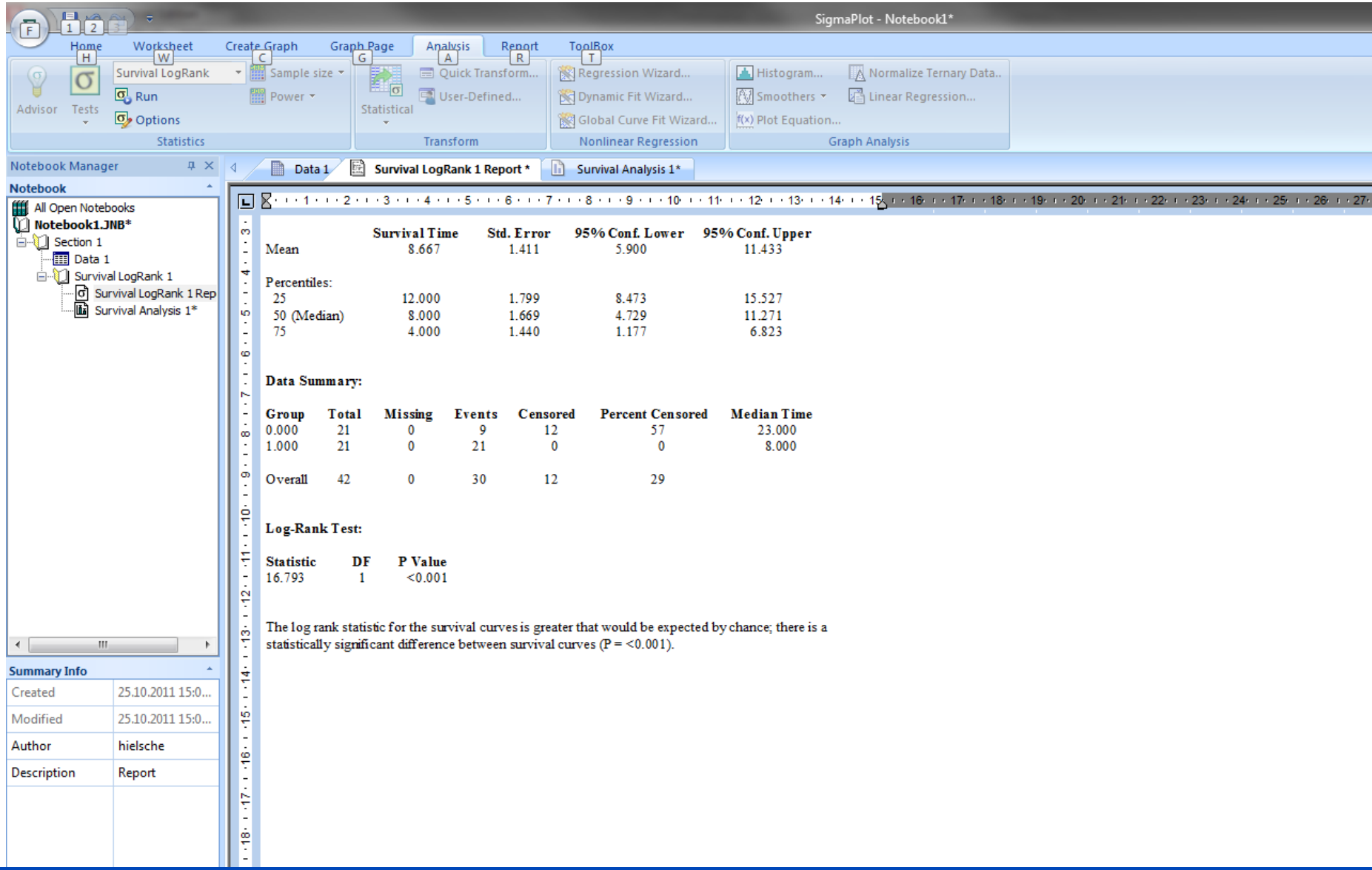
Log-rank test using SigmaPlot? (2)



Log-rank test using SigmaPlot? (3)



Log-rank test using SigmaPlot? (4)



	Survival Time	Std. Error	95% Conf. Lower	95% Conf. Upper
Mean	8.667	1.411	5.900	11.433
Percentiles:				
25	12.000	1.799	8.473	15.527
50 (Median)	8.000	1.669	4.729	11.271
75	4.000	1.440	1.177	6.823

Data Summary:

Group	Total	Missing	Events	Censored	Percent Censored	Median Time
0.000	21	0	9	12	57	23.000
1.000	21	0	21	0	0	8.000
Overall	42	0	30	12	29	

Log-Rank Test:

Statistic	DF	P Value
16.793	1	<0.001

The log rank statistic for the survival curves is greater that would be expected by chance; there is a statistically significant difference between survival curves (P = <0.001).

Time-dependent covariates - ‘immortal bias’

- Group assignment not known at start of time interval
  endpoint: time from therapy start to death
  covariates: response to treatment, transplantation, etc.
- If covariate is treated as if known in advance $\rightarrow$ biased estimates
- E.g. response to treatment: responders must live long enough for response to be observed; there is no such requirement for nonresponders.

Solution:

- Landmark analysis
- Methods to account for time-dependent variables (Simon-Makuch, Cox regression)

Categorization of continuous predictors

Issues:

- loss of power in case there is no real dichotomy: split at median is like discarding a third of the data
- choice of cutpoint
- generalization/validation of cutpoint

How to dichotomize?

- a priori known cutpoint
- distribution parameter (e.g. median): arbitrary
- optimal cutpoint: optimal discrimination w/r/t response, biased estimates/p-values due to type I error inflation

Power and Sample Size

- Number of events **not** number of subjects provides power:
  required events = $\frac{4 \cdot (z_\alpha + z_\beta)^2}{(log(HazardRatio))^2}$,
  for exponentially distributed survival times
- More events can be observed with
  - more subjects
  - longer follow-up
  - combined/early endpoints (PFS vs. OS)
  - prognostically selected population

References

Clark, Bradburn, Love and Altman, 2003. Survival Analysis Part I: Basic concepts and first analyses. British Journal of Cancer 89, 232-238.

Bland and Altman, 2004. The logrank test. British Medical Journal, 328.

SigmaPlot Statistics User Guide (PDF Manuals)

How to get support

- The biostatistics division C060 provides statistical support for all scientific activities of the DKFZ from in vitro and animal studies to human subject.
- Request statistical support via email to biostatistics-consulting@dkfz.de

Next week

- Hazard function
- Cox proportional hazard regression model
- Competing risk analysis