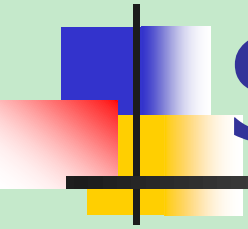
Saarland University, Medical Faculty
PhD-Program



Sample size calculation

PD Dr. Stefan Gräber
Institut für Medizinische Biometrie,
Epidemiologie und Medizinische Informatik
Universität des Saarlandes, Homburg

Sample size calculation

- is an answer to the question: “How large should my study be?”
- There are statistical tools (software) to enable a scientific answer, but
- they require basic considerations!

Sample size calculation

- = definition of sample size prior to the start of the study
- Part of study design
- Contradictory goals:
  - To get results accurate and reliable as possible the sample size should be large
  - To minimize possible risks for the partici-pants the sample size should be small

Kind of sample size calculation

- Estimation/determination: Estimation of sample size for achieving desired statistical assurance of accuracy and reliability
- Justification: Adaption to practical aspects, e.g. budget constraints or medical considerations
- Adjustment: e.g. for dropouts

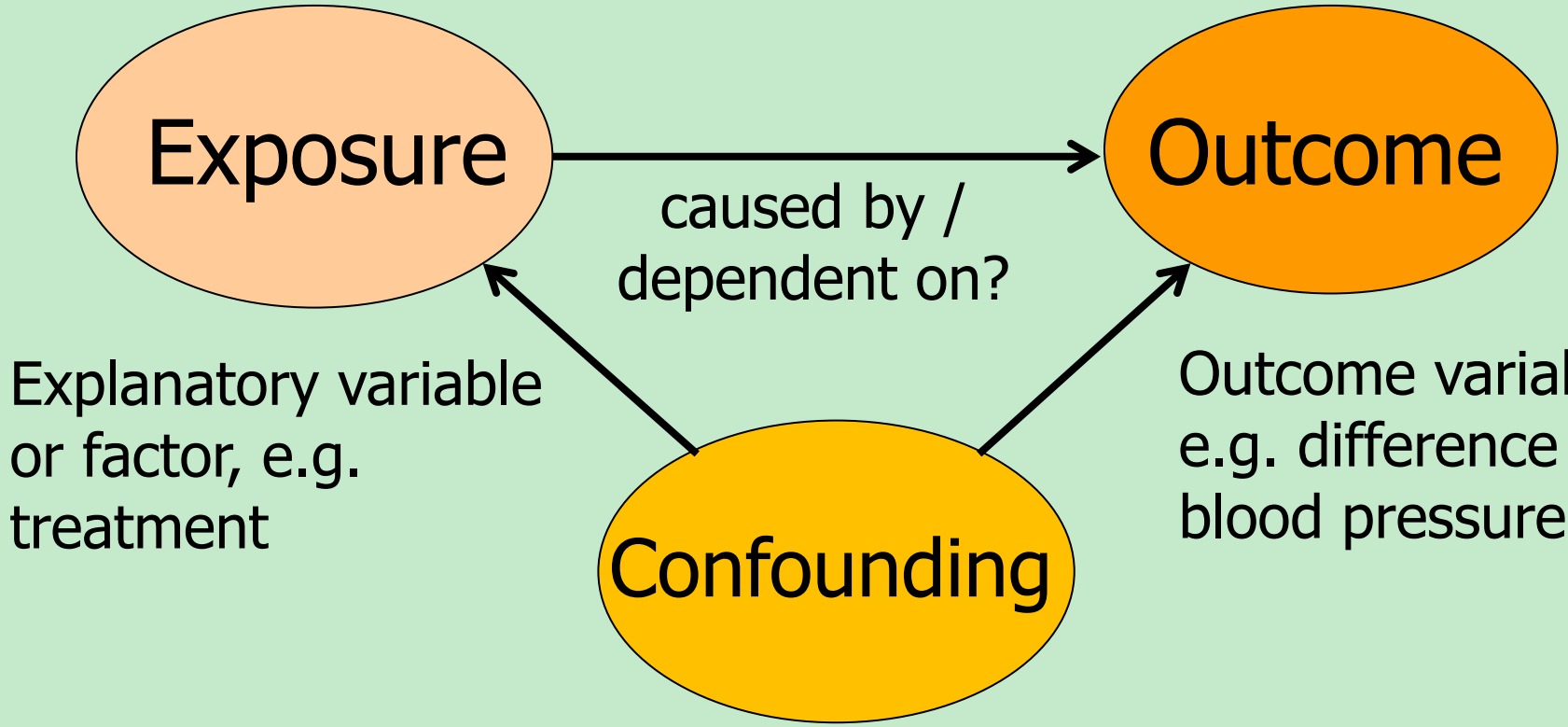
Kind of sample size calculation

- Re-estimation: interim analyses yield cumulative information for new estimation, i.e. you can assess if the selected sample size is sufficient to achieve a desired power at the end of the study ("adaptive study design")

Basic considerations (1)

- Clear definition of question which shall be answered with the study
- Example: Does the new drug A reduce the systolic blood pressure better as the established drug B in patients with hypertension?

Statistical model



Basic considerations (2)

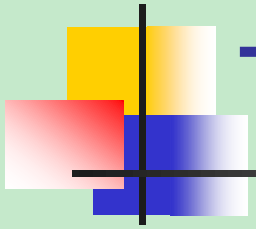
- Definition of primary outcome variable
- Example: The primary outcome variable is the difference of systolic blood pressure before – after the intake of drug (mmHg).

Example of data structure

Variable →

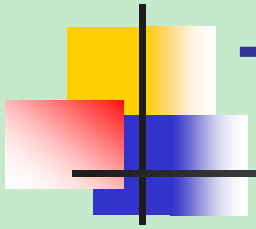
Patient ↓

ID	Treat-ment group	Gender	Age	Size	Weight	Systolic blood pressure before	Systolic blood pressure after	Diffe-rence
1	A	f	63	180	93,0	160	140	20
2	A	m	72	183	79,7	150	145	5
3	A	f	83	165	78,0	170	172	-2
...	...	...	...	...	...	...	...	...
61	B	f	61	165	64,0	150	155	-5
62	B	f	71	173	83,0	165	145	20
63	B	m	79	180	92,3	185	175	10
...	...	...	...	...	...	...	...	...



Types of variable

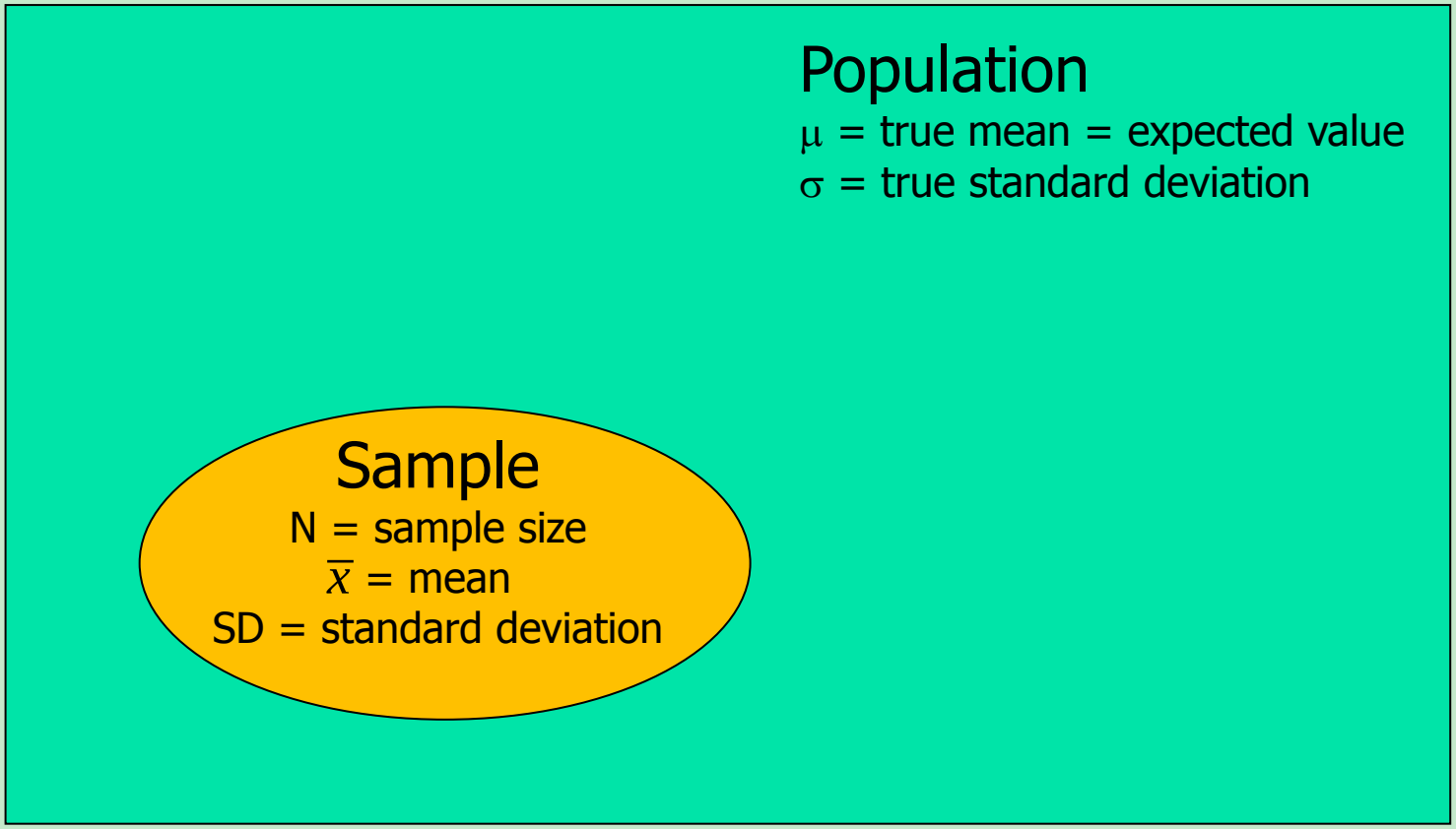
- Categorical (qualitative)
  - Nominal: categories are mutually exclusive and unordered, e.g. gender, eye colour Dichotomous or binary: two categories only, e.g. dead or alive, relapse y/n
  - Ordinal: categories are mutually exclusive and ordered, e.g. disease stage, education level, quality of life

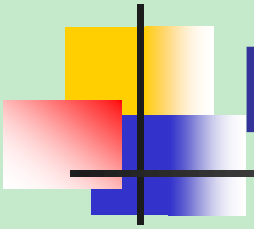


Types of variable

- Numerical (quantitative)
  - Counts (discrete): integer values, e.g. number of pregnancies, number of siblings
  - Continuous (measured): takes any value in a range of values (interval), e.g. blood pressure in mmHg, weight in kg, thickness in mm, age in years

Statistical inference





Example

- Sample: 200 patients with hypertension, i.e. sample size N = 200.
- When the superiority of drug A is proved (based on sample data), potentially all patients with hypertension (=population) could be prescribed the drug, i.e. the result of sample is generalized to the population (statistical inference).

Basic considerations (3)

- Statistical hypothesis = assumption about a circumstance in the population
- Hypotheses are defined using the outcome variable and the clinical meaningfully (relevant) difference.
- Example: The mean reduction of drug A is 20 mmHg and of drug B 10 mmHg, i.e. the clinical meaningfully difference is 20 - 10 = 10 mmHg

Kind of hypotheses

- Null hypothesis H_0 = status quo / no difference / no change / no dependency (converse of the study hypothesis)
- Alternative hypothesis H_1 = possible innovation / issue to be proved / difference / change / dependency (converse of null hypothesis)

Example H_0

- Null hypothesis: Drug A and drug B have the same effect, i.e.
  the mean reduction of blood pressure in the two groups is equal, i.e.
  $\mu_A = \mu_B$, i.e.
  $\delta = \mu_A - \mu_B = 0$

μ = true mean of blood pressure difference

Example H_1

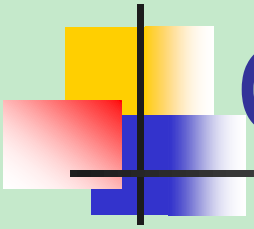
- Alternative hypothesis: Drug A and drug B have different effects, i.e. the mean reduction of blood pressure in the two groups is not equal, i.e. $\mu_A \neq \mu_B$, i.e. $\delta = \mu_A - \mu_B \neq 0$

μ = true mean of blood pressure difference

One-sided hypotheses

- H_0: The mean reduction of blood pressure in group A is lower or equal as in group B, i.e.
  $\mu_A \leq \mu_B$, i.e. $\delta = \mu_A - \mu_B \leq 0$
- H_1: The mean reduction of blood pressure in group A is greater as in group B, i.e.
  $\mu_A > \mu_B$, i.e. $\delta = \mu_A - \mu_B > 0$

μ = true mean of blood pressure difference

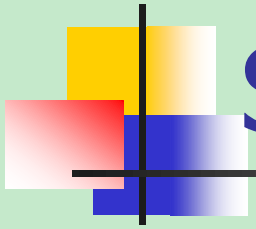


One-sided hypotheses

- are used in non-inferiority trials, i.e. you will prove that the new drug is at least as efficient as the well established drug.

Basic considerations (4)

- Study type: randomized controlled trial with two arms (parallel-group design)



Statistical test

- = statistical procedure to confirm or reject the null hypothesis

Errors with statistical test

Result of test	Population*	
(based on sample data)	H_0 is true	H_1 is true
Test confirms H_0	✓	Type II error
Test rejects H_0	Type I error	✓

* We don't really know whether H_0 is true or false!

Type I error

- = probability of rejecting H_0 although H_0 is true.
- The type I error is controlled by the significance level α, i.e. α is the probability of making type I error.
- Usual values for α are 1% (0,01), 5% (0,05) or 10% (0,1).

Type II error

- = β = probability of confirming H_0 although H_0 is false.
- Power = 1 - β = probability of rejecting H_0 (= obtaining a „statistically significant“ result) when H_0 is truly false.

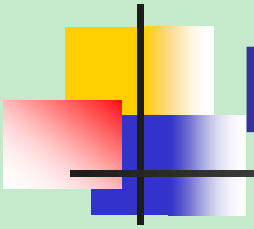
Type II error

- The type II error cannot be controlled because H_1 cannot be specified.
- Example H_1: Drug A and drug B have different effects, i.e. the mean reduction of blood pressure in the two groups is not equal, i.e. $\mu_A \neq \mu_B$, i.e. $\delta = \mu_A - \mu_B \neq 0$, but the true value of δ is unknown.

μ = true mean of blood pressure difference

Choice of statistical test depends on

- Type and distribution of outcome variable
- Kind of hypothesis
- Number of groups
- Paired (related) or independent samples



Example (1)

- Outcome variable is continuous and (approximately) normal distributed
- H_0: $\mu_A = \mu_B$, i.e. comparing means
- Two independent groups
- Appropriate statistical test: t-test for independent samples

Example (2)

- Explanatory and outcome variable are dichotomous
- H_0: Therapy and outcome variable are independent, i.e. comparing rates / proportions
- Appropriate statistical test: Chi-square test

Example Chi-square test

Purpose: To investigate the effect of a virostatic therapy on newborns of HIV-positive mothers.

H_0: P(Infection|Zidovudin)=P(Infection|Placebo)

Therapy	Infection yes	Infection no	Total
Zidovudin	13 (7,2%)	167 (92,8%)	180
Placebo	40 (21,9%)	143 (78,1%)	183
Total	53	310	363

p-value of Chi-square test = 0,000 < 0,05 = α,
i.e. H_0 is rejected

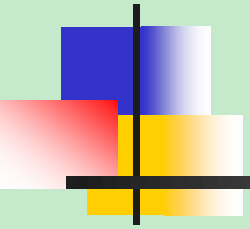
Summary: you must know resp. define

- Type and distribution of outcome variable, Hypotheses
- Information about variability (continuous outcome variable only)
- Kind of statistical test
- Clinical meaningfully difference
- Type I error, type II error

Example

- Comparison of proportions
- Chi-square-test
- Clinical meaningfully difference: 60% - 30% = 30%
- α = 0,05, β = 0,20

Saarland University, Medical Faculty
PhD-Program



Analysis

PD Dr. Stefan Gräber
Institut für Medizinische Biometrie,
Epidemiologie und Medizinische Informatik
Universität des Saarlandes, Homburg

The analysis depends on the type of outcome variable

- Nominal: comparison of proportions by Chi-square test or logistic regression
- Continuous:
  - Comparison of means by t-test or Mann-Whitney U-test
  - Time to event: Kaplan-Meier procedure or Cox-regression

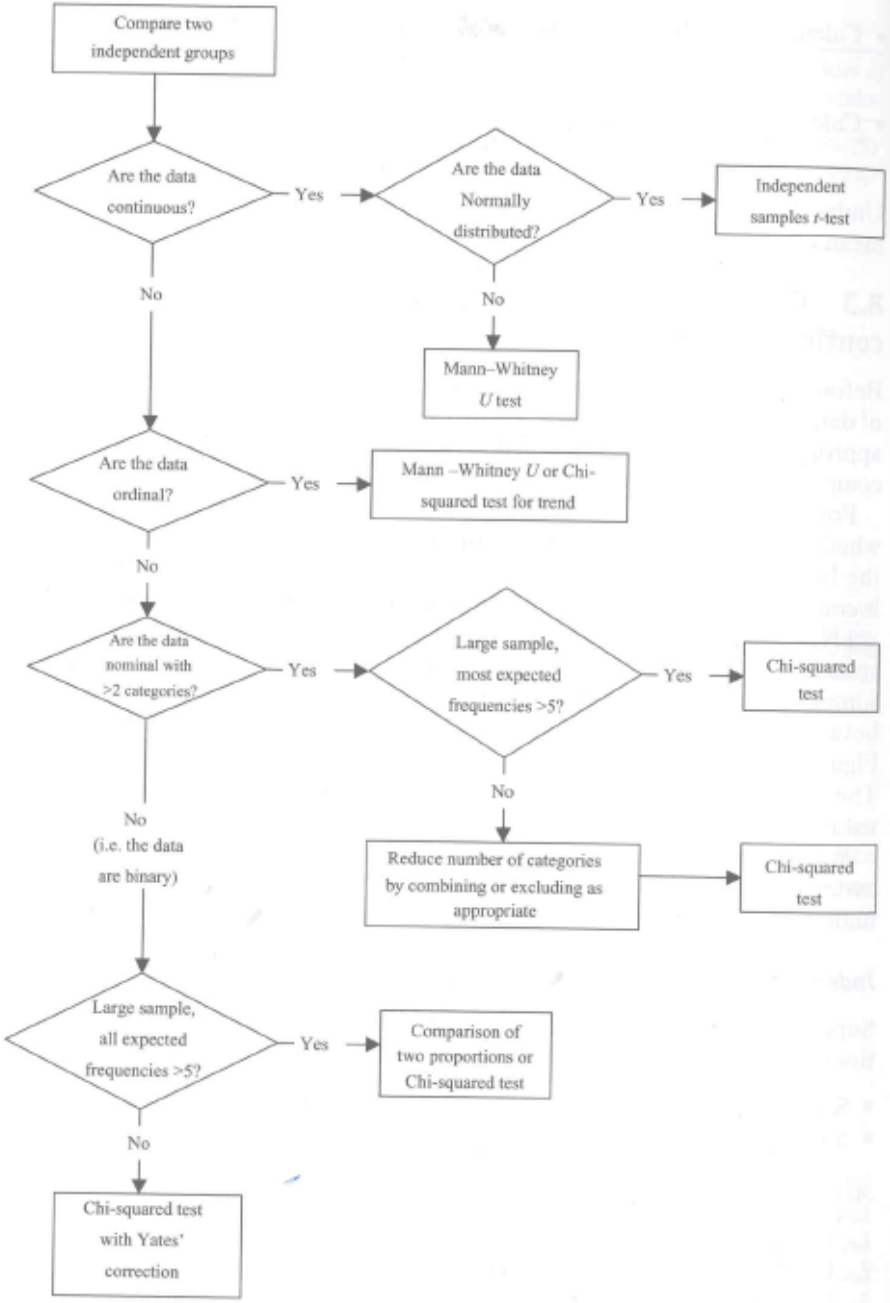


Figure 8.3 Statistical methods for comparing two independent groups or samples

Example: Hypericum study

- Purpose: To investigate the efficacy of hypericum extract LI160 (St John's wort) compared with placebo in patients with mild or moderate major depression.
- Design: Randomized double blind multicenter trial
- Setting: 3 psychiatric primary care units

Example: Hypericum study

- Participants: 89 adult outpatients with mild or moderate depression (Hamilton score < 17)
- Interventions: LI160 or placebo three times a day for four weeks
- Main outcome measure: Change in Hamilton score from baseline to day 28

Hamilton score

Hamilton Rating Scale for Depression

Anwendung: Depressives Syndrom, für psychiatrische Patienten entwickelt, aber auch für Patienten mit anderen Diagnosen verwendbar
Bereich: Überwiegend wissenschaftliche Untersuchungen
Dauer: 5–15 Min.
Ergebnisbereich: 0–55 Punkte: eine hohe Punktzahl charakterisiert einen hohen Schweregrad der Depression
Literatur: Hamilton M. Development of a rating scale for primary depressive illness. Br J Clin Psychol 1967; 6: 278–296.

1. Depressed mood

Sad, hopeless, helpless, worthless

0	Absent
1	Gloomy attitude, pessimism, hopelessness
2	Occasional weeping
3	Frequent weeping
4	Patient reports highlight these feeling states in his/her spontaneous verbal and non-verbal communication.

2. Feelings of guilt

0	Absent
1	Self-reproach, feels he/she has let people down
2	Ideas of guilt or rumination over past errors or sinful deeds
3	Present illness is punishment
4	Hears accusatory or denunciatory voices and/or experiences threatening visual hallucinations. Delusions of guilt

3. Suicide

0	Absent
1	Feels life is not worth living
2	Wishes he/she were dead, or any thoughts of possible death to self
3	Suicide, ideas or half-hearted attempt
4	Attempts at suicide (any serious attempt rates 4)

4. Insomnia, early

0	No difficulty falling asleep
1	Complaints of occasional difficulty in falling asleep i. e. more than half-hour
2	Complaints of nightly difficulty in falling asleep

5. Insomnia, middle

0	No difficulty
1	Patient complains of being restless and disturbed during the night
2	Waking during the night – any getting out of bed rates 2 (except voiding bladder)

6. Insomnia, late

0	No difficulty
1	Waking in the early hours of the morning but goes back to sleep
2	Unable to fall asleep again if he/she gets out of bed

7. Work and activities

0	No difficulty
1	Thoughts and feelings of incapacity related to activities: work of hobbies
2	Loss of interest in activity – hobbies or work – either directly reported by patient or indirectly seen in listlessness, in decisions and vacillation (feels he/she has to push self to work or activities)
3	Decrease in actual time spent in activities or decrease in productivity. In hospital, rate 3 if patient does not spend at least three hours a day in activities
4	Stopped working because of present illness. In hospital rate 4 if patient engages in no activities except supervised ward chores

8. Retardation

Slowness of thought and speech; impaired ability to concentrate; decreased motor activity.

0	Normal speech and thought
1	Slight retardation at interview
2	Obvious retardation at interview
3	Interview difficult
4	Interview impossible

9. Agitation

0	None
1	Fidgetness
2	Playing with hands, hair, obvious restlessness
3	Moving about; can't sit still
4	Hand wringing, nail biting, hair pulling, biting of lips, patient is on the run

10. Anxiety, psychic

Demonstrated by:
- *subjective tension and irritability, loss of concentration*
- *worrying about minor matters*
- *apprehension*
- *fears expressed without questioning*
- *feelings of panic*
- *feeling jumpy*

0	Absent
1	Mild
2	Moderate
3	Severe
4	Incapacitating

11. Anxiety, somatic

Physiological concomitants of anxiety such as:
- *gastrointestinal: dry mouth, wind, indigestion, diarrhoea, cramps, belching*
- *cardiovascular: palpitations, headaches*
- *respiratory: hyperventilation, sighing*
- *urinary frequency*
- *sweating*
- *giddiness, blurred vision*
- *tinnitus*

0	Absent
1	Mild
2	Moderate
3	Severe
4	Incapacitating

12. Somatic symptoms; gastrointestinal

0	None
1	Loss of appetite but eating without encouragement
2	Difficulty eating without urging. Requests or requires laxation or medication for GI symptoms

13. Somatic symptoms; general

0	None
1	Heaviness in limbs, back or head; backaches, headaches, muscle aches, loss of energy, fatiguability
2	Any clear-cut symptom rates 2

14. Genital symptoms

Symptoms such as: loss of libido, menstrual disturbances:

0	Absent
1	Mild
2	Severe

15. Hypochondriasis

0	Not present
1	Self-absorption (bodily)
2	Preoccupation with health
3	Strong conviction of some bodily illness
4	Hypochondriacal delusions

16. Loss of weight

Rate either „A" or „B":

A. When rating by history:

0	No weight loss
1	Probable weight loss associated with present illness
2	Definite (according to patient) weight loss

B. Actual weight changes (weekly):

0	Less than 1 lb (0,5 kg) weight loss in one week
1	1–2 lb (0,5–1,0 kg) weight loss in week
2	Greater than 2 lb (1 kg) weight loss in week
3	Not assessed

17. Insight

0	Acknowledges being depressed and ill
1	Acknowledges illness but attributes cause to bad food, overwork, virus, need for rest, etc.
2	Denies being ill at all

Kommentar: Die Skala versucht, den Schweregrad einer Depression zu erfassen. Dafür werden 17 unterschiedliche Items bewertet. Es handelt sich um eine Fremdbewertungsskala. Sie erfordert psychiatrische Vorkenntnisse bei dem Benutzer. Sie ist weitverbreitet und auch deshalb von Bedeutung.

Change in Hamilton score

Treatment arm	Baseline Mean ± SD	After 4 weeks Mean ± SD
Hypericum (N=42)	15,57 ± 4,10	7,10 ± 3,11
Placebo (N=47)	14,96 ± 4,82	10,45 ± 3,60
p-value Mann-Whitney U-test	0,531	0,000

Responder = Patient whose Hamilton score after 4 weeks was ≤ 8 or decreased at least 50%

Responder rate

Arm * Responder Kreuztabelle

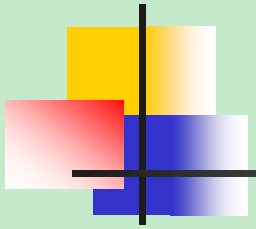
			Responder		
			no	yes	Gesamt
Arm	Hypericum	Anzahl	3	39	42
		% innerhalb von Arm	7,1%	92,9%	100,0%
	Placebo	Anzahl	17	30	47
		% innerhalb von Arm	36,2%	63,8%	100,0%
Gesamt		Anzahl	20	69	89
		% innerhalb von Arm	22,5%	77,5%	100,0%

OR = 0,136
[0,036; 0,506]

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)	Punkt-Wahrscheinlichkeit
Chi-Quadrat nach Pearson	10,727[a]	1	,001	,002	,001	
Kontinuitätskorrektur[b]	9,125	1	,003			
Likelihood-Quotient	11,714	1	,001	,001	,001	
Exakter Test nach Fisher				,002	,001	
Zusammenhang linear-mit-linear	10,606[c]	1	,001	,002	,001	,001
Anzahl der gültigen Fälle	89					

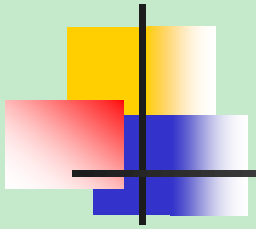
a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 9,44.
b. Wird nur für eine 2x2-Tabelle berechnet
c. Die standardisierte Statistik ist -3,257.



Regression analysis

- is a statistical procedure to answer the question: How do multiple explanatory variables (factors or covariables) influence one outcome variable?

Factor = categorical, Covariable = continuous
Notation: explanatory variables = independent variables, outcome variable = dependent variable



Regression analysis

Dependent on the type of outcome variables there are different procedures:

- Continuous: linear regression (or non-linear regression)
- Binary: logistic regression
- Time-to-event: Cox-Regression

Example logistic regression: APSAC study

	Heparin	APSAC
N	151	162
Men	129 (85%)	139 (86%)
Age [years]	55,9 (9,1)	56,5 (9,2)
Term between infarction and treatment [min]	158,5 (55,5)	172,4 (53,3)
Smoker	93 (61,6%)	95 (58,6%)
Anterior myocardial infarction	71 (47%)	73 (45,1%)

APSAC = Anisoylated Plasminogen Streptokinase Activator Complex

Example logistic regression: APSAC study

Mortality during hospital stay (28 days)	Died	Not died
APSAC	9	153
Heparin	19	132
Total	28	285

Odds ratio: $OR = \frac{9 \cdot 132}{19 \cdot 153} = 0,41$

95%-CI: [0,18;0,93]

Example logistic regression: APSAC study

Outcome variable: Mortality during hospital stay
Coding: 0 = not died, 1 = died

Factor / Covariable	**Coding**
Age	[years]
MI in history	0 = no, 1 = yes
Localisation	0 = anterior, 1 = posterior
Gender	0 = female, 1 = male
Term between infarction and treatment	[min]
Angina pectoris	0 = no, 1 = yes
Hypertension	0 = no, 1 = yes
Diabetes	0 = no, 1 = yes
Hypercholesterolemia	0 = no, 1 = yes
Smoking	0 = nonsmoker, 1 = smoker
Treatment	0 = Heparin, 1 = APSAC

Example logistic regression: APSAC study

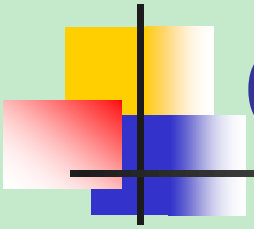
Factor / Covariable	Odds ratio	95%-CI
Treatment	0,37	[0,15;0,90]
Age	1,06	[1,01;1,12]
MI in history	8,09	[2,43;26,91]
Localisation	0,76	[0,33;1,79]
Gender	0,47	[0,17;1,35]
Term between infarction and treatment	1,01	[1,00;1,01]
Angina pectoris	0,88	[0,35;2,20]
Hypertension	1,24	[0,52;2,95]
Diabetes	0,88	[0,20;3,63]
Hypercholesterolemia	1,79	[0,53;6,05]
Smoking	0,88	[0,34;2,28]

Analysis of time to event (survival analysis)

- Outcome variable is the time upon an event occurs. Event may be death, relapse, remission, healing. If event = death: survival time
- Special features:
  - Measurement of time
  - Incomplete observations (censored data)
  - Mostly no normal distribution

Measurement of time

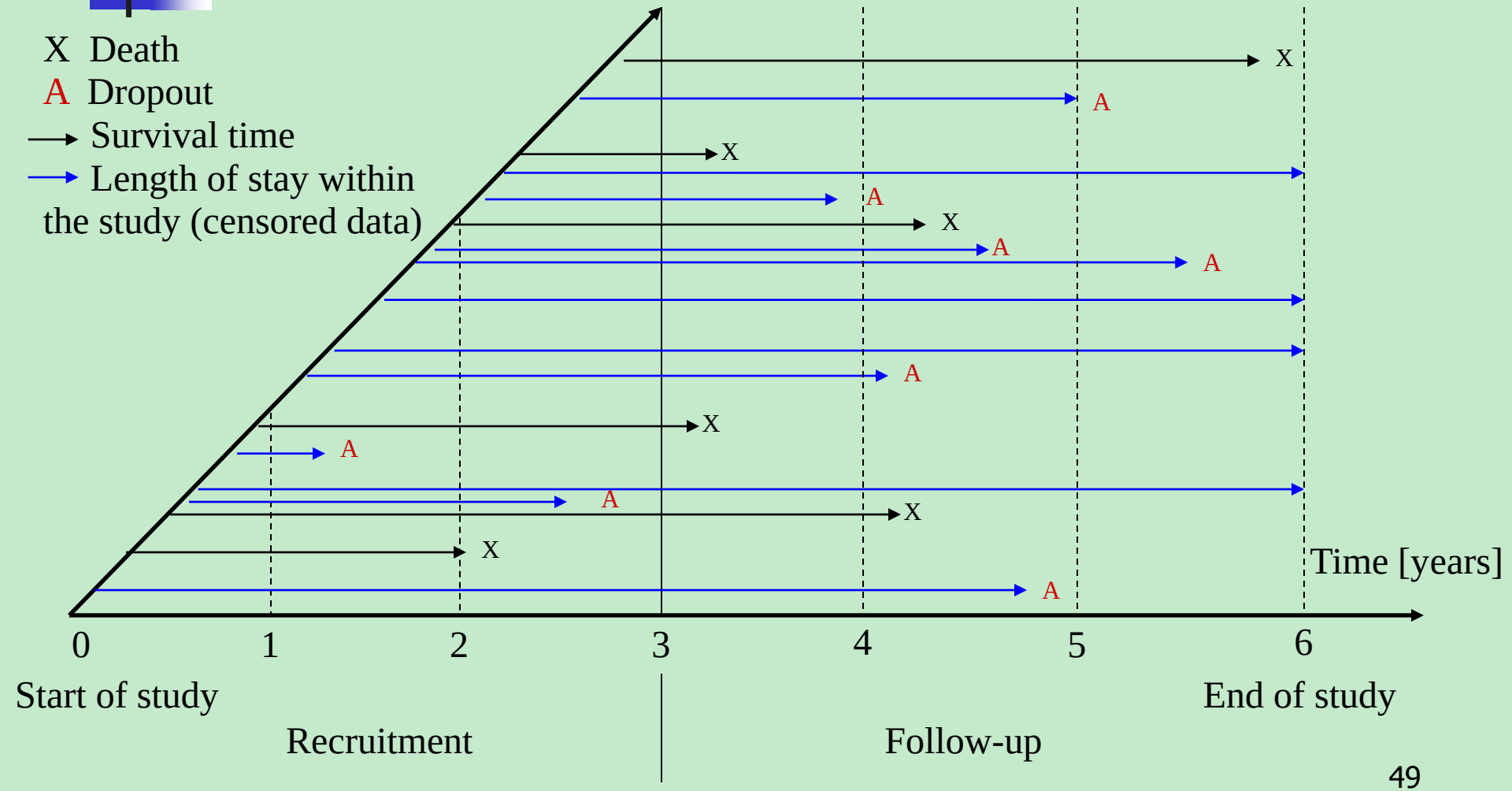
- Start
  - Date of randomization for RCTs
  - Date of diagnosis or therapy for observational studies
- End
  - is sometimes difficult to determine exactly, e.g. data of relapse

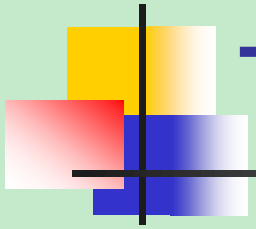


Censoring

- occurs when the value of time to event is only partially known.
- Reasons:
  - End of follow-up
  - Dropout
    - Due to therapy, e.g. severe allergy
    - Other reasons, e.g. lost of contact

Study history (example)



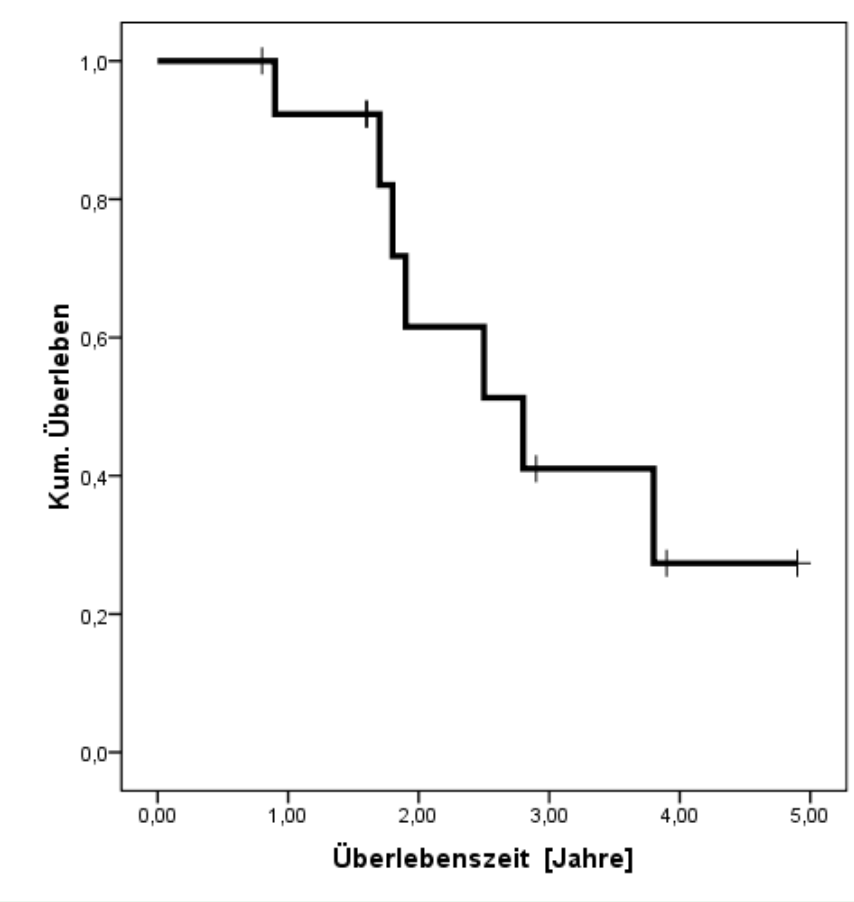


The Kaplan-Meier procedure

- allows the estimation of survival rate taking account of the information provided by censored observations.

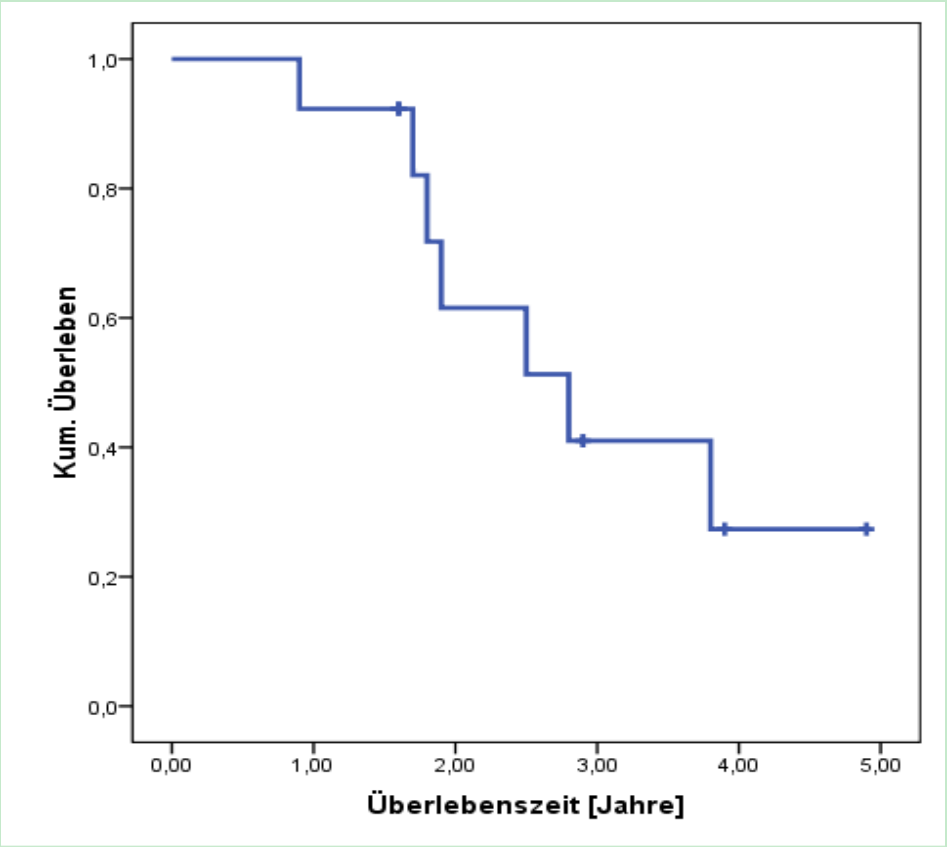
Example 1

	Zeit	Status	Kumulierter Anteil Überlebender zum Zeitpunkt	
			Schätzer	Standardfehler
1	,900	1,00	,923	,074
2	1,600	,00	.	.
3	1,600	,00	.	.
4	1,600	,00	.	.
5	1,700	1,00	,821	,117
6	1,800	1,00	,718	,140
7	1,900	1,00	,615	,153
8	2,500	1,00	,513	,158
9	2,800	1,00	,410	,156
10	2,900	,00	.	.
11	3,800	1,00	,274	,153
12	3,900	,00	.	.
13	4,900	,00	.	.

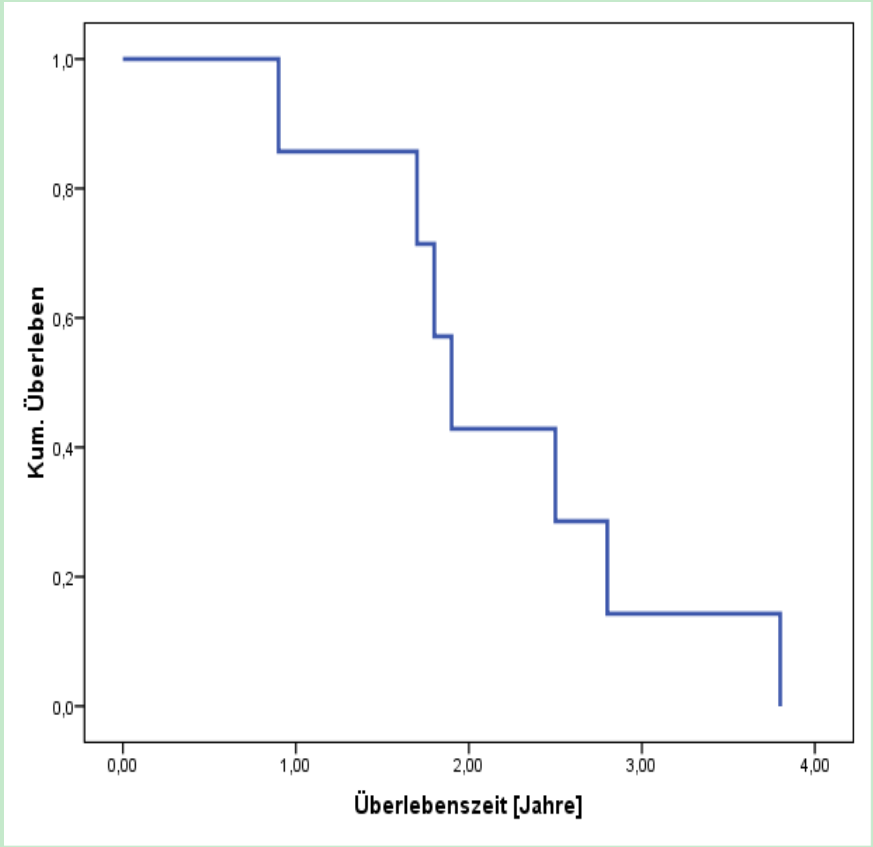


Kaplan-Meier survival curve

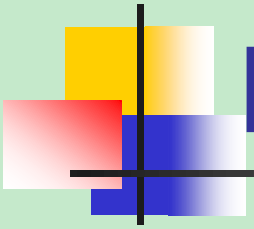
Example 1



With censored data



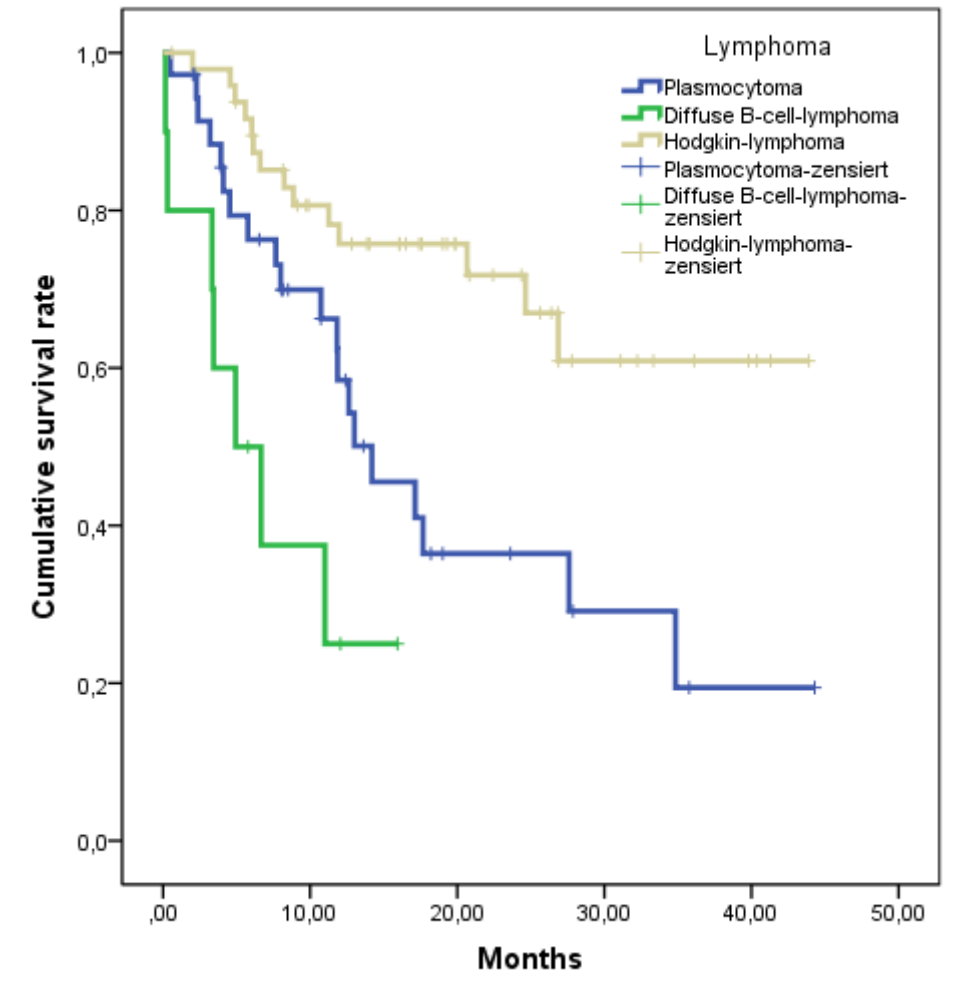
Without censored data



Kaplan-Meier curve with factor

- To compare several groups you can specify a categorical variable as factor, e.g. type of disease.
- Example: Comparison of three types of lymphoma: Hodgkin-lymphoma (N=49), plasmocytoma (N=36), and diffuse B-cell lymphoma (N=10)

Example 2: Survival curves



p-value Log-Rank-Test = 0,000

Cox-regression

- Is a method to analyze the impact of multiple factors or quantitative covariables (independent variables) to a time to event variable (dependent variable), i.e. you can include several factors in one statistical analysis.

Example 2: Cox regression

Factor / Covariable	Hazard ratio	95%-CI
Age	1,002	[0,980;1,023]
Hospital (1=Köln, 2=Hannover)	2,821	[1,225;6,494]
Lymphoma = 0 (Hodgkin-lymphoma)	1,000	Reference category
Lymphoma = 1 (Plasmocytoma)	1,844	[0,731;4,653]
Lymphoma = 2 (Diffuse B-cell lymphoma)	10,080	[3,347;30,360]
Therapy (1=A, 2=B)	0,545	[0,291;1,021]