

Medizinische Informatik Klinische Versuchsplanung

Definition: Klinische Studie

Eine klinische Studie untersucht die Wirkung (und ihr(e) Maß/Größe) einer (oder mehrerer) Intervention(en) an Menschen.

- Intervention (z.B.):
- Behandlung (Vergleich zw. Behandlungsmethoden)
 - Anwesenheit eines Faktors, Faktoren bei Krankheiten (z.B.: Übergewicht als Risiko für Herz-Kreislauf-Erkrankungen)

(Einige) Gebiete der klinischen (statistischen) Studien

- **Fallzahl-schätzung, Konfidenzintervalle**
- Survival-Analyse
- Äquivalenznachweise
- Randomisation und Verblindung
- Phase I-IV - Studien

Entscheidung	Wirklichkeit	
	$H_0: \pi_1 = \pi_2$	$H_0: \pi_1 \neq \pi_2$
Nicht-Ablehnung von H_0 (Annahme von H_0)	Richtige Nicht-Ablehnung	Falsche Nicht-Ablehnung (β-, Typ II-Fehler)
Ablehnung von H_0	Falsche Ablehnung (α-, Typ I-Fehler)	Richtige Ablehnung

Asymptotischer 4-Felder Test zum Vergleich von 2 Raten

		Erfolg		
		+	-	
Gruppe	1	a	b	N_1
	2	c	d	N_2
				N

$$N = N_1 + N_2;$$

$$N_1 = a + b;$$

$$N_2 = c + d;$$

$$r_1 = a / (a + b) = a / N_1; \quad r_2 = c / (c + d) = c / N_2$$

für genügend große N-werte:
 $r_1 \sim \pi_1; r_2 \sim \pi_2$

Testhypothesen:
 $H_0: \pi_1 = \pi_2$ vs. $H_1: \pi_1 \neq \pi_2$

Transformation der Raten zur Prüfgröße:

$$Z = \frac{(r_1 - r_2)}{\sqrt{\frac{r_1(1-r_1)}{N_1} + \frac{r_2(1-r_2)}{N_2}}}$$

$$Z = \frac{(r_1 - r_2)}{\sqrt{\frac{r_1(1-r_1)}{N_1} + \frac{r_2(1-r_2)}{N_2}}}$$

Entscheidungsregel über Normal-Approximation:

$H_0: Z \rightarrow \mathcal{N}(0,1)$ Standardnormalverteilung

$$|Z| \geq Z_{(1-\frac{\alpha}{2})}: H_0: \text{ablehnen}$$

Beispiele:

	Erfolg	kein Erfolg	
Intervention	80	20	100
Kontrolle	70	30	100

$$Z_{\text{errechnet}} = 1.6440$$

$$\alpha = 0.05 \quad 1-\alpha/2 = 0.975 \quad |Z| \geq Z_{(1-\frac{\alpha}{2})}: Z_{0,975} = 1.9600$$

	Erfolg	kein Erfolg	
Intervention	800	200	1000
Kontrolle	700	300	1000

$$Z_{\text{errechnet}} = 5.1988$$

$$\text{wenn } |Z| \geq Z_{(1-\frac{\alpha}{2})}: H_0: \text{ablehnen}$$

mit Excel:

Erfolg	kein Erfolg		
80	20	100	NI
70	30	100	NK
150	50	200	
$r_{\text{Intervention}}=r_1= 0.8$			
$r_{\text{Kontrolle}}=r_2= 0.7$			
$\alpha = 0.05$			
$1-\alpha/2 = 0.975$			
$ Z _{\text{errechnet}} = 1.6440$			$Z = \frac{(r_1 - r_2)}{\sqrt{\frac{r_1(1-r_1)}{N_1} + \frac{r_2(1-r_2)}{N_2}}}$
$Z_{(1-\alpha/2)=0.975} = 1.9600$			

$Z_{\text{errechnet}} < Z_{(1-\alpha/2)}: H_0$ zu behalten

$$r_1=80/100; r_2=70/100; N_1=NI; N_2=NK$$

Fallzahlschätzung für Differenz von 2 (unbekannten) Raten

Raten (d.h. relative Häufigkeiten):

- r_I : Erfolg bei Interventionen/Anzahl-Interventionen
- r_K : Erfolg bei Kontrollen/Anzahl-Kontrollen
- r_u : untere Grenze des Konfidenzintervalls **für Differenz der Raten**
- r_o : obere Grenze des Konfidenzintervalls **für Differenz der Raten**

$$r_u = (r_I - r_K) - z_{(1-\frac{\alpha}{2})} \sqrt{\frac{r_I(1-r_I)}{N_I} + \frac{r_K(1-r_K)}{N_K}}$$

$$r_o = (r_I - r_K) + z_{(1-\frac{\alpha}{2})} \sqrt{\frac{r_I(1-r_I)}{N_I} + \frac{r_K(1-r_K)}{N_K}}$$

$$r_u = (r_I - r_K) - d \quad r_o = (r_I - r_K) + d$$

$$KI_{(1-\alpha)} = (r_I - r_K) \pm d$$

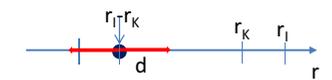
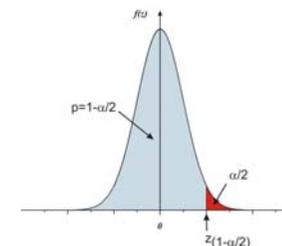
Zweiseitiges Konfidenzintervall für Differenz von 2 (unbekannten) Raten

Signifikanzniveau: α (an beiden Enden der Verteilungsfunktion $\alpha/2$)

Konfidenzniveau: $\gamma=1-\alpha$

$$KI_{\gamma=(1-\alpha)} = (r_I - r_K) \pm d \quad d = z_{(1-\frac{\alpha}{2})} \sqrt{\frac{r_I(1-r_I)}{N_I} + \frac{r_K(1-r_K)}{N_K}}$$

$z_{(1-\frac{\alpha}{2})}$: z-Wert der Standardnormalverteilung bei $p=1-\alpha/2$



„d“ ist auch eine relative Häufigkeit!

Beispiel:

Erfolgsrate bei neuer Intervention: $r_1=0.8$;

Erfolgsrate mit alter Methode: $r_2=0.6$;

In beiden Untersuchungen sei der Stichprobenumfang 100;

Wie groß ist das Konfidenzintervall mit einer $\gamma=0,95$ (d.h. $Kl_{0,95}$) für die Differenz der Raten?

E1	80		d=	0.1240
E2	60		$r_{\text{untere}}=$	0.076
N1	100		$r_{\text{obere}}=$	0.324
N2	100			
r1	0.8			
r2	0.6			
r1-r2=	0.2			
alpha=	0.05			
$1-\alpha/2=$	0.975			
$z_{1-\alpha/2}=$	1.960			
$r=(r_1+r_2)/2=$	0.7			

Problem:

Wir möchten die **Länge** (2d) des Konfidenzintervalls verkleinern (d.h. bei dem selben Konfidenzniveau $\gamma=0,95$). Wie viele Fälle soll man beobachten (wie groß muss der Umfang sein)?

Fallzahlschätzung für Differenz von 2 (unbekannten) Raten

$$d = z_{\left(1-\frac{\alpha}{2}\right)} \sqrt{\left[\frac{r_I(1-r_I)}{N_I} + \frac{r_K(1-r_K)}{N_K} \right]} \quad \text{mit } N_I=N_K=N$$

$$N = \frac{z_{\left(1-\alpha/2\right)}^2}{d^2} \sqrt{[r_I(1-r_I) + r_K(1-r_K)]}$$

Beispiel:

Erfolgsrate bei neuer Intervention: $r_1=0.8$;

Erfolgsrate mit alter Methode: $r_2=0.7$;

$d=0.05$

$N=?$



4-Felder-Test Fallzahlschätzung für Differenz von 2 (unbekannten) Raten

	Erfolg	kein Erfolg	
Intervention	80	20	100
Kontrolle	70	30	100

$$r_1=0.8; r_k=0.7; \alpha=0.05; \gamma=0.95; (1-\alpha/2)=0.975; \quad (1-\alpha/2) = \frac{1+\gamma}{2}$$

Formulierung der Fragestellung:

- die alte Methode (z.B. Behandlung) ergibt 70% Erfolg
- mit neuer Methode ist 80% Erfolg erwartet
- **wie groß muss der Stichprobenumfang sein, diesen Unterschied statistisch abzusichern?**
- **wir möchten einen α -Fehler [α] und einen β -Fehler [β] erreichen.**

H_0 :

- **Die zwei Raten übereinstimmen, d.h. die Differenz der Raten/Häufigkeiten tritt wegen Zufall auf** (andere Formulierung: die Differenz der Raten/Häufigkeiten weicht nicht signifikant von Null ab.)
- Diese Differenz kann sowohl kleiner, als auch größer sein.
- Zweiseitige Fragestellung.

Formel für asymptotischen Fall:

Bezeichnungen:

N_1, N_2 (z.B.: N_I, N_K) — Vereinfachung: $N_1=N_2=N$;

π_1, π_2 (z.B.: r_I, r_K);

$\pi = (\pi_1 + \pi_2) / 2$;

α und β als entsprechende Fehler;

zweiseitige Fragestellung:

$$N \approx (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2 [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)] / (\pi_1 - \pi_2)^2$$

einseitige Fragestellung:

$$N \approx (z_{1-\alpha} + z_{1-\beta})^2 [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)] / (\pi_1 - \pi_2)^2$$

Güte $1-\beta$, auf Grund N $z_{1-\beta} = \frac{|\pi_2 - \pi_1| N^{\frac{1}{2}}}{(2\pi(1 - \pi))^{1/2}} - z_{1-\frac{\alpha}{2}}$

Beispiel 1.:

$\pi_1=0,8$;
 $\pi_2=0,7$;
 $\alpha = 0,05$ (2-seitig)
 $\beta = 0,05$ ($1-\beta=0,95$)

$p_1 =$	0.8	$z_{1-\alpha/2}$	1.960
$p_2 =$	0.7	$z_{1-\beta}$	1.645
$\alpha =$	0.05	Term 1.	12.99
$\beta =$	0.05	Term 2.	0.37
$1-\beta =$	0.95	Term 3.	0.01
		$N =$	481.00



Nerrechnet=	481	Zaehler=	2.19317122
p1=	0.8	Nenner=	0.612372436
p2=	0.7	Z(1-beta) berechnet,N=	1.621
alpha=	0.05	beta(berechnet, N)=	0.052
beta (gezielt)=	0.05		
pi=	0.75		
$z_{1-\alpha/2}$	1.960		
$z_{1-\beta}$ (gezielt)	1.645		

Literatur:

<http://imsieweb.uni-koeln.de/lehre/q1/Q1-06-Konfidenzintervalle.pdf>

<http://imsieweb.uni-koeln.de/lehre/klinstud/KlinStud07-Fallzahlschaetzungen.pdf>

http://de.wikipedia.org/wiki/Konfidenzintervall_einer_unbekannten_Wahrscheinlichkeit

Überlebenszeitanalyse: Der Log-Rang-Test

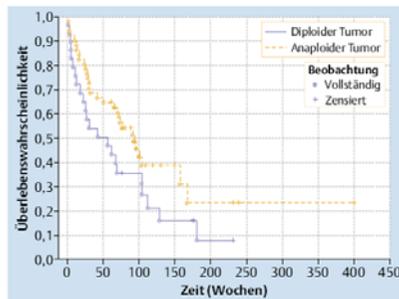


Abb. 1 Kaplan-Meier Kurven für die Überlebenszeit der 80 Zungenkrebspatienten. Es wird in orange/blau die Wahrscheinlichkeit gezeigt, dass ein Patient mit aploidem/diploidem Tumor eine Zeit (in Wochen) überlebt.

- Fallzahlschätzung, Konfidenzintervalle
- **Survival-Analyse**
- Äquivalenznachweise
- Randomisation und Verblindung
- Phase I-IV - Studien

Dtsch Med Wochenschr 2007; 132: e39–e41 · A. Ziegler et al., Überlebenszeitanalyse: Der Log-Rang-Test

Überlebens-Daten

Sind Zeiten bis zum Auftreten eines Ereignisses, Lebensdauern/Zeitdauern (Zeitintervalle) von Start- bis Ziel- bzw. Endereignis

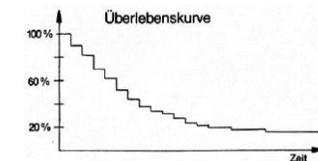
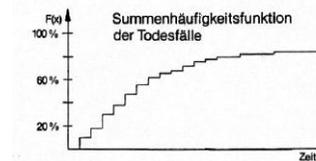
Ist das Endereignis „Tod“ aufgetreten, → „Lebensdauern/Überlebenszeiten“

Definitionen:

$F(t)$ gibt den Anteil der Individuen an, die zum Zeitpunkt $t, t \geq 0$, bereits gestorben sind.

$S(t) = 1 - F(t)$ ist die Überlebenskurve. Sie gibt den Anteil der Individuen zum Zeitpunkt t an, die noch leben.

Auf Grund der Definitionen, die $F(t)$ ist eine Summenhäufigkeitskurve.



Kaplan-Meier-Kurve

Begriffe/Definitionen:

Ereigniszeiten t_j ; $0 <= t_1 < t_2 < \dots < t_j$ Zeitpunkte der Beobachtungen;

n_j : Anzahl der Individuen unter Risiko zum Zeitpunkt t_j (einschließlich später zensierter);

d_j : Anzahl der Ereignisse (Todesfälle) zum Zeitpunkt t_j ;

$r_j = d_j / n_j$ ist der Anteil der zum Zeitpunkt t_j Gestorbenen unter denjenigen, die t_{j-1} überlebt haben (bzw. die zu t_j unter Risiko stehen);

$s_j = (n_j - d_j) / n_j = (1 - d_j / n_j)$ ist der Anteil der den Zeitpunkt t_j Überlebenden unter denjenigen, die t_{j-1} überlebt haben (bzw. zu t_j unter Risiko stehen);

$S(t_i) = s_1, S(t_2) = S(t_1) \cdot s_2 = s_1 \cdot s_2, S(t_j) = S(t_{j-1}) \cdot s_j = s_1 \cdot s_2 \cdot \dots \cdot s_j$ für $t_j \leq t$

$S(t_j)$ ist ein Punkt der Überlebenskurve — K-M-Kurve

Mediane Überlebenszeit: der Zeitpunkt t_M , an dem die Hälfte der Patienten/Individuen noch lebt.

Logrank-Test (Mantel-Haenszel)

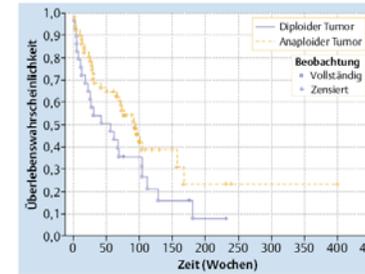


Abb. 1 Kaplan-Meier Kurven für die Überlebenszeit der 80 Zungenkrebspatienten. Es wird in orange/blau die Wahrscheinlichkeit gezeigt, dass ein Patient mit aploidem/diploidem Tumor eine Zeit (in Wochen) überlebt.

Der Log-Rang-Test ist das Standardverfahren in der Überlebenszeitanalyse für einfache Gruppenvergleiche in klinisch-therapeutischen Studien. Mit diesem nichtparametrischen Test lässt sich statistisch überprüfen, ob das Mortalitätsrisiko in zwei oder mehr Gruppen verschieden ist.

Der Logrank-Test ist ein (nichtparametrischer) Signifikanztest, der 2 (unabhängige) Survivalfunktionen auf Gleichheit überprüft, unter Berücksichtigung von zensierten Daten.

Zensierte Daten (auch trunkierte (gestutzte) Daten) sind Daten, bei denen nicht alle Werte einer Statistische Variablen bekannt sind; z. B.: das Ereignis ist noch nicht eingetreten/wenn Kontakt zum Patienten abgebrochen ist.

Literatur:

Dtsch Med Wochenschr 2007; 132: e39–e41 · A. Ziegler et al., Überlebenszeitanalyse:

Der Log-Rang-Test;

<http://imsieweb.uni-koeln.de/lehre/klinstud/KlinStud04-Survival.pdf>

Äquivalenznachweise

- Fallzahlschätzung, Konfidenzintervalle
- Survival-Analyse
- **Äquivalenznachweise**
- Randomisation und Verblindung
- Phase I-IV - Studien

Problem/Frage:

Ist ein statistisch signifikanter Unterschied auch klinisch relevant?

$\delta = \Delta_r = \mu_1 - \mu_2$: ist die kleinste Überlegenheit von μ_1 vs. μ_2 , die klinisch relevant ist.

$A = (-\varepsilon - \delta)$ ist Äquivalenzbereich, z.B.: $A = (-\delta - \delta)$

μ : Mittelwert, Häufigkeit,...

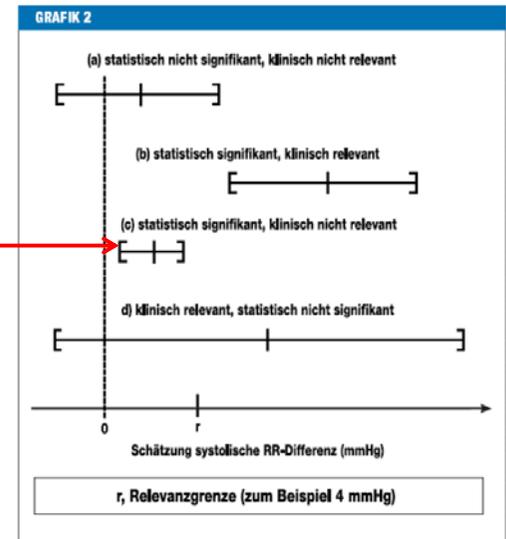
$KI_{(1-2\alpha)}$ -Konfidenzintervall für $\Delta = \mu_T - \mu_R$ (T: treatment, R-Referenz)

obere- und untere einseitige $(1-\alpha)$ -KI

Behandlungen sind „äquivalent“ mit Sicherheitswahrscheinlichkeit $(1-\alpha)$, wenn A beinhaltet das ganze $(1-2\alpha)$ -KI

du Prel J-B, Hommel G, Röhrig B, Blettner M, 2009: Konfidenzintervall oder p-Wert? Teil 4 der Serie zur Bewertung wissenschaftlicher Publikationen. Dtsch Arztebl 106, 335–9

KI enthält nicht "die Null" als Unterschied, erreicht aber die Relevanzgrenze nicht.



Statistische Signifikanz und klinische Relevanz

Bioäquivalenz

Quotient der Erwartungswerte der beiden Behandlungen $\Delta = \mu_T / \mu_R$

Bei Bioäquivalenz gibt der Äquivalenzbereich $A = (80\% - 125\%)$ den bioäquivalenten Bereich an.

Ursache: $\lg(0,8) \sim -0,1$ $\lg(1,25) \sim +0,1$;

→ die Abweichung ist kleiner als $\pm 0,1$ für einen Quotient.

Randomisation (Randomisierung/Zufallszuteilung)

Sie ist eine Technik für jeden Patienten zufällig eine Zuteilung zu einer Behandlungsgruppe zu realisieren, wobei alle Gruppen mit gleicher Wahrscheinlichkeit ausgewählt werden können.

- Fallzahlschätzung, Konfidenzintervalle
- Survival-Analyse
- Äquivalenznachweise
- **Randomisation und Verblindung**
- Phase I-IV - Studien

Ziel:

- bewussten und unbewussten Störgrößen auszuschalten;
- zu verhindern den Prüfer-Bias (z.B.: abhängig von weiteren Umständen teilt der Prüfer einen Patienten zu einer oder anderer Gruppe zu: Verzerrungen (Bias) können in Behandlungseffekten auftreten);
- zu garantieren, dass statistische Tests gültige/valide Signifikanz-Niveaus ergeben.

Randomisierungstechniken

- **einfache Randomisierung:** Zuteilung auf Grund erzeugten Zufallszahlen
 - Münzwurf, Würfeln, Zufallszahlentabellen,....;
 - Nachteil (z.B.):
 - Umfänge der Behandlungsgruppen kann unterschiedlich sein — Verschlechterung der Teststärke (Güte $1-\beta$)
 - Ungleichverteilung der Störfaktoren in den Gruppen
- **Blockrandomisation:** in einem Block bestimmter Länge (z.B. 2,4,6...) sind die Patienten zufällig einer Behandlungsart zugeteilt (AABB, ABAB...);
 - Balanciertheit der Gruppen/Zuteilungen ist möglich
 - Nachteil (z.B.):
 - zufällige Blocklänge
 - bei bekannten Blocklänge kann Code aufgebrochen werden
- **Stratifizierte (geschichtete) Randomisierung:**
Ursache: bekannte/unbekannte Risiko-Faktoren, Unbalanciertheiten können die Auswertungen/Ergebnisse verfälschen.
 - Schichten auf Grund z.B.: Alter, Geschlecht, Stadium, Zentren, Ausgangslage,...
 - Randomisierung innerhalb einer Strata
 - Vorteil (z.B.): Varianz und die Einflüsse der Störfaktoren nehmen ab.

Verblindung

In klinischen Studien spricht man von **Verblindung**, wenn

- die Prüfarzte, das Pflegepersonal,
- die teilnehmende Patienten und auch
- die Personen, die mit dem Monitoring, dem Datenmanagement und der Auswertung der Studie betraut sind,

nicht über die individuelle Behandlungszuteilung der Patienten **informiert sind** und ihre Handlungen somit nicht durch dieses Wissen beeinflusst sein können.

Die Verblindung hat das Ziel, systematischen Unterschieden in der Behandlung der Patienten oder der Bewertung des Therapieerfolgs vorzubeugen.

Verblindungsniveaus:

unverblindet (offen)

- Chirurgie, Diät etc.
- Teilnehmer und Untersucher kennen Therapiegruppe
- Bias in beiden Richtungen

einfach-blind

- Teilnehmer blind

doppel-blind

- Teilnehmer und Untersucher blind

dreifach-blind

- Teilnehmer, Untersucher, Auswerter etc. blind

Notfälle - zum Beispiel schwere Nebenwirkungen
- können jedoch bei einzelnen Probanden die
verfrühte Aufdeckung der Zuordnung zu den
Untersuchungsgruppen notwendig machen
(Entblindung)

Phase I-IV - Studien

Phasenmodell klinischer Studien

- Fallzahlschätzung, Konfidenzintervalle
- Survival-Analyse
- Äquivalenznachweise
- Randomisation und Verblindung
- **Phase I-IV - Studien**

Phase 0: Präklinische Entwicklung

Ziele:

1. Abklärung möglicher **toxischer Effekte**, wie Einfluss auf zahlreiche in Laboruntersuchungen bestimmte Größen (Klinische Chemie, Hämatologie), Fertilität, Embryotoxizität/Teratogenität, Cancerogenität
2. Abklärung **sicherheitspharmakologischer Aspekte**, wie Beeinträchtigung von Herz/Kreislauf, Einfluss auf Körpergewicht
3. Hinweise auf **erwünschte pharmakologische Effekte** in vitro/in vivo

Phase I: Erstanwendung am Menschen

Meist gesunde Freiwillige („Probanden“), gegebenenfalls besondere Patientengruppe (z. B. bei Studien mit Zytostatika)

Ziele:

- Verträglichkeit, Pharmakokinetik/-dynamik
- Hinweis auf wirksame Dosis (eventuell)/ Arzneimittelinteraktionen

Phase II: Einstieg in die therapeutische Anwendung am Patienten

Begrenzte Zahl von Patienten der anvisierten Indikation

Ziele:

- Verträglichkeit und Dosisfindung
- Wirkung (pharmakologische Effekte)/Wirksamkeit (Heilerfolg)
- Pharmakokinetik in Spezialfällen (z. B. Leber-, Nierenerkrankung)

Phase III: Breite Anwendung im anvisierten Indikationsgebiet, Beleg für die

Einsetzbarkeit als Arzneimittel [Zulassung] Patienten der anvisierten Indikation in Klinik/Praxis

Ziele:

- Beleg der Wirksamkeit an Patienten in unterschiedlichen Populationen
- Ausreichende Beurteilung der Verträglichkeit, besondere Patientengruppen
- Verhalten unter Langzeitbehandlung, Vergleich mit etablierter Therapie

Phase IV: Klinische Prüfung nach der Zulassung: Erkenntniserweiterung über die Substanz, Einsatz unter Praxisbedingungen

Einsatz an großer Zahl von Patienten entsprechend den Vorgaben der Zulassungsbehörden (unter Praxisbedingungen)

Ziele:

- Quantifizierung seltener Nebenwirkungen
- Detailuntersuchungen in bestimmten Patientengruppen
- Einfluss auf Spätfolgen einer Erkrankung (Folgemorbidität, Letalität)
- Tatsächlicher Einsatz des Präparates (→ „Anwendungsbeobachtungen“)
- Hinweis auf weitere Indikationen, zu modifizierende Dosis (→ Phase II)