

Liczebność próbki w randomizowanych próbach klinicznych

- ◆ Testowanie hipotez i p-stwo błędu
- ◆ Kryteria oceny skuteczności leczenia o rozkładzie normalnym
 - wyznaczanie liczebności próbki
 - nierównoliczne grupy
 - nie-podrzędność i równoważność
 - więcej niż dwie grupy leczenia
- ◆ Kryteria binarne
- ◆ Czas do wystąpienia zdarzenia
- ◆ Nieistotne statystycznie wyniki próby

Testowanie hipotez

- ◆ Próby fazy III są porównawcze.
- ◆ Wnioskowanie z reguły oparte o testy istotności statystycznej.
- ◆ Dwie hipotezy:
 - zerowa (np. “brak różnicy w skuteczności leczenia”)
 - alternatywna (np. “różnica w skuteczności leczenia”)
- ◆ $p\text{-value} = P(\text{Test statistic} > \text{observed} \mid H_0) < \alpha$
 - zazwyczaj $\alpha = 0.05$

$$\alpha = 5\%$$

The value for which $P=0.05$, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation ought to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant. Using this criterion we should be led to follow up a false indication only once in 22 trials, even if the statistics were the only guide available. Small effects will still escape notice if the data are insufficiently numerous to bring them out, but no lowering of the standard of significance would meet this difficulty.

R.A. Fisher, Statistical Methods for Research Workers (1925)

Dwa rodzaje błędu

	Prawda (nieznana)	
Wynik testu (znany)	Hipoteza zerowa	Hipoteza alternatywna
Istotny ($p < \alpha$) (odrzucaamy h. zerową)	BŁĄD I RODZAJU (α)	OK $(1 - \beta)$
Nieistotny ($p > \alpha$) (nie odrzucaamy h. zerowej)	OK $(1 - \alpha)$	BŁĄD II RODZAJU (β)

- ◆ *Wynik testu ZAWSZE może być błędny !*
- ◆ Możemy jedynie kontrolować p-stwo błędu

Kontrola prawdopodobieństwa błędu

- ◆ Błąd I rodzaju: użycie odpowiedniego poziomu istotności testu.
- ◆ Błąd II rodzaju: p-stwo (β) zależy od
 - poziomu istotności testu;
 - (oczekiwanej) różnicy w skuteczności leczenia;
 - zmienności oszacowań miar efektu leczenia;
 - liczebności próbki.

Wyznaczanie liczebności próbki

- ◆ Próba ze zbyt małą liczbą chorych może nie wykryć różnicy w skuteczności leczenia
 - problem etyczny: chorzy zostaną niepotrzebnie poddani ryzyku efektów niepożądanych itp.
- ◆ Próby kliniczne są kosztowne
 - równowaga między liczebnością próbki a kosztem
- ◆ Liczebność wyznaczana w oparciu o planowaną metodę analizy

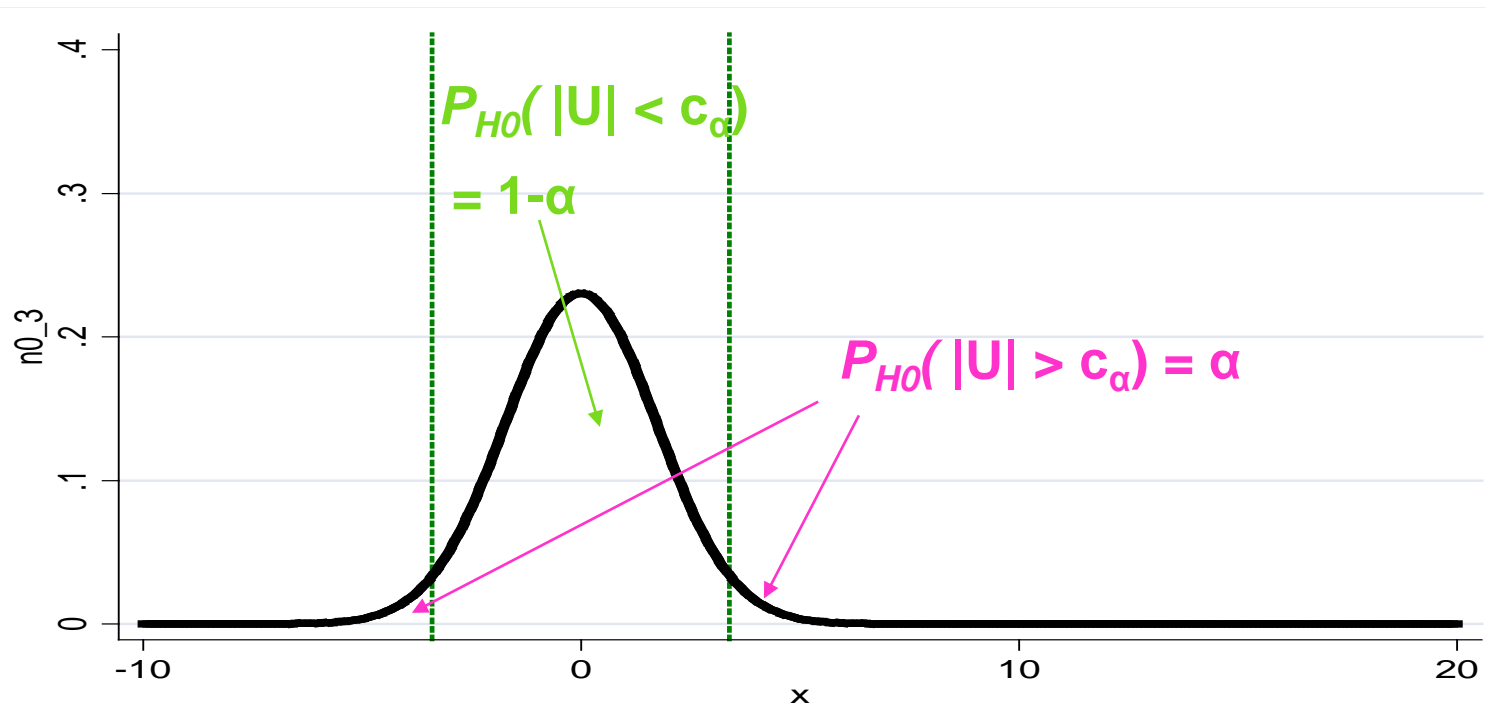
Kryteria oceny skuteczności leczenia o rozkładzie normalnym

- ◆ Dwie grupy
- ◆ W eksperymentalnej, kryterium $\sim N(\mu_E, \sigma^2)$
- ◆ W kontrolnej, kryterium $\sim N(\mu_C, \sigma^2)$
- ◆ Liczebność próbki w każdej z grup = n
- ◆ Niech $\Delta = \mu_E - \mu_C$
- ◆ Rozważamy statystyki:

$$U = \bar{X}_E - \bar{X}_C \sim N\left(\Delta, \frac{2\sigma^2}{n}\right)$$
$$Z = \frac{U}{\sigma\sqrt{2/n}} \sim N\left(\frac{\Delta}{\sigma\sqrt{2/n}}, 1\right)$$

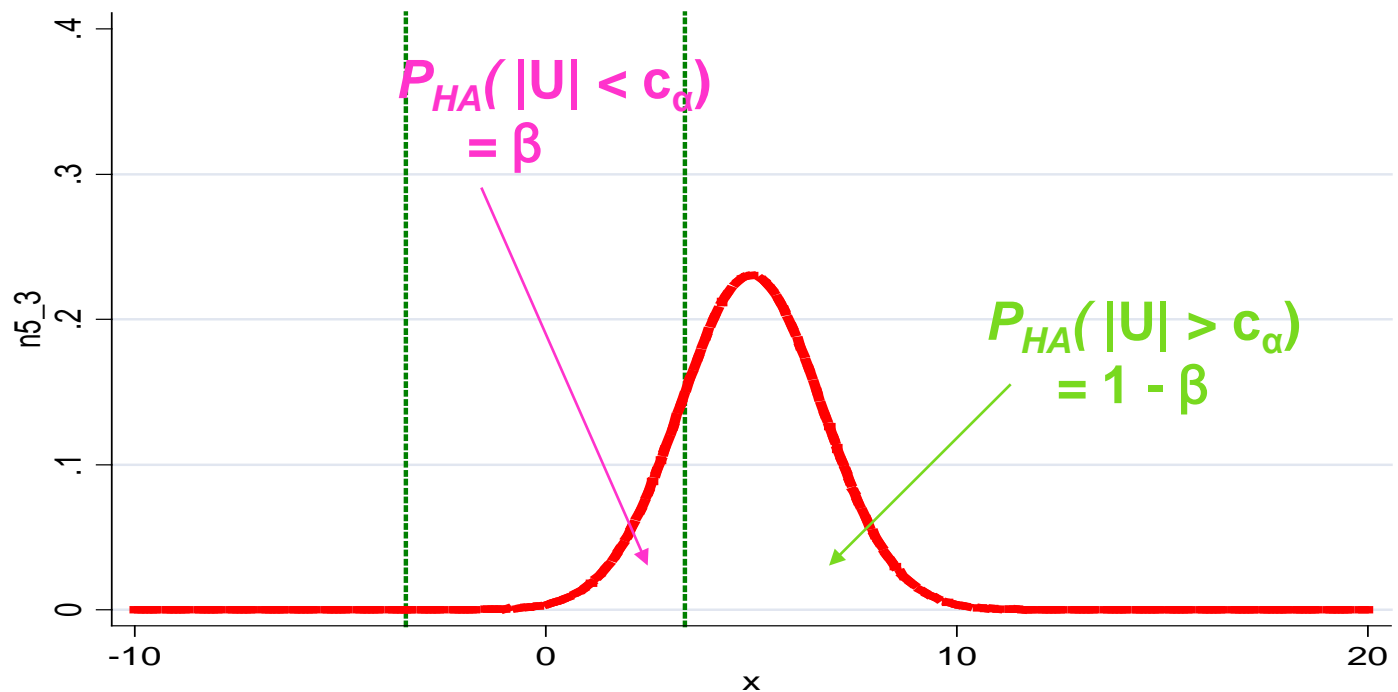
Kryteria oceny skuteczności leczenia o rozkładzie normalnym

- ◆ Dla H_0 : $\Delta = 0$, $U \sim N(0, 2\sigma^2/n)$, $Z \sim N(0,1)$
- ◆ Test dwustronny: H_A : $\Delta \neq 0$
- ◆ Błąd I rodzaju: $\alpha = P_{H_0}(|Z| > z_{1-\alpha/2}) = P_{H_0}(|U| > (2\sigma^2/n)^{1/2} z_{1-\alpha/2})$
 - $z_{1-\alpha/2} = 1.96 \rightarrow \alpha = 0.05$; $z_{1-\alpha/2} = 2.57 \rightarrow \alpha = 0.01$



Kryteria oceny skuteczności leczenia o rozkładzie normalnym

- ◆ Rozważmy H_A z $\Delta > 0$. Wówczas
 $U \sim N(\Delta, 2\sigma^2/n)$, $Z \sim N(\Delta / (2\sigma^2/n)^{1/2}, 1)$
- ◆ Błąd II rodzaju: $\beta = P_{HA}(|Z| < z_{1-\alpha/2}) = P_{HA}(|U| < (2\sigma^2/n)^{1/2} z_{1-\alpha/2})$



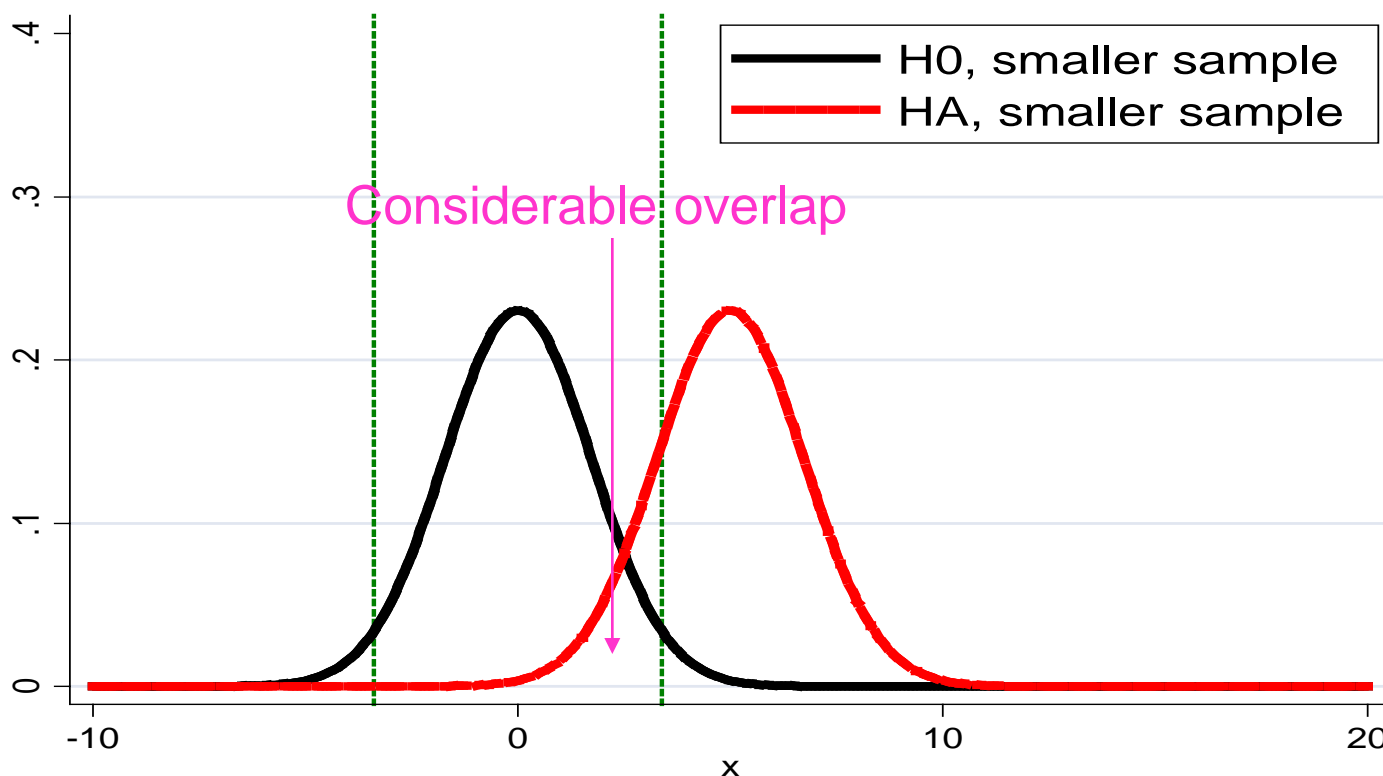
Prawdopodobieństwo błędu II rodzaju

$$\begin{aligned}\beta &= P_{HA} \left(|U| \leq z_{1-\alpha/2} \sigma \sqrt{\frac{2}{n}} \right) = P_{HA} \left(-z_{1-\alpha/2} \sigma \sqrt{\frac{2}{n}} \leq U \leq z_{1-\alpha/2} \sigma \sqrt{\frac{2}{n}} \right) \\ &= P_{HA} \left(-z_{1-\alpha/2} - \frac{\Delta}{\sigma \sqrt{2/n}} \leq \frac{U - \Delta}{\sigma \sqrt{2/n}} \leq z_{1-\alpha/2} - \frac{\Delta}{\sigma \sqrt{2/n}} \right) \\ &= \Phi \left(z_{1-\alpha/2} - \frac{\Delta}{\sigma \sqrt{2/n}} \right) - \Phi \left(-z_{1-\alpha/2} - \frac{\Delta}{\sigma \sqrt{2/n}} \right)\end{aligned}$$

- ◆ β zależy od α , Δ , σ^2 oraz n .
 - możemy sensownie zmieniać jedynie n !

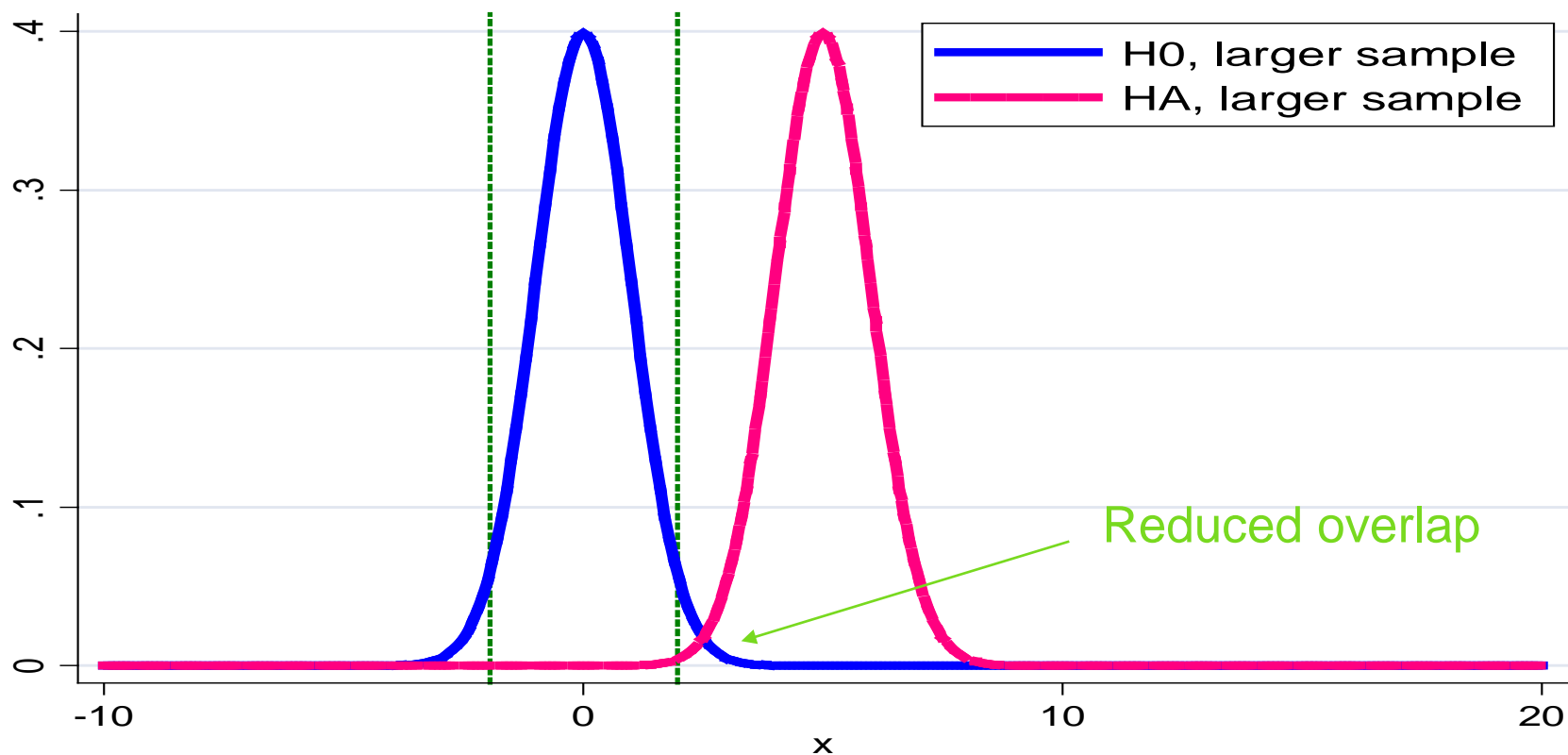
Prawdopodobieństwo błędu II rodzaju

- ◆ Przyjmijmy $\alpha=0.05$, $\Delta = 5$, $\sigma^2 = 15$.
- ◆ Niech $n = 10 \rightarrow 2\sigma^2/n = 3 \rightarrow \Delta / (2\sigma^2/n)^{1/2} = 2.89$.
 $\beta = \Phi(1.96-2.89) - \Phi(-1.96-2.89) = 0.177 - 10^{-6} \approx 0.177$

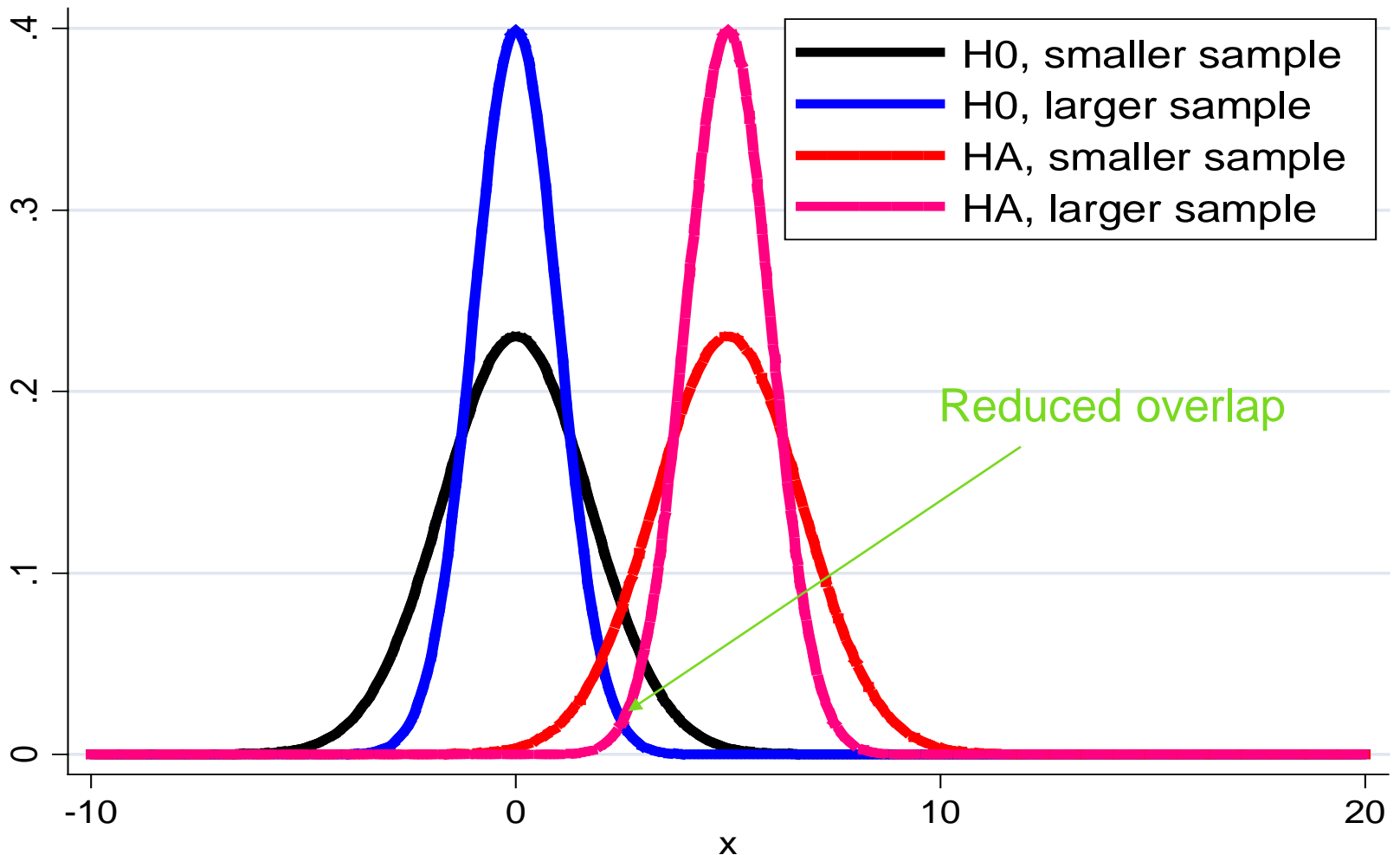


Prawdopodobieństwo błędu II rodzaju

- ◆ Przyjmijmy $\alpha=0.05$, $\Delta = 5$, $\sigma^2 = 15$.
- ◆ Niech $n = 30 \rightarrow 2\sigma^2/n = 1 \rightarrow \Delta / (2\sigma^2/n)^{1/2} = 5$.
 $\beta = \Phi(1.96-5) - \Phi(-1.96-5) \approx 0.001$



Prawdopodobieństwo błędu II rodzaju: wpływ liczebności próbki



Wymagana liczebność próbki

- ♦ Jak duże powinno być n dla założonej wartości β ?
- ♦ Ustalamy α , Δ oraz σ^2 .

$$\beta \approx \Phi\left(z_{1-\alpha/2} - \frac{\Delta}{\sigma\sqrt{2/n}}\right)$$

\Leftrightarrow

$$z_{\beta} = z_{1-\alpha/2} - \frac{\Delta}{\sigma\sqrt{2/n}} \Rightarrow$$

$$n = \frac{2(z_{1-\beta} + z_{1-\alpha/2})^2}{\left(\frac{\Delta}{\sigma}\right)^2}$$

Wymagana liczebność próbki: przykład

- ◆ Przyjmijmy $\alpha=0.05$ ($z_{1-\alpha/2}=1.96$), $\Delta = 5$, $\sigma^2 = 15$ ($\Delta / \sigma = 1.29$).
- ◆ Założone p-stwo błędu II rodzaju: $\beta = 0.2$ ($z_{1-\beta}=0.84$).
 - Jeśli $\Delta = 5$, próba daje 80% p-stwo odrzucenia hipotezy zerowej.

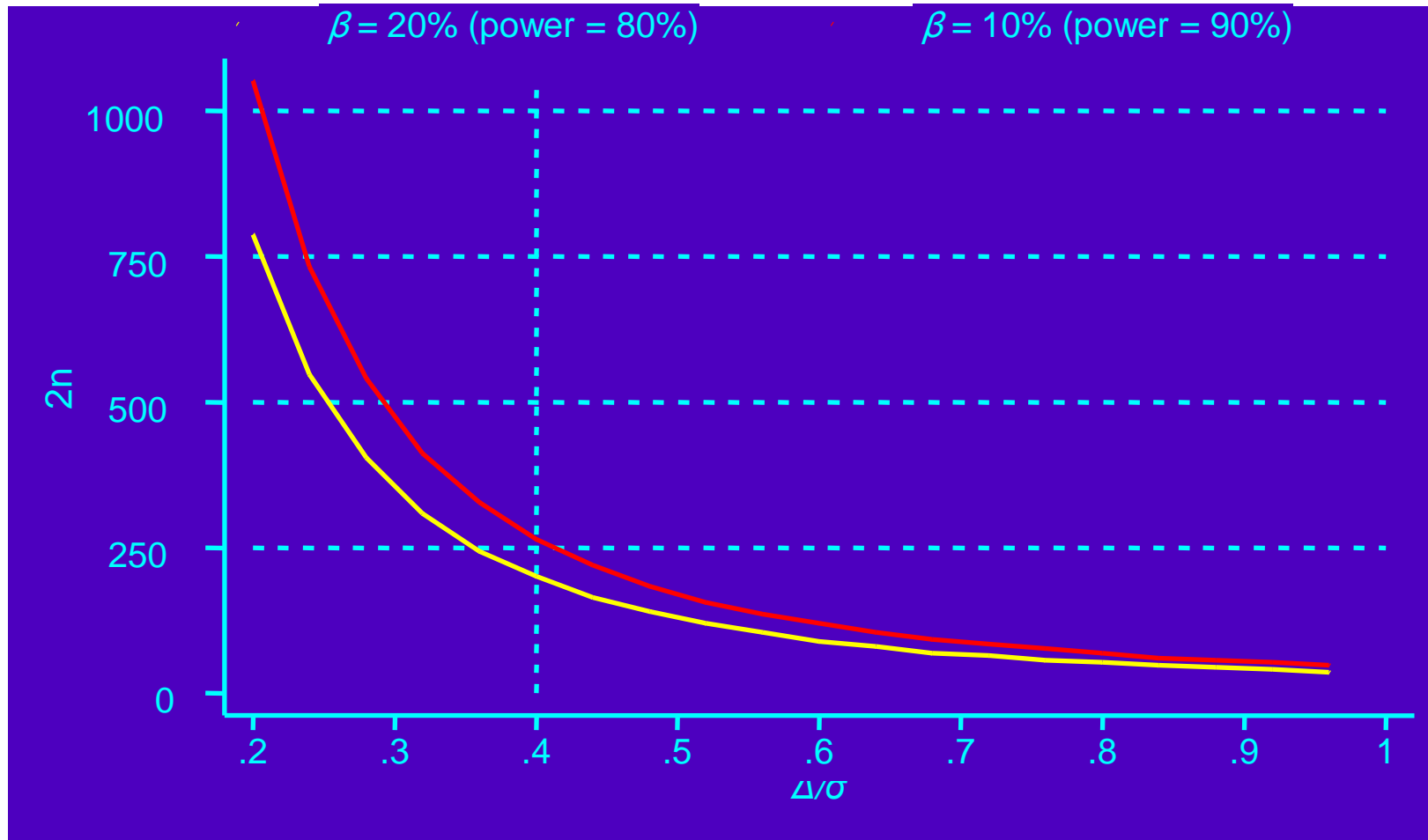
$$n = 2(1.96+0.84)^2/1.29^2 = 9.4 \approx 10.$$

- ◆ Dla $\beta = 0.1$ ($z_{1-\beta}=1.28$)

$$n = 2(1.96+1.28)^2/1.29^2 \approx 13.$$

Zależność liczebności próbki od parametrów testu

- ◆ $H_0: \Delta = 0$ vs. $H_A: \Delta \neq 0$ (zakładamy > 0).
- ◆ $\alpha = 0.05$ (dwustronny).



Wymagana liczebność próbki

- ◆ Uwzględniamy chorych niespełniających kryteria doboru
 - zwiększamy liczebność o 5% lub 10%
- ◆ Dla testu jednostronnego, przyjmujemy $z_{1-\alpha}$.
- ◆ Najczęściej, $\beta = 0.2, 0.1$, lub 0.05 , oraz $\alpha = 0.05$ lub 0.01 .
- ◆ $n = 2 \times A(\alpha, \beta) / (\text{stand. diff})^2$, gdzie $A(\alpha, \beta)$ dane w tabeli

β	α	
	0.01	0.05
0.05	17.8	13
0.1	14.9	10.5
0.2	11.7	7.85

Nierównoliczne grupy

- ◆ Przyjmijmy liczebności n oraz qn .

- ◆ Wówczas

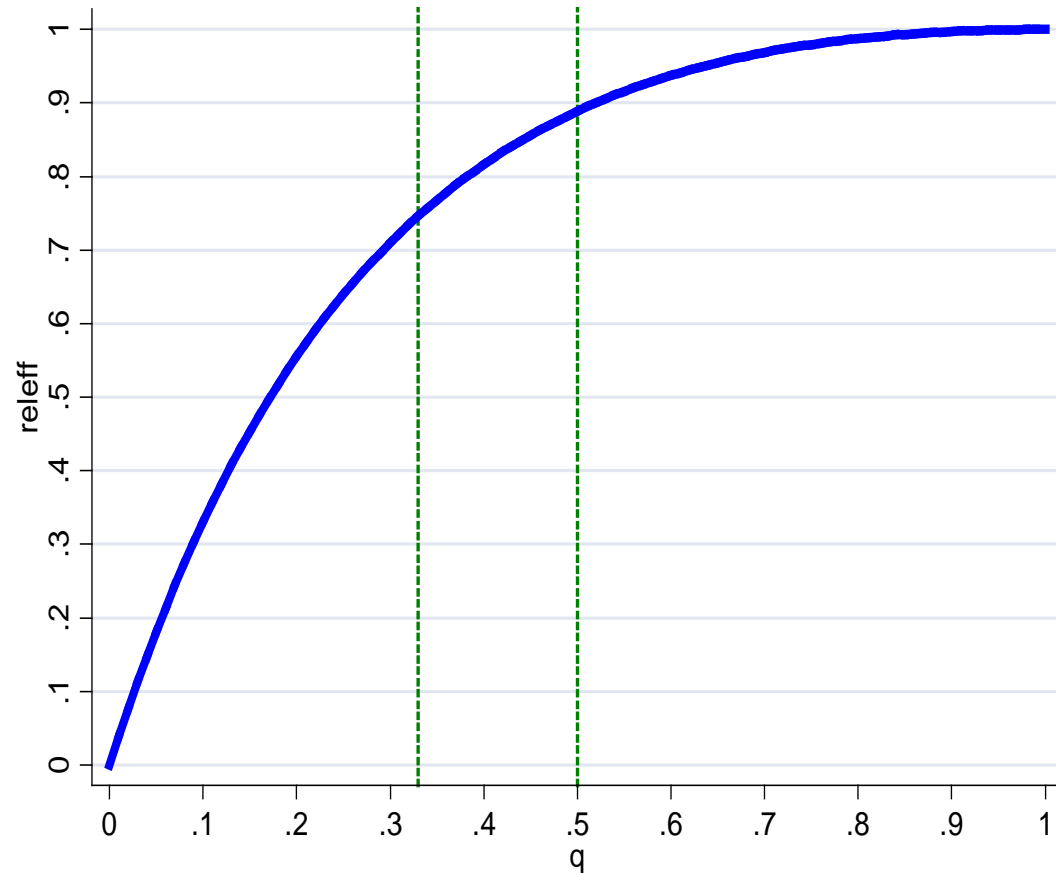
$$n = \frac{q+1}{q} \frac{(z_{1-\beta} + z_{1-\alpha/2})^2}{\left(\frac{\Delta}{\sigma}\right)^2}$$

- ◆ Równa liczebność: $q = 1 \rightarrow (q+1)/q = 2$ (jak poprzednio)
- ◆ Randomizacja 2:1: $q = 1/2 \rightarrow (q+1)/q = 3$.
- ◆ Względna efektywność (dla tej samej mocy):

$$\frac{\text{Total sample balanced}}{\text{Total sample unbalanced}} = \frac{4q}{(q+1)^2} < 1$$

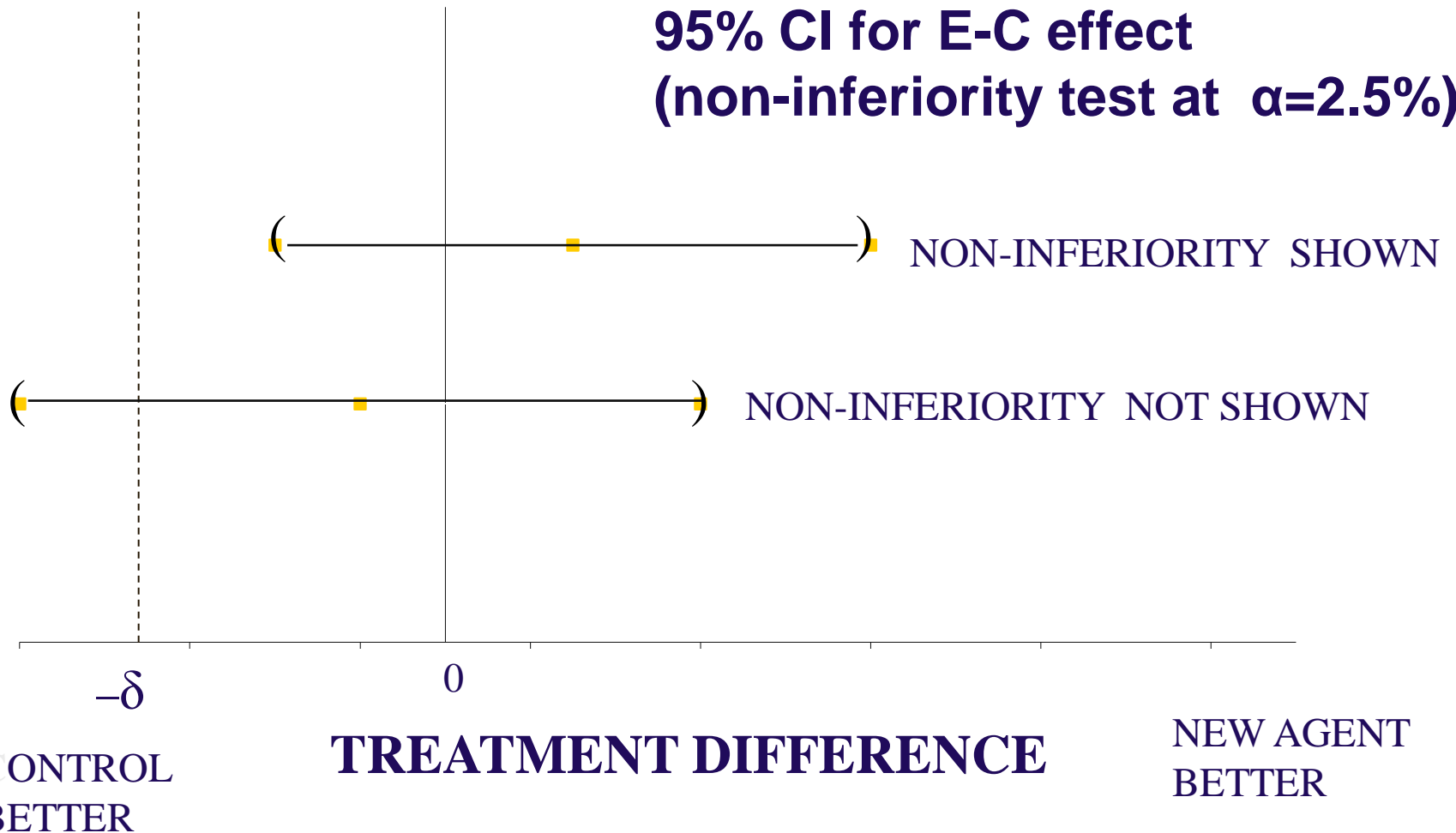
Nierównoliczne grupy: efektywność

- ♦ Względna efektywność vs. q
- ♦ $q=1/2 \rightarrow \text{rel.eff.} \approx 90\%$
- ♦ $q=1/3 \rightarrow \text{rel.eff.} \approx 75\%$
- ♦ Niewielka strata efektywności dla $q \geq 1/2$.



Nie-podrzędność

**95% CI for E-C effect
(non-inferiority test at $\alpha=2.5\%$)**



Próby kliniczne nie-podrzędności

- ◆ $H_0: \mu_E - \mu_C \leq -\delta$ $H_A: \mu_E - \mu_C > -\delta$
(podrzędność) (nie-podrzędność)

- δ to klinicznie istotna różnica

- ◆ W obliczeniach

$$H_0: \mu_E - \mu_C = -\delta \qquad H_A: \mu_E - \mu_C = \Delta > -\delta$$

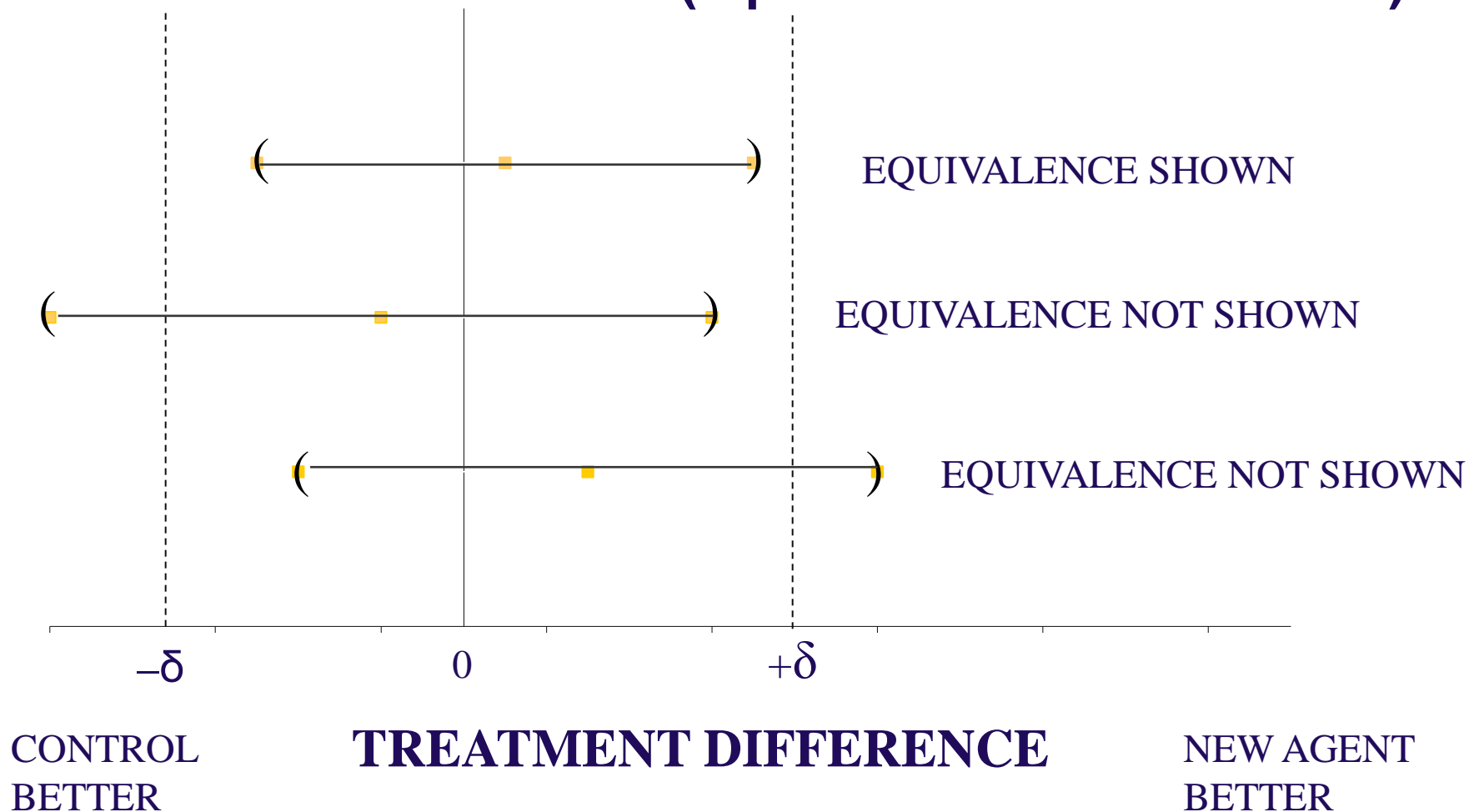
$$n = \frac{q+1}{q} \frac{(z_{1-\beta} + z_{1-\alpha})^2}{\left(\frac{\Delta + \delta}{\sigma}\right)^2}$$

Table V. Sample sizes (n_A) for one arm of an parallel group **non-inferiority study** with equal allocation ($r = 1$) for different standardized equivalence limits ($\delta = d/\sigma$) and true mean differences (as a percentage of δ) for 90 per cent power and a type I error rate of 2.5 per cent.

δ	Percentage mean difference										
	-25	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%	25%
0.05	5381	5839	6358	6949	7626	8407	9316	10 379	11 636	13 136	14 945
0.10	1346	1461	1590	1738	1908	2103	2330	2596	2910	3285	3737
0.15	599	650	708	773	849	935	1036	1155	1294	1461	1662
0.20	338	366	399	436	478	527	584	650	729	822	935
0.25	217	235	256	279	306	338	374	417	467	527	599
0.30	151	164	178	194	213	235	260	290	325	366	417
0.35	111	121	131	143	157	173	192	213	239	270	306
0.40	86	93	101	110	121	133	147	164	183	207	235
0.45	68	74	80	87	96	105	116	130	145	164	186
0.50	55	60	65	71	78	86	95	105	118	133	151
0.55	46	50	54	59	64	71	78	87	98	110	125
0.60	39	42	46	50	54	60	66	74	82	93	105
0.65	33	36	39	43	47	51	57	63	70	79	90
0.70	29	31	34	37	40	44	49	54	61	68	78
0.75	25	27	30	32	35	39	43	48	53	60	68
0.80	23	24	26	29	31	34	38	42	47	53	60
0.85	20	22	23	26	28	31	34	37	42	47	53
0.90	18	20	21	23	25	27	30	34	37	42	48
0.95	16	18	19	21	23	25	27	30	34	38	43
1.00	15	16	17	19	21	23	25	27	31	34	39
1.05	14	15	16	17	19	21	23	25	28	31	35
1.10	13	14	15	16	17	19	21	23	26	29	32
1.15	12	13	14	15	16	17	19	21	23	26	30
1.20	11	12	13	14	15	16	18	20	22	24	27
1.25	10	11	12	13	14	15	16	18	20	23	25
1.30	10	10	11	12	13	14	15	17	19	21	24
1.35	9	10	10	11	12	13	14	16	17	20	22
1.40	8	9	10	10	11	12	13	15	16	18	21
1.45	8	9	9	10	11	12	13	14	15	17	19
1.50	8	8	9	9	10	11	12	13	14	16	18

Równoważność

95% CI for E-C effect
(equivalence test at $\alpha=2.5\%$)



Próby kliniczne równoważności

- ◆ $H_0: |\mu_E - \mu_C| \geq \delta$ $H_A: |\mu_E - \mu_C| < \delta$
(nie-równoważność) (równoważność)
 - δ to klinicznie istotna różnica

- ◆ W obliczeniach

$$H_0: |\mu_E - \mu_C| = \delta \quad H_A: |\mu_E - \mu_C| = \Delta < \delta$$

$$1 - \beta = \Phi\left(\sqrt{\frac{(\Delta - \delta)^2 qn}{(q + 1)\sigma^2}} - Z_{1-\alpha}\right) + \Phi\left(\sqrt{\frac{(\Delta + \delta)^2 qn}{(q + 1)\sigma^2}} - Z_{1-\alpha}\right) - 1$$

Próby kliniczne równoważności

- ◆ Przybliżone obliczenia:

$$n = \frac{q+1}{q} \frac{(z_{1-\beta} + z_{1-\alpha})^2}{\left(\frac{\Delta - \delta}{\sigma}\right)^2}$$

- ◆ Gdy $\Delta = 0$

$$1 - \beta = 2\Phi\left(\sqrt{\frac{\delta^2 q n}{(q+1)\sigma^2}} - Z_{1-\alpha}\right) - 1 \rightarrow \sqrt{\frac{\delta^2 q n}{(q+1)\sigma^2}} - Z_{1-\alpha} = Z_{\beta/2}$$

- ◆ Stąd, dla $\Delta = 0$

$$n = \frac{q+1}{q} \frac{(z_{1-\beta/2} + z_{1-\alpha})^2}{\left(\frac{\Delta - \delta}{\sigma}\right)^2}$$

Table III. Sample sizes (n_A) for one arm of a parallel group equivalence study with equal allocation ($r=1$) for different standardized equivalence limits ($\delta=d/\sigma$) and true mean differences (as a percentage of δ) for 90 per cent power and a type I error rate of 2.5 per cent.

δ	Percentage mean difference				
	0%	10%	15%	20%	25%
0.05	10397	11042	11915	13218	14960
0.10	2600	2762	2980	3306	3741
0.15	1157	1228	1325	1470	1664
0.20	651	691	746	827	936
0.25	417	443	478	530	600
0.30	290	308	332	369	417
0.35	214	227	245	271	307
0.40	164	174	188	208	235
0.45	130	138	149	165	186
0.50	105	112	121	134	151
0.55	87	93	100	111	125
0.60	74	78	84	93	105
0.65	63	67	72	80	90
0.70	55	58	62	69	78
0.75	48	51	54	60	68
0.80	42	45	48	53	60
0.85	37	40	43	47	53
0.90	34	36	38	42	48
0.95	30	32	34	38	43
1.00	27	29	31	35	39
1.05	25	27	29	31	35
1.10	23	24	26	29	32
1.15	21	22	24	26	30
1.20	20	21	22	24	27
1.25	18	19	21	23	25
1.30	17	18	19	21	24
1.35	16	17	18	20	22
1.40	15	16	17	18	21
1.45	14	15	16	17	19
1.50	13	14	15	16	18

Więcej niż dwie porównywane grupy

- ◆ K grup, kryterium $\sim N(\mu_i, \sigma^2)$
- ◆ $H_0: \mu_1 = \dots = \mu_K$ $H_A: \mu_i \neq \mu_j$ dla pewnych i, j
- ◆ Przyjmijmy równą liczebność grup n
- ◆ Jednoczynnikowa ANOVA:
$$E(MSB) = \sigma^2 + \frac{n \sum_i (\mu_i - \bar{\mu})^2}{K - 1}$$
$$E(MSE) = \sigma^2$$
- ◆ Dla H_0 , $MSB / MSE \sim$ centralny $F(K-1, K(n-1))$

Uwaga: MSB = Mean Square Between, MSE = Mean Square Error

Więcej niż dwie porównywane grupy

- ◆ Dla H_A , $MSB / MSE \sim$ niecentralny $F(K-1, K(n-1), \theta^2)$, z parametrem nie-centralności

$$\theta^2 = \frac{n \sum_i (\mu_i - \bar{\mu})^2}{\sigma^2}$$

- ◆ Niech $K=3$, $n=25$, $\mu_1 = 2$, $\mu_2 = 4$, $\mu_3 = 6$, $\sigma^2=15$
- ◆ $\theta^2 = 25 \times 8 / 15 = 13.333$
- ◆ Dla $\alpha = 0.05$, wartość krytyczna dla $F(2,72,0)$ to 3.124
- ◆ $\beta = P(F(2,72,13.333) < 3.124) = 0.097$; moc = 0.903
 - ponieważ df zależą od n , procedura iteracyjna
 - sformułowanie H_A jest trudne z praktycznego punktu widzenia

Binarne kryteria oceny skuteczności leczenia

- ◆ Dwie grupy
- ◆ W eksperymentalnej, liczba odpowiedzi $\sim B(\pi_E, n)$
- ◆ W kontrolnej, liczba odpowiedzi $\sim B(\pi_C, n)$
- ◆ Niech $\Delta = \pi_E - \pi_C$
- ◆ $H_0: \Delta = 0$ $H_A: \Delta \neq 0$ (przyjmujemy > 0)
- ◆ Rozważamy test χ^2 dla dwóch odsetków

Kryteria binarne: liczebność próbki

- ◆ Fleiss, Tytun and Ury (1980):

$$n = \frac{\left[\sqrt{A} + \sqrt{A + 2(q+1)\Delta} \right]^2}{4q\Delta^2}$$

$$A = \left[z_{1-\alpha/2} \sqrt{(q+1)\bar{\pi}(1-\bar{\pi})} + z_{1-\beta} \sqrt{q\pi_E(1-\pi_E) + \pi_C(1-\pi_C)} \right]^2$$

$$\bar{\pi} = \frac{\pi_C + q\pi_E}{1+q} = \pi_C + \frac{q}{1+q} \Delta$$

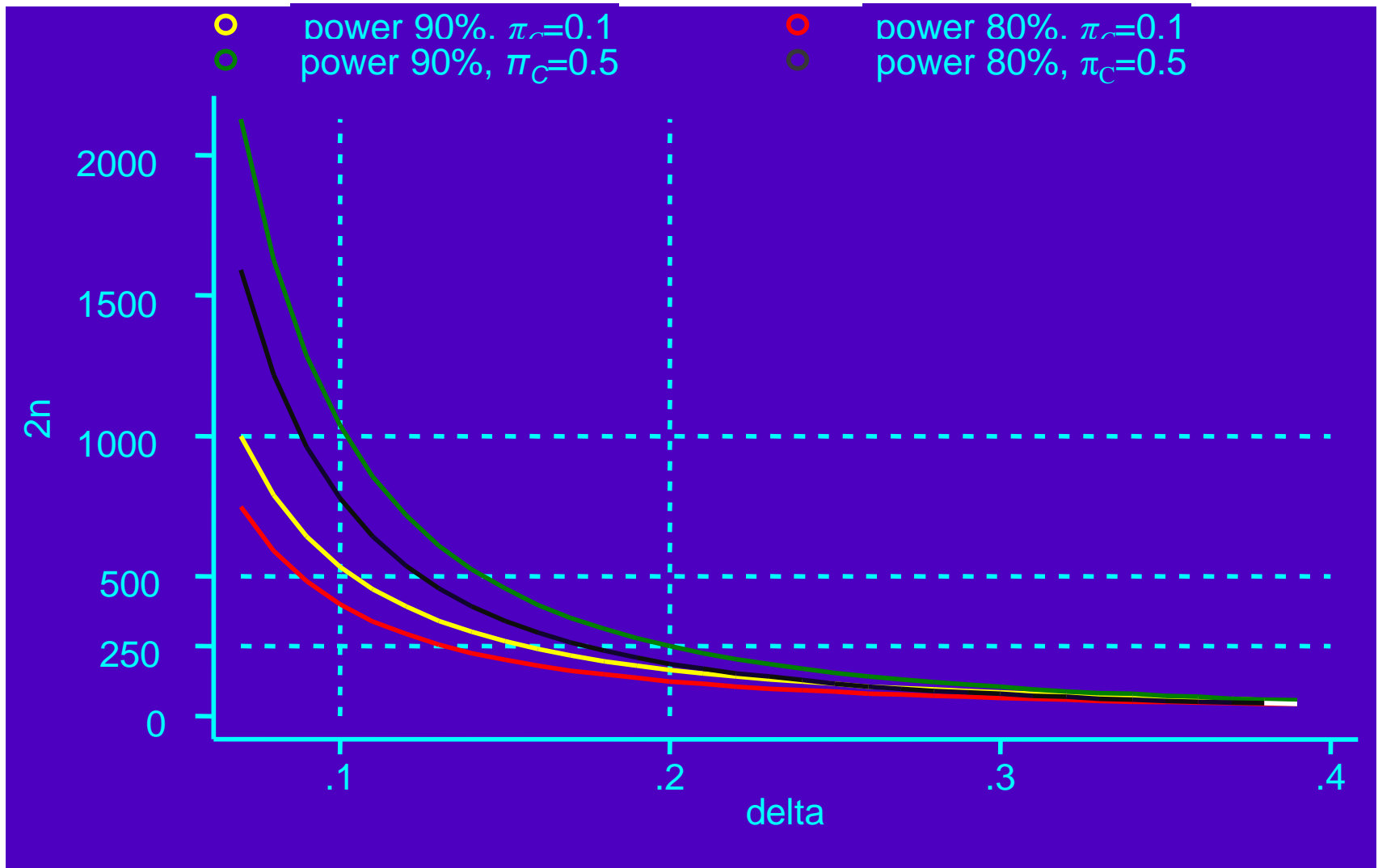
- ◆ n zależy od π_C ($\alpha = 0.05$, $\beta = 0.2$, $\Delta = 0.05$, $q = 1$):

$$\pi_C = 0.10 \rightarrow n = 580$$

$$\pi_C = 0.20 \rightarrow n = 901$$

$$\pi_C = 0.50 \rightarrow n = 1273$$

Binarne kryteria oceny skuteczności leczenia



Binarne kryteria oceny skuteczności leczenia: $K > 2$ grupy

- ◆ Rozważmy “najmniej korzystną konfigurację”
 - ustalamy minimalne i maksymalne p-stwo odpowiedzi
 - dla $K-2$ pozostałych grup p-stwo odpowiedzi przyjmujemy pomiędzy tymi dwoma wartościami
- ◆ $n = C(K, \alpha, \beta) \times (\text{liczebność próbki dla } K=2)$
 - $C(K, \alpha, \beta)$ z Tabeli 18.2 w artykule George’a
- ◆ Problem: bardzo konserwatywne podejście

Binarne kryteria oceny skuteczności leczenia: nie-podrzędność

♦ $H_0: \pi_E - \pi_C = -\delta$ $H_A: \pi_E - \pi_C = \Delta < -\delta$

$$n = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 \left(\pi_E(1 - \pi_E) + \pi_C(1 - \pi_C)\right)}{(\Delta + \delta)^2}$$

Czas do wystąpienia zdarzenia: równoliczne grupy

- ♦ Ważna liczba zaobserwowanych zdarzeń (nie liczba chorych)
- ♦ Niech Δ = iloraz hazardu (eksperymentalna/kontrolna)
 - stały w czasie
- ♦ $H_0: \Delta = 1$ $H_A: \Delta \neq 1$ (przyjmujemy < 1)
- ♦ Rozważamy dwustronny test logrank
- ♦ Wymagana CAŁKOWITA liczba zdarzeń (Freedman 1982):

$$d = \left(z_{1-\beta} + z_{1-\alpha/2} \right)^2 \left(\frac{1+\Delta}{1-\Delta} \right)^2 \approx \left(z_{1-\beta} + z_{1-\alpha/2} \right)^2 \frac{4}{(\ln \Delta)^2}$$

Czas do wystąpienia zdarzenia: nierównoliczne grupy

- ◆ Przyjmujemy względną liczebność 1: q
- ◆ Wymagana CAŁKOWITA liczba zdarzeń:

$$d = \left(z_{1-\beta} + z_{1-\alpha/2} \right)^2 \frac{(1+q)^2}{q(\ln \Delta)^2}$$

- ◆ Dla $q=1$, $(1+q)^2/q = 4$.
- ◆ Przyjęcie d jako docelowej liczebności próby oznacza konieczność obserwacji zdarzenia dla wszystkich chorych.

Czas do wystąpienia zdarzenia: liczba chorych

- ◆ Przyjmijmy okres rekrutacji T i okres obserwacji τ .
- ◆ Wymagana liczba chorych może być wyznaczona jako
 $d / P(\text{zdarzenie})$

gdzie

$$P(\text{zdarzenie}) = 1 - \frac{1}{6} [\bar{S}(\tau) + 4\bar{S}(\tau + 0.5T) + \bar{S}(\tau + T)]$$
$$\bar{S}(t) = \frac{S_E(t) + S_C(t)}{2}$$

Czas do wystąpienia zdarzenia: liczba chorych

- ♦ Zakładając rozkład wykładniczy

$$\bar{S}(t) = \frac{e^{-\lambda_E t} + e^{-\lambda_C t}}{2}$$

- ♦ λ_E i λ_C można uzyskać np. z założonej mediany czasu:

$$\lambda_E = \ln 2 / \text{median}_E$$

Czas do wystąpienia zdarzenia: jak długa rekrutacja?

- ◆ Przyjmijmy stały współczynnik intensywności rekrutacji γ .
- ◆ Oczekiwana liczba zdarzeń

$$\gamma T \times P(\text{zdarzenie})$$

- ◆ Szukamy takich τ oraz T , że

$$d - \gamma T \times P(\text{zdarzenie}) \approx 0$$

- zazwyczaj trzeba rozważyć kilka kombinacji wartości τ i T

Czas do wystąpienia zdarzenia: przykład

- ◆ Czas w miesiącach

$$S_C(12)=0.68, S_C(36)=0.42, \\ S_C(48)=0.25, S_C(60)=0.20$$

- ◆ Zakładamy model
PH: $S_E(t)=S_C(t)^\Delta$

- ◆ Niech $S_E(48)=0.40$:
 $\Delta=\ln(0.4)/\ln(0.25)=.6$

$$S_E(12)=0.76, S_E(36)=0.56, \\ S_E(48)=0.40, S_E(60)=0.35$$

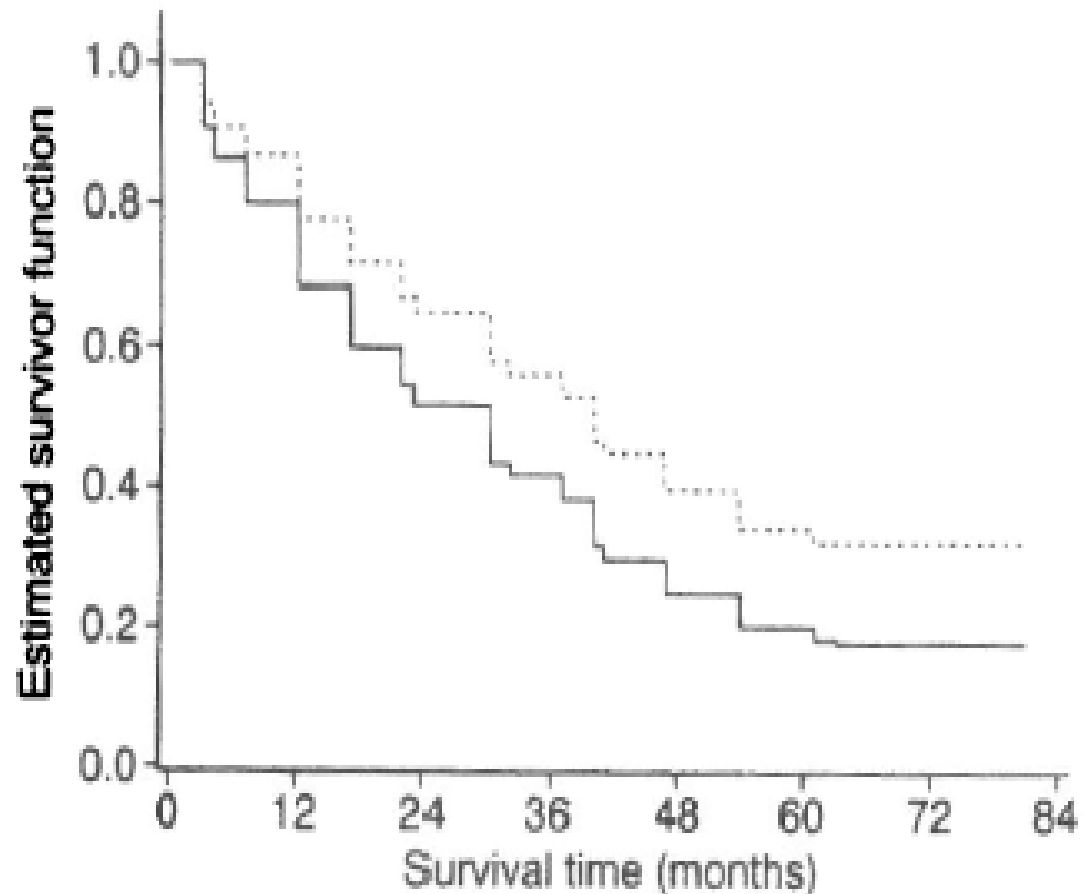


Figure 1 Estimated survivor functions for individuals on the standard treatment (—) and the new treatment (···)

Czas do wystąpienia zdarzenia: przykład

- ♦ Ustalając $\alpha=0.05$ (dwustronne) oraz $\beta=0.1$ otrzymujemy

$$d = (1.96 + 1.28)^2 \frac{4}{(\ln 0.66)^2} = 243$$

Czas do wystąpienia zdarzenia: przykład

- ◆ Niech okres rekrutacji $T=18$, a obserwacji $\tau=24$ mies.
- ◆ Otrzymujemy

$$\bar{S}(24) = \frac{S_E(24) + S_C(24)}{2} = \frac{0.52 + 0.64}{2} = 0.58$$

$$\bar{S}(24 + 9) = \frac{S_E(33) + S_C(33)}{2} = \frac{0.42 + 0.56}{2} = 0.49$$

$$\bar{S}(24 + 18) = \frac{S_E(42) + S_C(42)}{2} = \frac{0.30 + 0.44}{2} = 0.37$$

$$P(\text{zdarzenie}) = 1 - \frac{1}{6} [\bar{S}(24) + 4\bar{S}(33) + \bar{S}(42)] = 0.515$$

$$2n = \frac{243}{0.515} = 472$$

Czas do wystąpienia zdarzenia: przykład

- ◆ 472 chorych w 18 mies. daje $\gamma \approx 26$ chorych/mies
- ◆ Załóżmy, że tylko 15 chorych/mies będzie dostępnych
- ◆ Ustalamy $T=30$ oraz $\tau=24$ mies.
 - co daje całkowitą liczbę chorych równą 450

$$P(\text{zdarzenie}) = 1 - \frac{1}{6} [\bar{S}(24) + 4\bar{S}(39) + \bar{S}(54)] = 0.552$$
$$2n = \frac{243}{0.552} \approx 440$$

Czas do wystąpienia zdarzenia: $K > 2$ grupy

- ◆ Jak dla kryteriów binarnych
- ◆ Rozważamy “najmniej korzystną konfigurację”
- ◆ $n = C(K, \alpha, \beta) \times (\text{liczebność próbki dla } K=2)$
 - $C(K, \alpha, \beta)$ z Tabeli 18.2 w artykule George’a
- ◆ Problem: bardzo konserwatywne podejście

Liczebność próbki a nieistotny statystycznie wynik testu

- ◆ Niestotny statystycznie wynik testu nie dowodzi prawdziwości hipotezy zerowej.
- ◆ Jeśli liczebność próbki była zbyt mała, test (próba kliniczna) mógł mieć zbyt małą moc dla “rozsądnych” hipotez alternatywnych.
- ◆ W takim przypadku „negatywny” wynik testu jest obciążony dużym p-stwem błędu II rodzaju.

Przykład: 5-FU vs. 5-FU+L

◆ Di Constanzo *et al.*, *Proc ASCO* (1989)

Grupa	Odpowiedzi
5-FU	14/90 (15.6%)
5-FU+L	12/91 (13.2%)

◆ $\pi_E = \pi_C$?

◆ Oszacowana różnica ryzyka: $\pi_E - \pi_C = 2.4\%$

◆ 95% CI dla $\pi_E - \pi_C$:

[- 8.0%, 12.8%].

◆ Zawiera 0 \rightarrow nie możemy odrzucić hipotezy, że $\pi_E = \pi_C$ na poziomie istotności 5%.

Przykład: 5-FU vs. 5-FU+L

- ◆ By osiągnąć moc 90% dla różnicy 10% na korzyść 5-FU+L (dla $\pi_C=10\%$), próba powinna liczyć $2n > 500$ chorych.
- ◆ W GOIRC chorych było 181... $\beta = 47.5\%$, moc= 52.5%.
- ◆ Prawdziwa różnica ryzyka może być co najmniej 10% (95% CI dla $\pi_E - \pi_C : [- 8.0\%, 12.8\%]$).

Wybór parametrów próby

- ◆ Zazwyczaj, kontrola α jest najważniejsza
 - błędny wniosek o korzystnym działaniu leczenia
- ◆ σ^2 uzyskane z np. poprzednich prób klinicznych
 - czasem problematyczne; próby pilotażowe lub układy dwupoziomowe
- ◆ Δ
 - the Smallest Clinically Meaningful Difference
 - the Anticipated Treatment Effect

Software

- ◆ nQuery
- ◆ EaSt
- ◆ STATA
- ◆ StatXact
- ◆ *Power & Precision*
- ◆ *Many web sites...*