




Rekrutacja chorych

- Dobór chorych
 - rekrutacja (accrual)
 - kryteria włączania/wyłączania (inclusion/exclusion)
 - próby pragmatyczne vs. objaśniające
- „Populacje” chorych
 - zgodna z intencją leczenia (intention-to-treat)
 - zgodna z protokołem (per protocol)
 - bezpieczeństwa (safety)
- Podgrupy chorych
 - analizy podgrup
 - interakcje
- Czynniki prognostyczne i predykcyjne (prognostic & predictive factors)

Rekrutacja chorych

- Chorzy są włączani do próby jeśli spełniają kryteria włączenia/wyłączenia (inclusion/exclusion)
- Kryteria zależą od celów próby
- Mają wpływ na
 - tempo rekrutacji i czas trwania próby
 - moc statystyczną próby
 - możliwość uogólnienia wyników próby

Tempo rekrutacji

-  Często mocno przeszacowane
-  Długie próby są trudne do ukończenia
-  Zmiany w kryteriach doboru w trakcie trwania próby

Kryteria doboru chorych (1)

- Związane z
 - stopniem zaawansowania choroby
 - przeciwwskazaniami
 - oczekiwanej gotowości do stosowania się do wymogów leczenia („okresy rozbiegowe”)

Kryteria doboru chorych (2)

- Bezpośrednie (dorośli, wiek poniżej 75 lat, z udokumentowaną chorobą w określonym stopniu zaawansowania, nie w ciąży...)
 - chorzy, u których leczenie prawdopodobnie nie zadziała lub może zaszkodzić, powinni zostać wyłączeni
- Pośrednie
 - częste wizyty kontrolne → miejsce zamieszkania blisko ośrodka klinicznego?
 - świadoma zgoda → lepiej wykształceni?

Kryteria doboru chorych (3)

- Szczegółowe redukują zmienność (zwiększają moc), ale
 - utrudniają rekrutację
 - źródła zmienności często nie są znane
- Szerokie umożliwiają większą liczebność próbki (zwiększają moc) i skracają czas trwania
- Generalnie, im szersze, tym lepiej

Próby pragmatyczne vs. objaśniające (1)

- Objaśniające (explanatory): potwierdzenie skuteczności (efektu biologicznego) leczenia
⇒ wąskie kryteria doboru
- Pragmatyczne: potwierdzenie efektywności leczenia
⇒ szerokie kryteria doboru chorych

Uogólnianie wyników: uczestnicy próby stanowią wyselekcjonowaną grupę

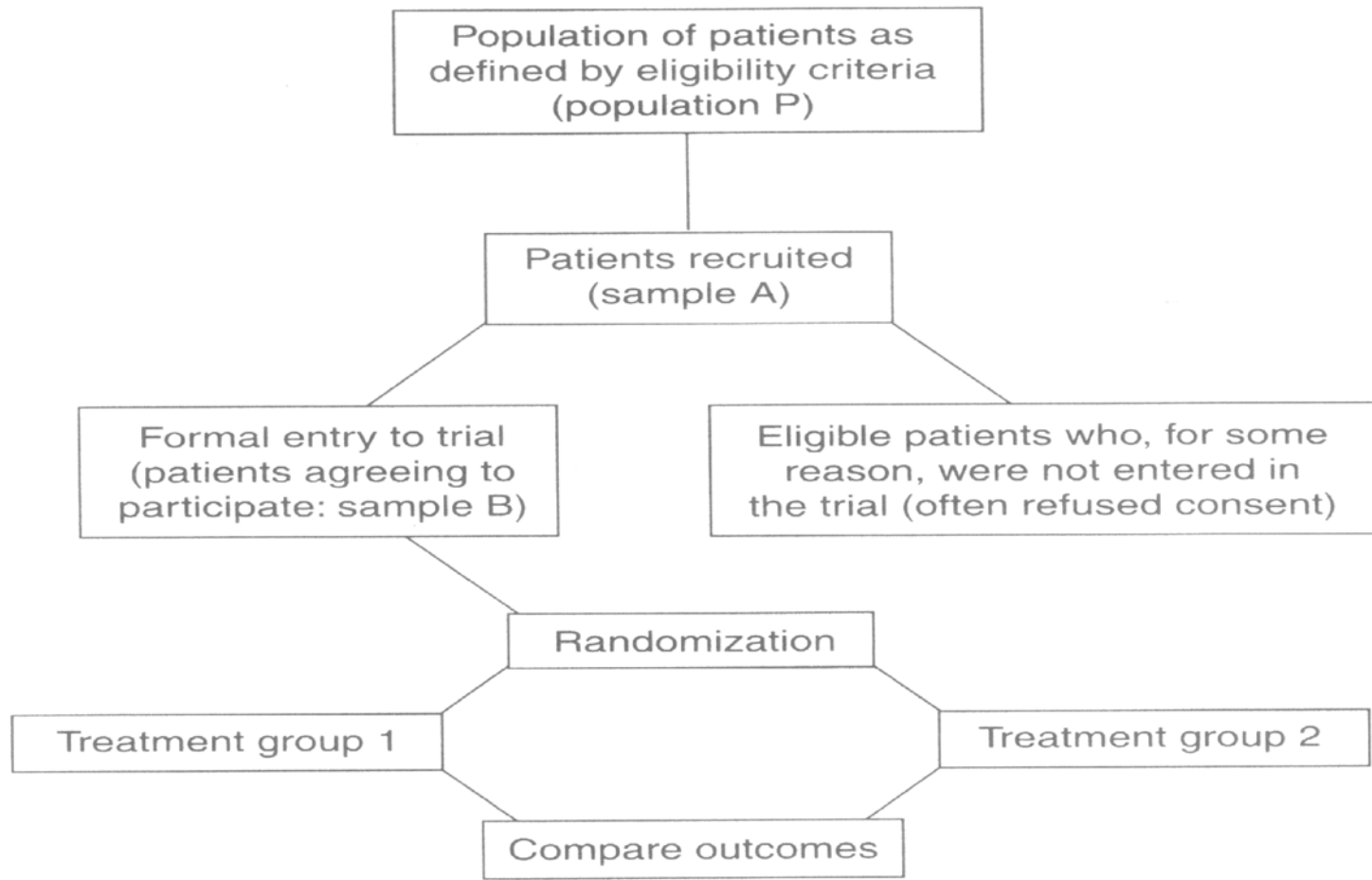


Figure 1.1

Uogólnianie wyników: uczestnicy próby stanowią wyselekcjonowaną grupę

- Chorzy włączeni do próby nie są ani losową, ani reprezentatywną próbką z docelowej populacji chorych.
- Randomizacja zapewnia wewnętrzną trafność (brak obciążenia porównania wyników leczenia), ale nie zapewnia trafności zewnętrznej (możliwości uogólnienia rezultatów dla przyszłych chorych). Ta druga jest oceniana na podstawie charakterystyki próby i chorych, którzy z próbie wzięli udział.

„Populacje” chorych

- *Intencja leczenia (Intention-to-treat, ITT)* : wszyscy zrandomizowani chorzy, analiza zgodnie z wynikiem randomizacji
- *Full-analysis set* : ICH E9 – zbiór tak bliski ITT jak to możliwe
- *Per-protocol* : podzbiór „full-analysis”, leczenie zgodne z protokołem
- *As-treated* : chorzy analizowani zgodnie z otrzymanym (a nie randomizowanym) leczeniem
- *Safety* : as-treated + minimalna dawka leku

Intention-To-Treat (ITT)

- Konserwatywna w próbach nadrzędności
 - poważne odstępstwa od protokołu (np. zamiana leczenia lub strata z obserwacji) generalnie zmniejszają różnicę w skuteczności
- Liberalna w próbach nie-podrzędności
- W idealnej sytuacji, wszyscy randomizowani chorzy powinni być leczeni i obserwowani zgodnie z protokołem
 - w praktyce, na ogół są z tym problemy.

Full-Analysis Set (ICH E9)

- Możliwe wykluczenia z ITT
 - pogwałcenie kryteriów doboru
 - nieotrzymanie leczenia
 - brak jakichkolwiek danych po randomizacji
- Zawsze muszą być uzasadniane !
- Obciążenie, które może być spowodowane wykluczeniami, musi być ocenione w analizie (sensitivity analyses)

Full-Analysis Set: Eligibility Violations (ICH E9)

- Wykluczenia możliwe jeśli:
 - kryterium doboru było ocenione przed randomizacją;
 - ocena była całkowicie obiektywna;
 - wszyscy uczestnicy próby byli oceniani tak samo pod kątem odstępstw od kryteriów doboru (problem z próbami otwartymi, a nawet z podwójnie ślepyimi, jeśli zaślepienie zostało usunięte przed oceną odstępstw);
 - wszystkie przypadki odstępstw dla danego kryterium są wykluczone

Full-Analysis Set: No Trial Medication (ICH E9)

- Wykluczanie chorych, którzy nie otrzymali leczenia
 - decyzja o rozpoczęciu leczenia nie mogła być spowodowana znajomością przydzielonego leczenia („selection bias”)

Full-Analysis Set: Missing Data (ICH E9)

- Wykluczanie chorych bez danych po randomizacji, lub straconych z obserwacji (lost to follow-up)
- Należy podejmować próby uzyskania chociażby częściowej informacji
 - metody powinny być opisane w protokole
- Metody uzupełniania (imputation) danych mogą być użyte
 - opisane i uzasadnione w protokole (założenia !!)
 - w celu wykazania „odporności” wyników próby
- Najgorszy możliwy scenariusz („worst case scenario”)
 - na ogół bezużyteczny


Per-Protocol

- Podzbiór „full analysis”
- Chorzy leczeni zgodnie z protokołem, np.:
 - otrzymali wymaganą minimalną dawkę leku;
 - uzyskano dla nich pomiary dla wymaganego kryterium oceny skuteczności leczenia;
 - brak poważnych odstępstw od protokołu.
- Może stwarzać największą szansę na ujawnienie się efektu leczenia.
- Wyniki analizy mogą być mocno obciążone, jeśli stosowanie się do protokołu było związane z leczeniem i jego wynikami.

Analiza „zgodna z protokołem”

Niebezpieczeństwa


	Anturan continued	Anturan withdrawn
Nr of patients	563	220
Nr of deaths	44	20
Mortality (/100 patient-years)	5.4%	9.0%



Better outcome among compliers
⇒ is compliance to treatment important?

Analiza „zgodna z protokołem”

	Anturan continued	Anturan withdrawn	Placebo continued	Placebo withdrawn
Nr of patients	563	220	580	195
Nr of deaths	44	20	62	23
Mortality (/100 pt-yrs)	5.4%	9.0%	7.6%	11.7%



Difference is also seen on Placebo
⇒ this suggests selection bias

As-Treated

- Chorzy analizowani zgodnie *otrzymanym leczeniem* (jeśli otrzymali przynajmniej jedną dawkę)
- Osoby, które nie biorą leku zgodnie z protokołem prawie na pewno nie są losową próbką z całej populacji chorych
- Raczej dla potrzeb analizy bezpieczeństwa leczenia

ITT/Per-Protocol/As-Treated

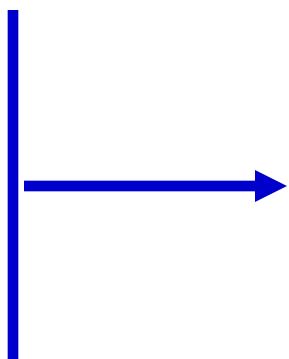
- W próbach nadrzędności, analiza ITT jest prawie zawsze najbardziej wskazana statystycznie (brak obciążenia) oraz klinicznie (odpowiada praktyce)
- Inne podpopulacje mogą być używane dla oceny „odporności” wyników próby

Próby pragmatyczne vs. objaśniające (2)

<u>Type of trial</u>	<u>Explanatory</u>	<u>Pragmatic</u>
Use	Regulatory	Public health
Patient selection	Strict	Broad
Experimental conditions	Well-defined	As in practice
Control group	Untreated (placebo)	Current standard
Data collected	Many	Few
Main analysis	Per protocol	Intention to treat

Podgrupy chorych

Pytanie Czy leczenie przynosi korzyść *generalnie* ?
statystyczne: *(a jeśli tak, to jak dużą?)*



Czy leczenie przynosi korzyść *jakiejś grupie* ?
(a jeśli tak, to jak dużą?)

Pytanie Czy leczenie przynosi korzyść *danemu* choremu ?
medyczne: *(a jeśli tak, to jak dużą?)*

Strategie szukania efektów dla podgrup

Jeśli uważa się, że leczenie może z większym p-stwem przynosić korzyść jakiejś grupie chorych, możemy

1. odpowiednio ograniczyć kryteria doboru chorych w próbie, lub
2. użyć szerokich kryteriów i szukać efektów leczenia w podgrupach

Próba ograniczona do podgrupy chorych

1. Zwiększenie mocy (potencjalne)
2. Szczegółowe wskazanie dla leczenia chorych w przyszłości
3. Mniejsza ogólność wyników
4. Brak informacji o efektach leczenia w innych podgrupach

Próba z analizą podgrup

1. Niebezpieczeństwo wyników fałszywie dodatnich
2. Mniejsza liczba chorych w podgrupach
3. Problem wielokrotnych porównań

Niebezpieczeństwa związane z analizą podgrup

Dwie próby podjęte w celu oceny nowego leczenia
zwyrodnienia plamki żółtej związanego z wiekiem

Każda próba z 500 chorymi, jedna w USA, druga w Europie

W obu próbach efekt leczenia był istotny statystycznie
($P < 0.01$)

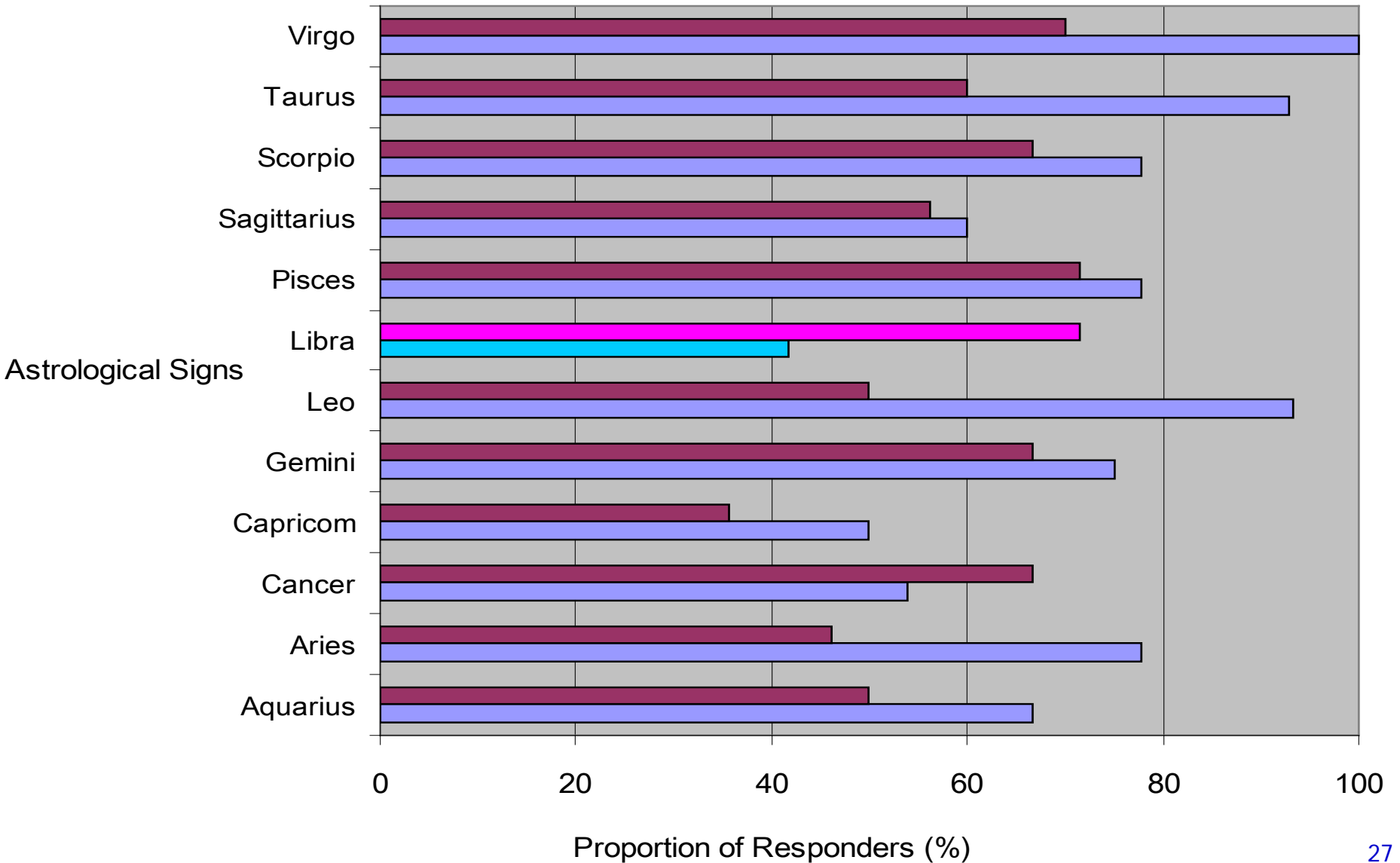
Analiza wyników próby europejskiej według znaku
astrologicznego wskazała na poważny, **szkodliwy** efekt
leczenia chorych spod znaku Wagi

Próba w USA potwierdziła ten wynik !

(Czy zaniechał(a)byś leczenia chorych spod znaku Wagi?!)

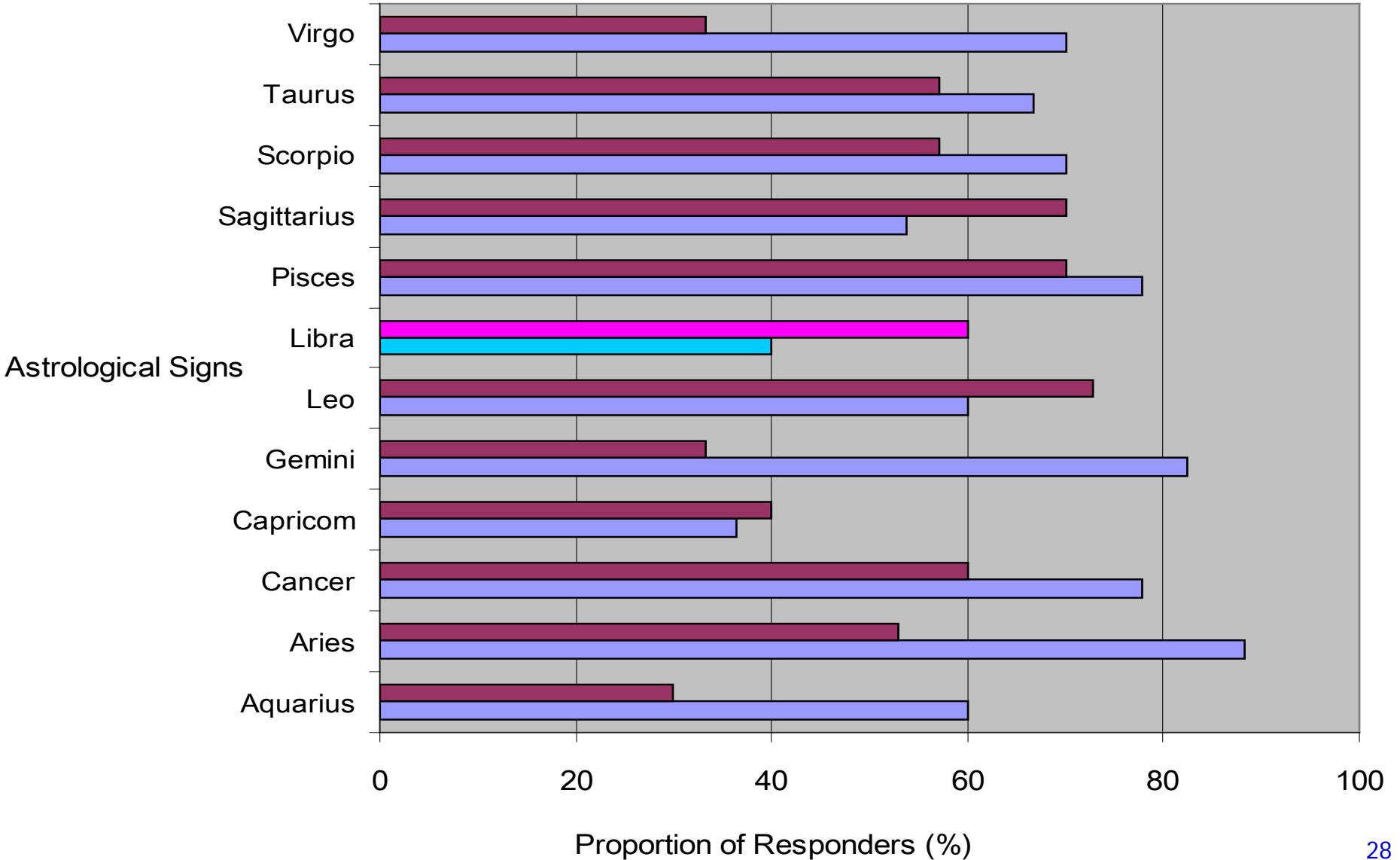
Proportion of Responders by Astrological Signs - European Trial

0.3 mg Sham



Proportion of Responders by Astrological Signs - US Trial

0.3 mg Sham



Testowanie efektów w podgrupach

- W większości przypadków efekty są przeszacowane – powinny być traktowane jako hipotezy
- Wiarygodność większa jeśli analiza była ograniczona do głównego kryterium oceny skuteczności w jedynie kilku zdefiniowanych z góry podgrupach
- Ostrożność zalecana w sytuacji, gdy efekt leczenia w całej próbie nie jest istotny statystycznie
- Poprawka poziomów istotności testów na wielokrotne testowanie
- Analiza powinna opierać się na testach interakcji

Analiza podgrup

- Cockburn *et al.* (1980): suplement witaminy D/placebo w prewencji hipokalcemii noworodków

Table 8.1. *Data from neonatal hypocalcaemia trial: all calcium levels in mmol/l*

	Breast-fed		Bottle-fed	
	Supplement	Placebo	Supplement	Placebo
Treatment mean	2.445	2.408	2.300	2.195
Number of babies <i>n</i>	64	102	169	285
SE	0.0365	0.0311	0.0211	0.0189
Treatment effect	0.037		0.105	
SE	0.0480		0.0283	
<i>P</i> -value	0.44		0.0002	

$p = 0.44$ vs. $0.0002 \rightarrow$ efekt leczenia zależy od sposobu karmienia? (czy to prawdopodobne i statystycznie sensowne?)

Test interakcji

- Statystyka testowa zdefiniowana jako

$$X^2 = \sum_{i=1}^S (\tau_i - \tau)^2 / s_i^2$$

gdzie τ_i jest miarą efektu leczenia dla i-tej podgrupy, s_i jest błędem standardowym τ_i , a τ jest miarą ogólnego efektu leczenia.

- X^2 ma rozkład χ^2 z $(S - 1)$ stopniami swobody.
- Dla dwóch podgrup, X^2 może być równoważnie wyrażone jako

$$X^2 = (\tau_1 - \tau_2)^2 / (s_1^2 + s_2^2)$$

Test interakcji

Table 8.1. *Data from neonatal hypocalcaemia trial: all calcium levels in mmol/l*

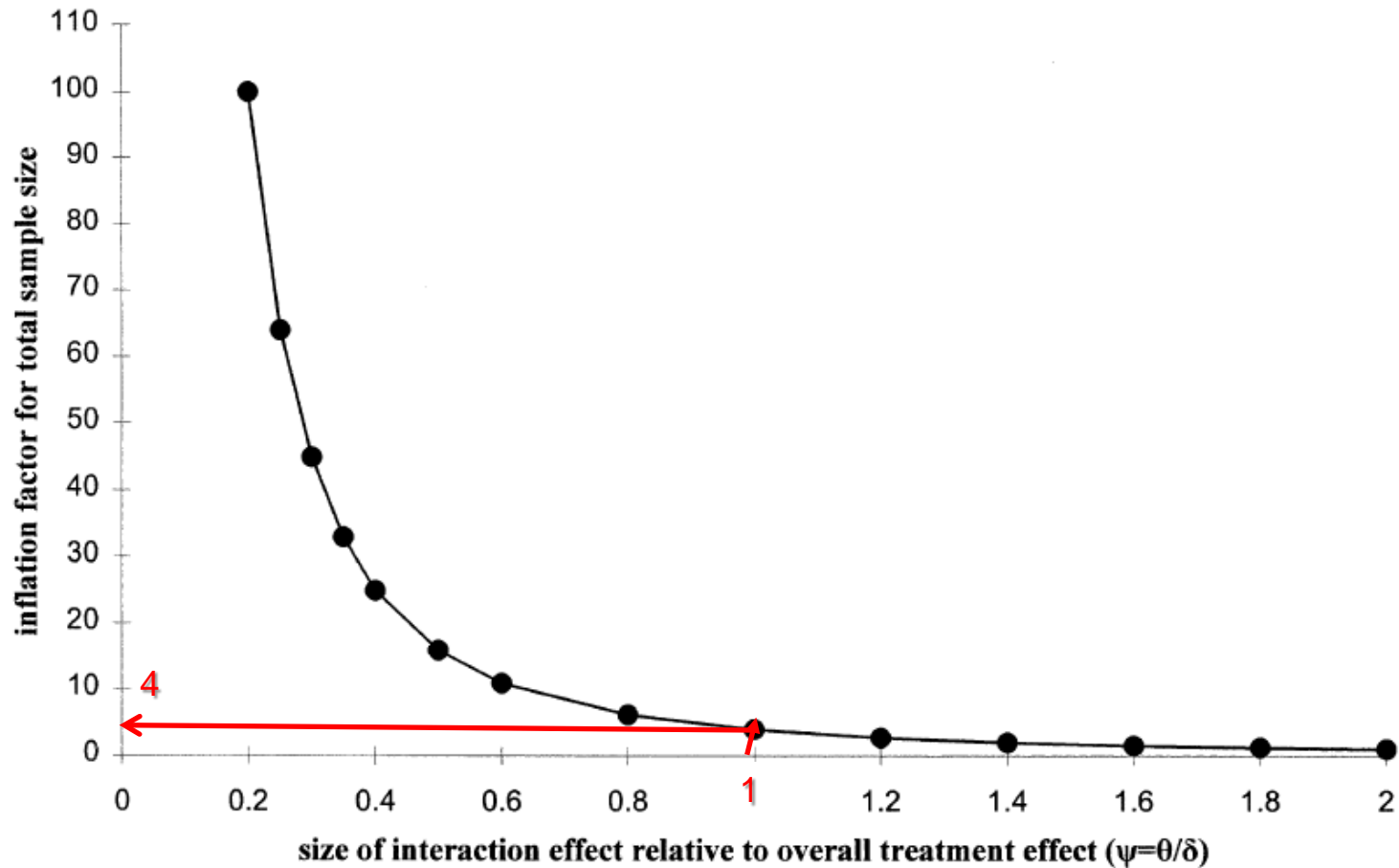
	Breast-fed		Bottle-fed	
	Supplement	Placebo	Supplement	Placebo
Treatment mean	2.445	2.408	2.300	2.195
Number of babies <i>n</i>	64	102	169	285
SE	0.0365	0.0311	0.0211	0.0189
Treatment effect	0.037		0.105	
SE	0.0480		0.0283	
<i>P</i> -value	0.44		0.0002	

- $Z = (.037 - .105)^2 / (.048^2 + .0283^2) = 1.49$
- $p = P(\chi_1^2 > 1.49) = 0.22$; brak dowodów na interakcję
- $p = 0.44$ może wynikać z małej liczebności grupy



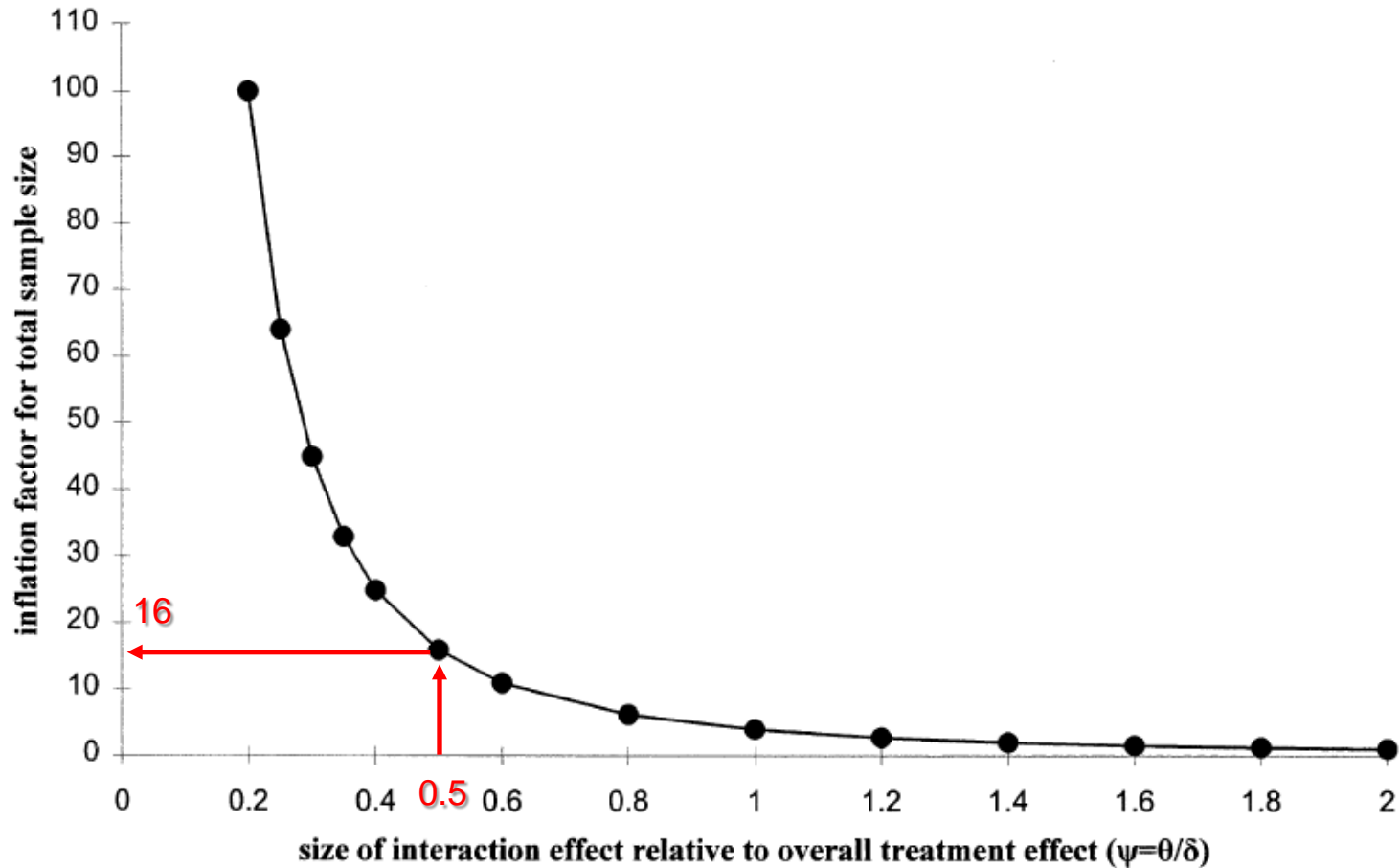
Moc testu interakcji jest mniejsza niż testu ogólnego efektu leczenia

Test interakcji ma ograniczoną moc



Inflation factor required to increase the sample size so that the interaction test has the same power as the original sample size had for the overall treatment effect.

Test interakcji ma ograniczoną moc



Inflation factor required to increase the sample size so that the interaction test has the same power as the original sample size had for the overall treatment effect.

Jakościowe/ilościowe interakcje

- **I**lościowa interakcja oznacza różnicę w oszacowanych wartościach miary efektu leczenia, ale bez zmiany znaku.
 - oczekiwana (do pewnego stopnia)
 - może wynikać ze skali pomiarowej
- **J**akościowa interakcja oznacza różnicę w oszacowanych wartościach miary efektu leczenia ze zmianą znaku, czyli przeciwstawne efekty.
 - nieczęsta
 - nie może być wynikiem skali
 - generalnie ważniejsza klinicznie (jeśli prawdziwa)

Jakościowe/ilościowe interakcje

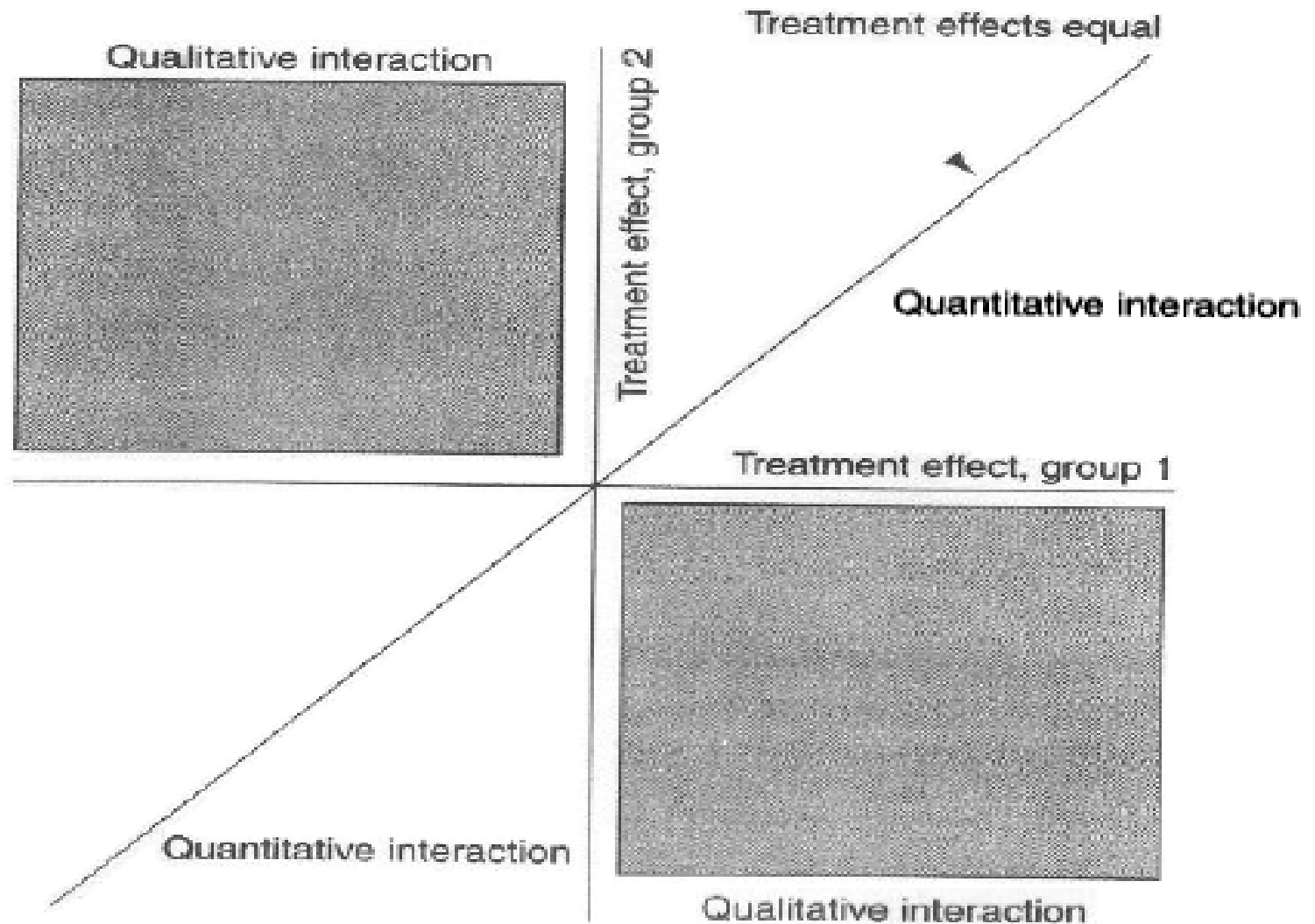


Figure 8.1

Jakościowa interakcja: przykład

- Gail and Simon (1985): PF vs. PF+tamoxifen for breast cancer.

Table 8.3. *Results from the National Surgical Adjuvant Breast and Bowel Project trial*

	Age < 50 PR < 10		Age ≥ 50 PR < 10		Age < 50 PR ≥ 10		Age ≥ 50 PR ≥ 10	
	PF	PFT	PF	PFT	PF	PFT	PF	PFT
Proportion disease free at 3 years	0.599	0.436	0.526	0.639	0.651	0.698	0.639	0.790
SE	0.0542	0.0572	0.0510	0.0463	0.0431	0.0438	0.0386	0.0387
D	0.163		-0.114		-0.047		-0.151	
s	0.0788		0.0689		0.0614		0.0547	
D^2/s^2	4.28		2.72		0.59		7.58	

Jakościowa interakcja: przykład

Prosty test interakcji bierze pod uwagę wszystkie hipotezy alternatywne

Test jakościowej interakcji bierze pod uwagę ograniczoną hipotezę alternatywną:

$$H_A: \sim [\{ \theta_i \geq 0 \ \forall i \} \cup \{ \theta_i \leq 0 \ \forall i \}]$$

W przykładzie, $p = 0.01$

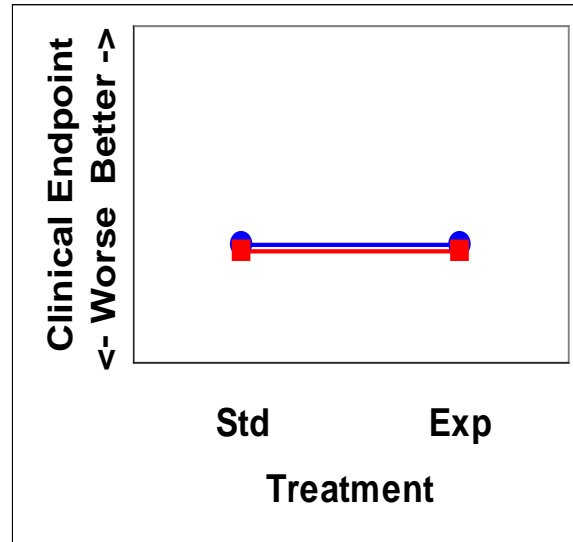
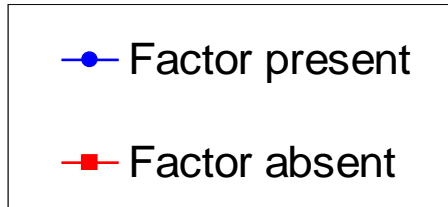
\Rightarrow Mocna wskazówka, że efekt leczenia zależy od wieku i stanu receptorów progesteronu

Zasady interpretacji wyników analizy podgrup

Podstawowe zasady interpretacji:

1. Podgrupy powinny być zdefiniowane a priori by zredukować ryzyko wyników fałszywie dodatnich
2. Podgrupy powinny być uzasadnione biologicznie i klinicznie
3. Poziom istotności powinien być właściwie skorygowany
4. Analizy powinny dawać zgodne wyniki dla różnych prób klinicznych

Definicje podgrup



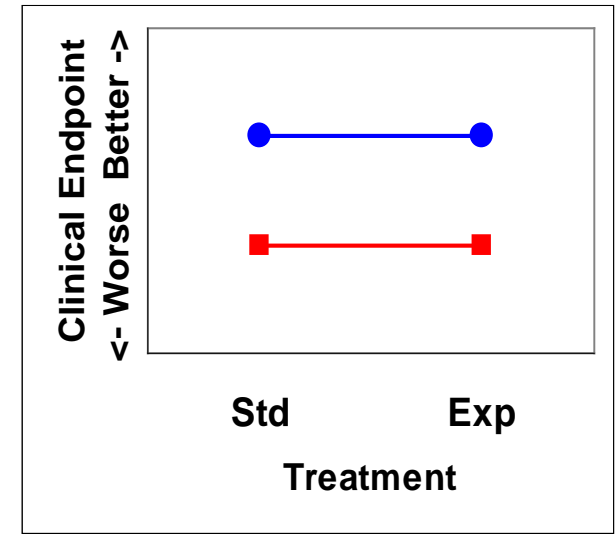
Wyniki te same w grupie z czynnikiem i bez, niezależnie od leczenia. Czynniki nie jest ani prognostyczny, ani predykcyjny.

Czynnik prognostyczny

● Factor present

■ Factor absent

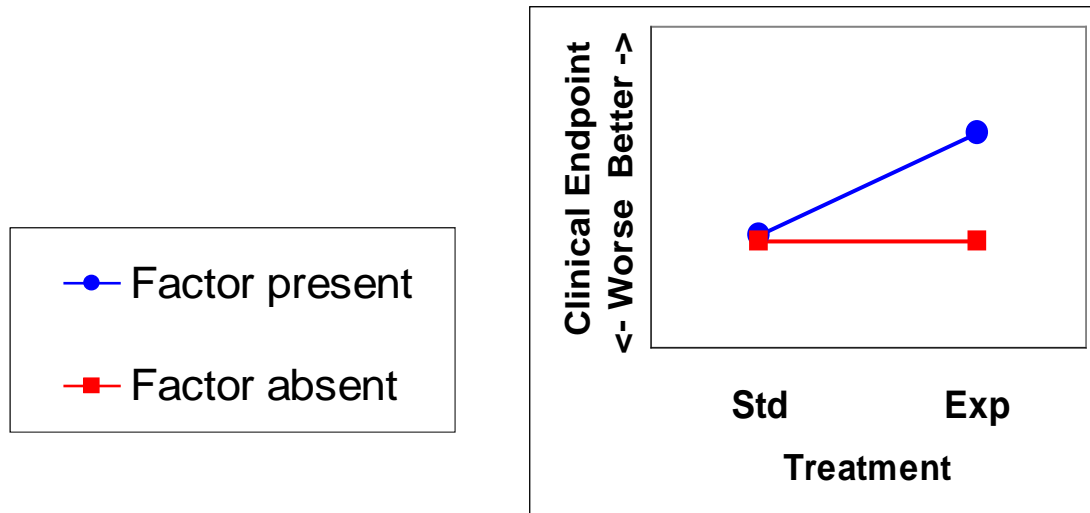
Wyniki lepsze w grupie z czynnikiem, niezależnie od leczenia. Czynniki jest prognostyczny.



Przykład dla wczesnego raka piersi:
zajęcie węzłów chłonnych

Czynnik predykcyjny

Przykład dla wczesnego raka piersi:
stan receptorów estrogenu lub progesteronu i
skuteczność terapii hormonalnej



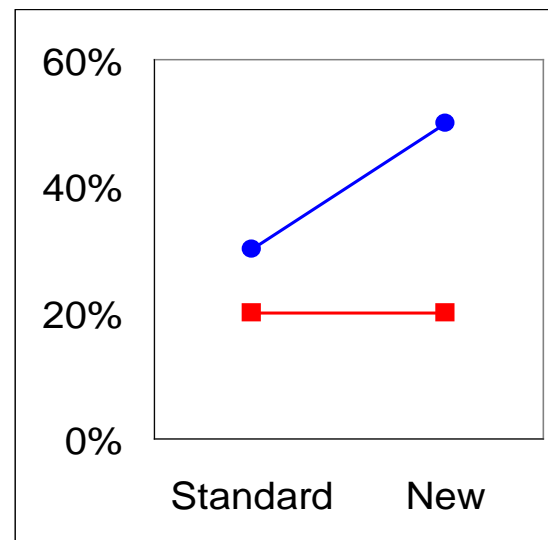
Wyniki lepsze w leczonej grupie z czynnikiem. Czynniki jest predykcyjny dla nowej metody leczenia.

Czynnik prognostyczny i predykcyjny

Przykład dla wczesnego raka piersi:
Podwyższona ekspresja genu HER2-neu
i skuteczność trastuzumabu

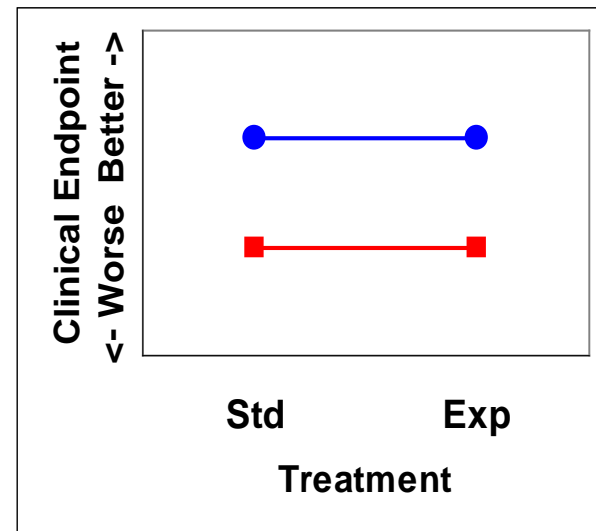
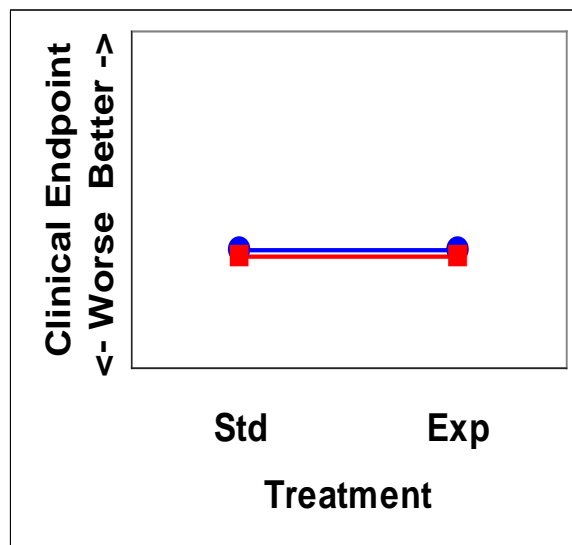
—●— Factor present
—■— Factor absent

Wyniki średnio
lepsze w grupie z
czynnikiem i większy
efekt dla nowej
metody leczenia.
Czynnik jest
prognostyczny i
predykcyjny.

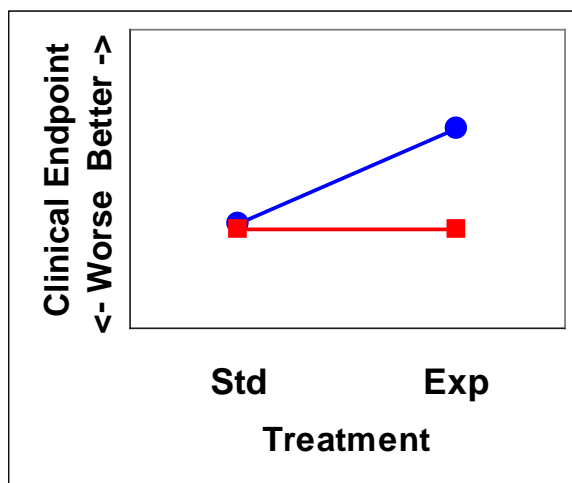


Czynniki prognostyczne i predykcyjne

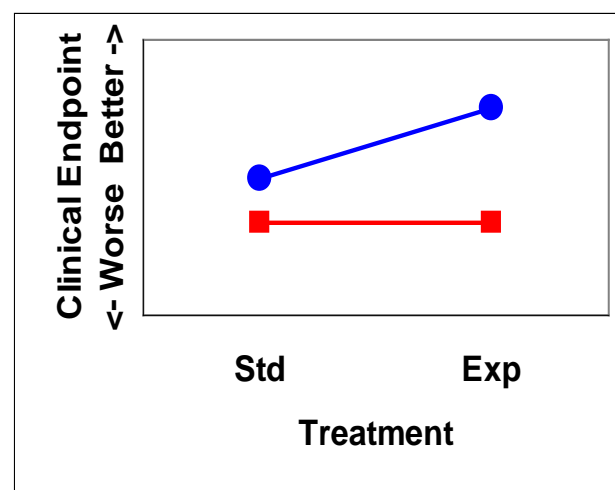
Prognostyczny



Predykcyjny



Prognostyczny i predykcyjny



—●— Factor present

—■— Factor absent

A Predictive Marker in Non-small Cell Lung Cancer

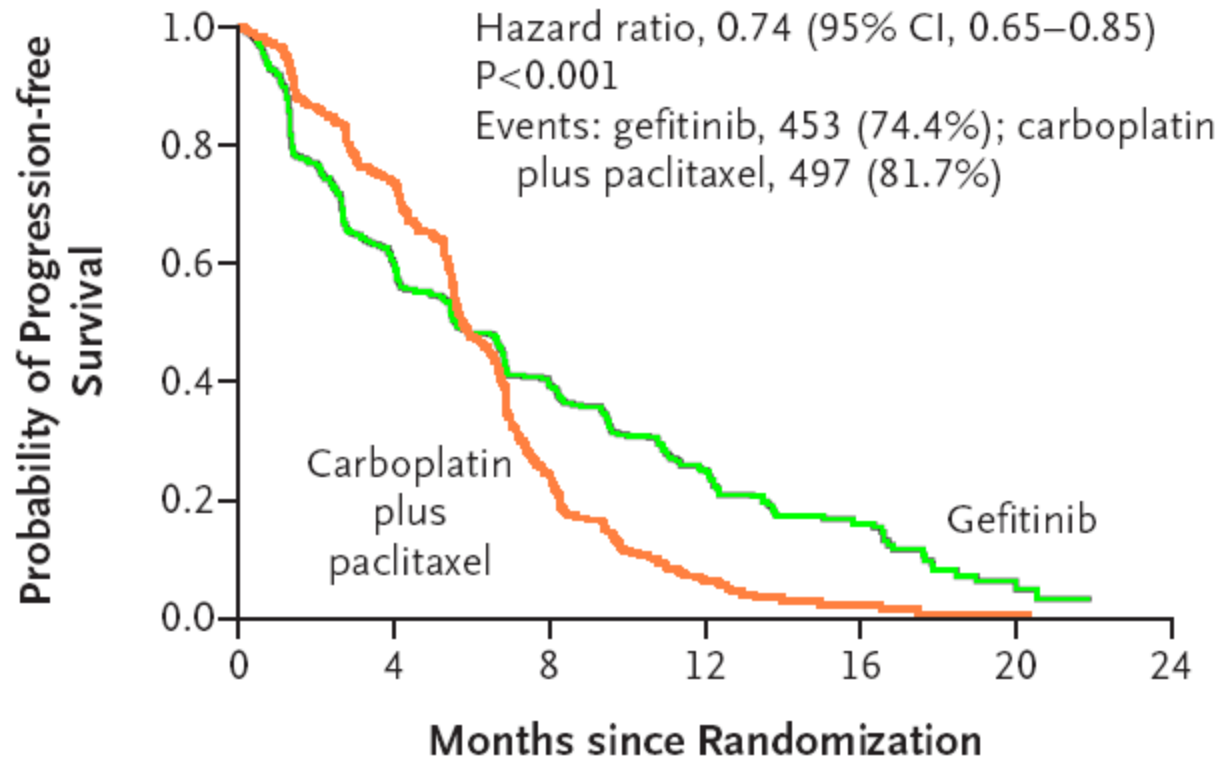
The NEW ENGLAND
JOURNAL *of* MEDICINE

Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma

Tony S. Mok, M.D., Yi-Long Wu, M.D., F.A.C.S., Sumitra Thongprasert, M.D., Chih-Hsin Yang, M.D., Ph.D.,
Da-Tong Chu, M.D., Nagahiro Saijo, M.D., Ph.D., Patrapim Sunpaweravong, M.D., Baohui Han, M.D.,
Benjamin Margono, M.D., Ph.D., F.C.C.P., Yukito Ichinose, M.D., Yutaka Nishiwaki, M.D., Ph.D.,
Yuichiro Ohe, M.D., Ph.D., Jin-Ji Yang, M.D., Busyamas Chewaskulyong, M.D., Haiyi Jiang, M.D.,
Emma L. Duffield, M.Sc., Claire L. Watkins, M.Sc., Alison A. Armour, F.R.C.R., and Masahiro Fukuoka, M.D., Ph.D.

Gefitinib vs. chemotherapy w NSCLC

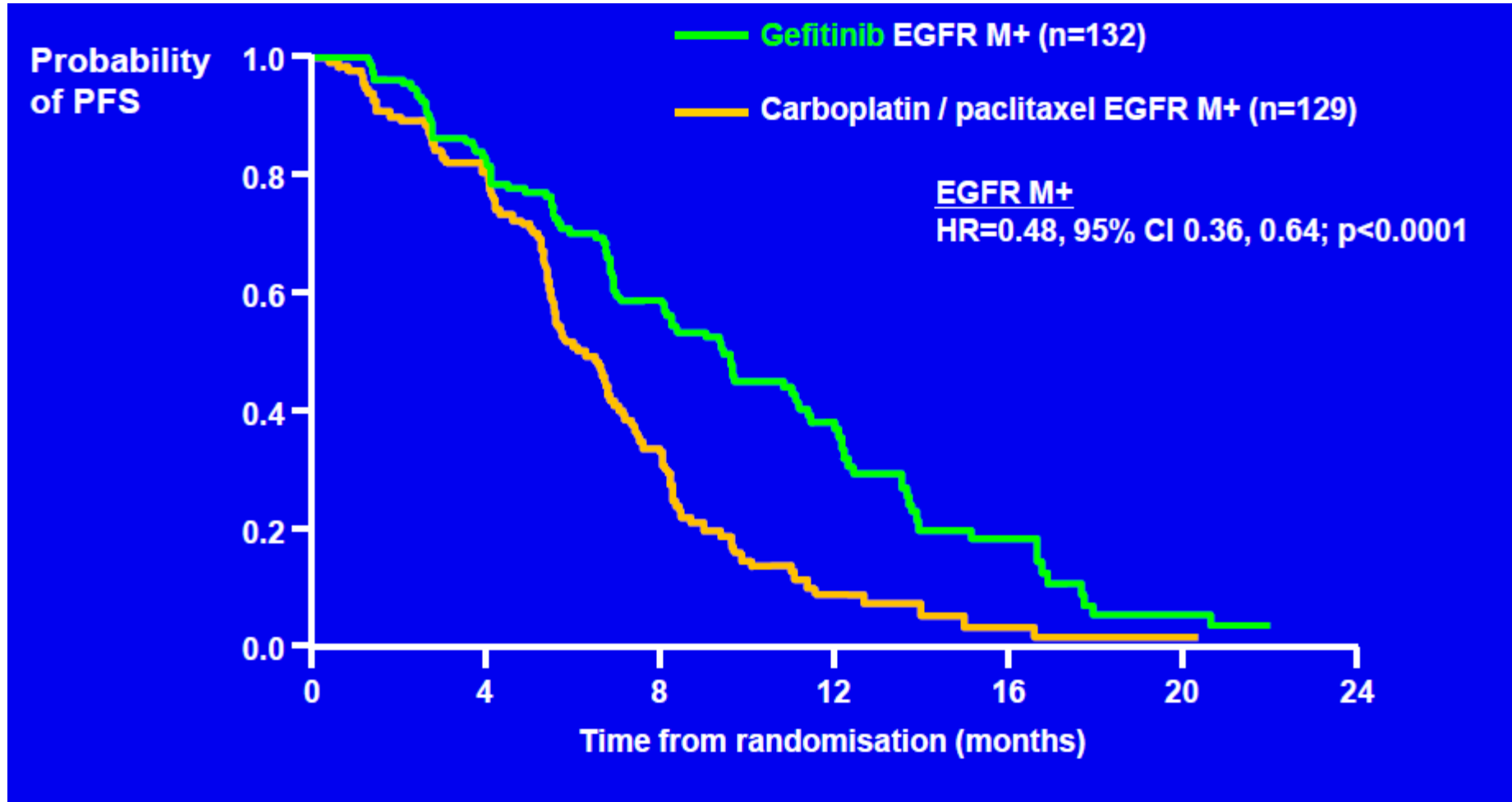
A Overall



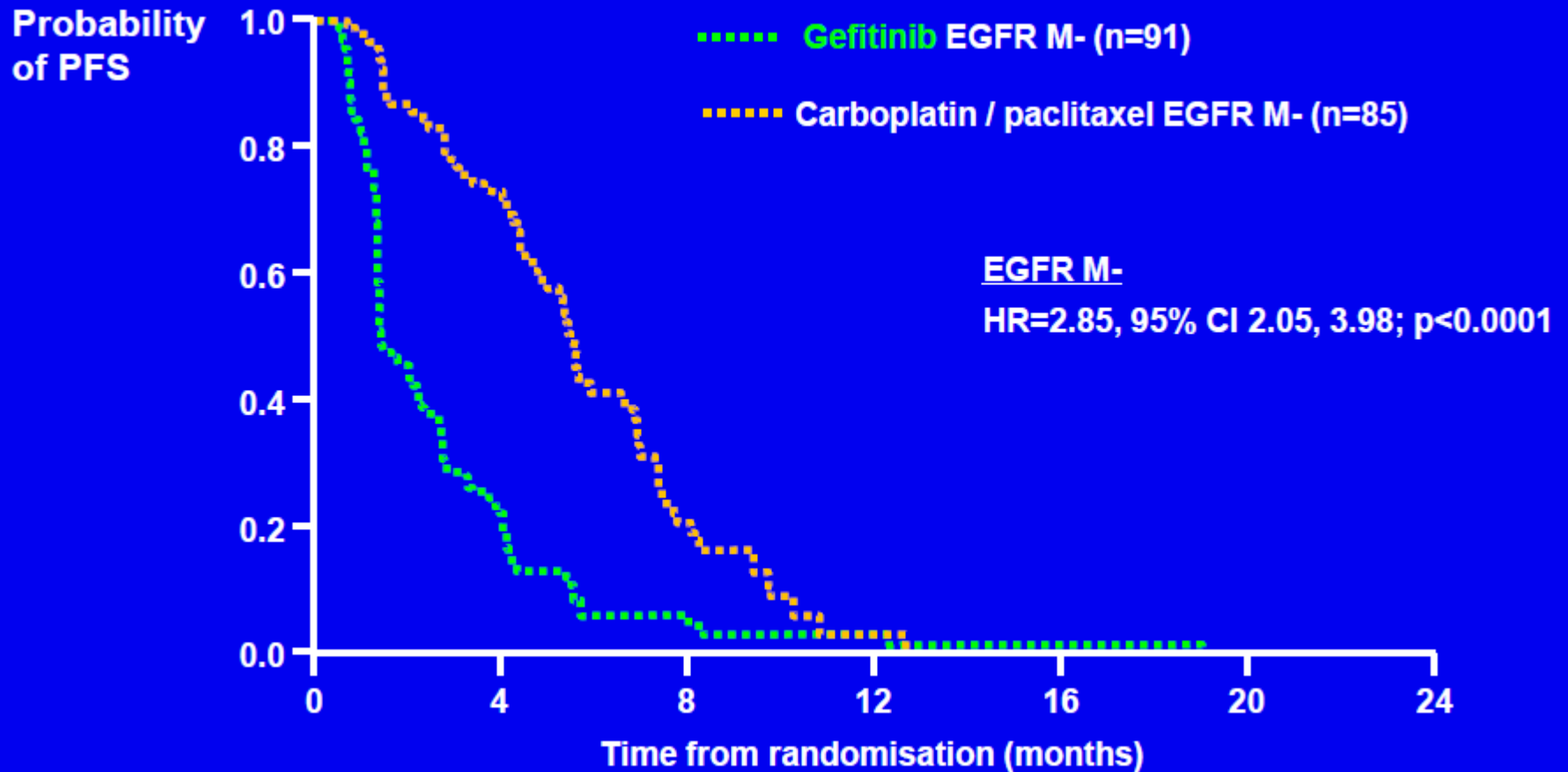
No. at Risk

Gefitinib	609	363	212	76	24	5	0
Carboplatin plus paclitaxel	608	412	118	22	3	1	0

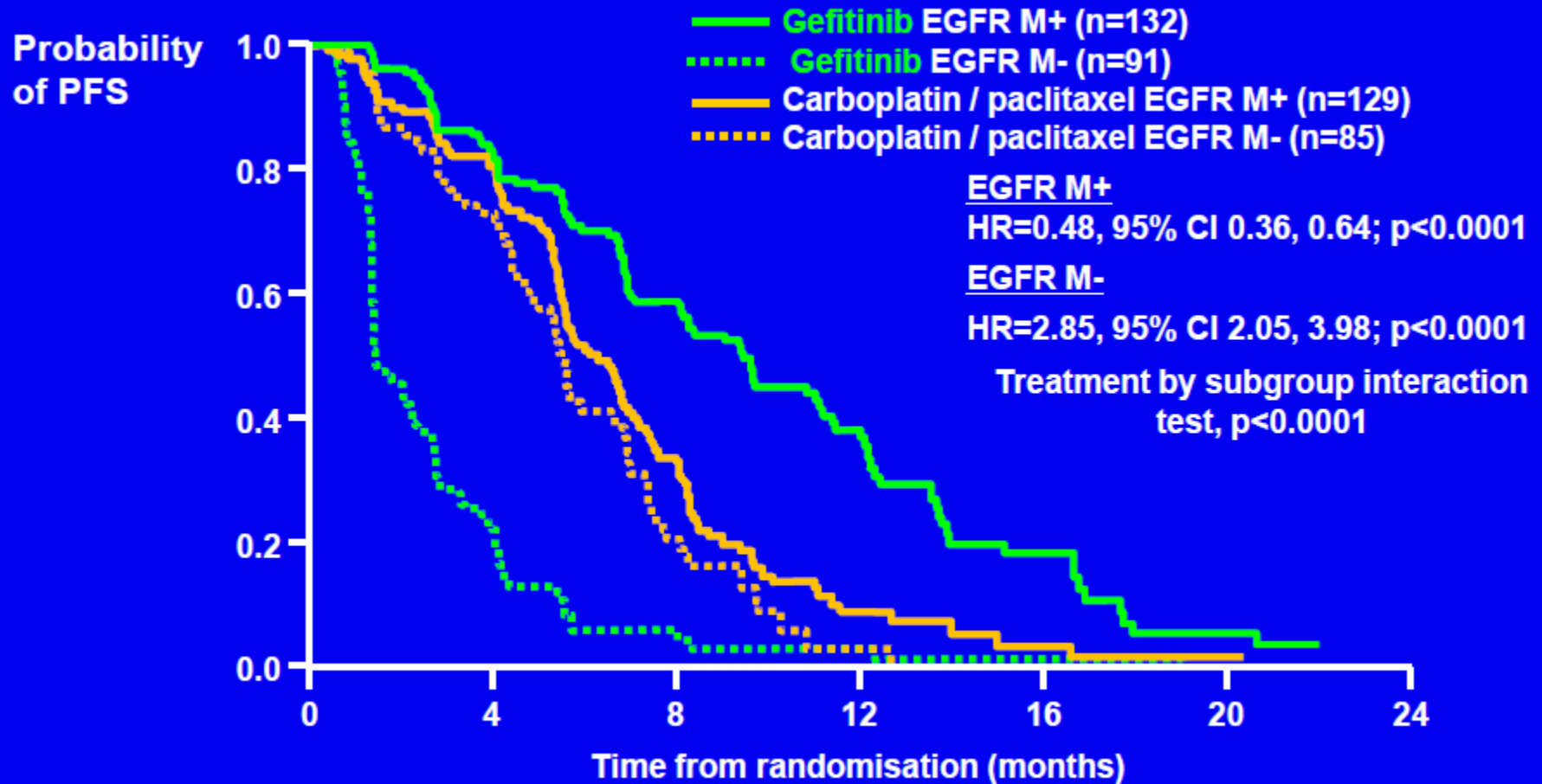
Gefitinib vs. chemioterapia u chorych z mutacją genu EGFR



Gefitinib vs. chemioterapia u chorych bez mutacji genu EGFR



Gefitinib pomaga lub szkodzi w zależności od mutacji EGFR (Qualitative Interaction)



Gefitinib pomaga lub szkodzi w zależności od mutacji EGFR (Qualitative Interaction)

Treatment HR = 0.74; $\ln(0.74) = -0.3$

Interaction HR = HR(EGFR M+) / HR (EGFR M-)
 $= 0.48 / 2.85 = 0.17$; $\ln(0.17) = -1.77$

⇒ interaction effect size = 6 × treatment effect size

⇒ interaction highly significant

Utrata mocy związana z czynnikami predykcyjnymi

- Porównujemy leczenie kontrolne z eksperymentalnym
- Dwa genetyczne podtypy chorych (przyjmujemy 50% dla każdego z podtypów) z różną prognozą
- Mediana czasu przeżycia: 4 lata dla leczenia kontrolnego, 6 lat dla chorych z dobrą prognozą, 2 lata dla chorych ze złą prognozą

Scenariusz I – ten sam efekt leczenia w dla obu podtypów

Genetic subtype	Median Control	Median Experimental	Median Diff	Hazard ratio
Good prognosis	6 y	9 y	3 y	1.5
Poor prognosis	2 y	3 y	1 y	1.5

Genetic subtype is prognostic but not predictive

Scenariusz II – brak efektu leczenia dla jednego z podtypów

Genetic subtype	Median Control	Median Experimental	Median Diff	Hazard ratio
Good prognosis	6 y	9 y	3 y	1.5
Poor prognosis	2 y	2 y	0 y	1.0

Genetic subtype is prognostic and predictive

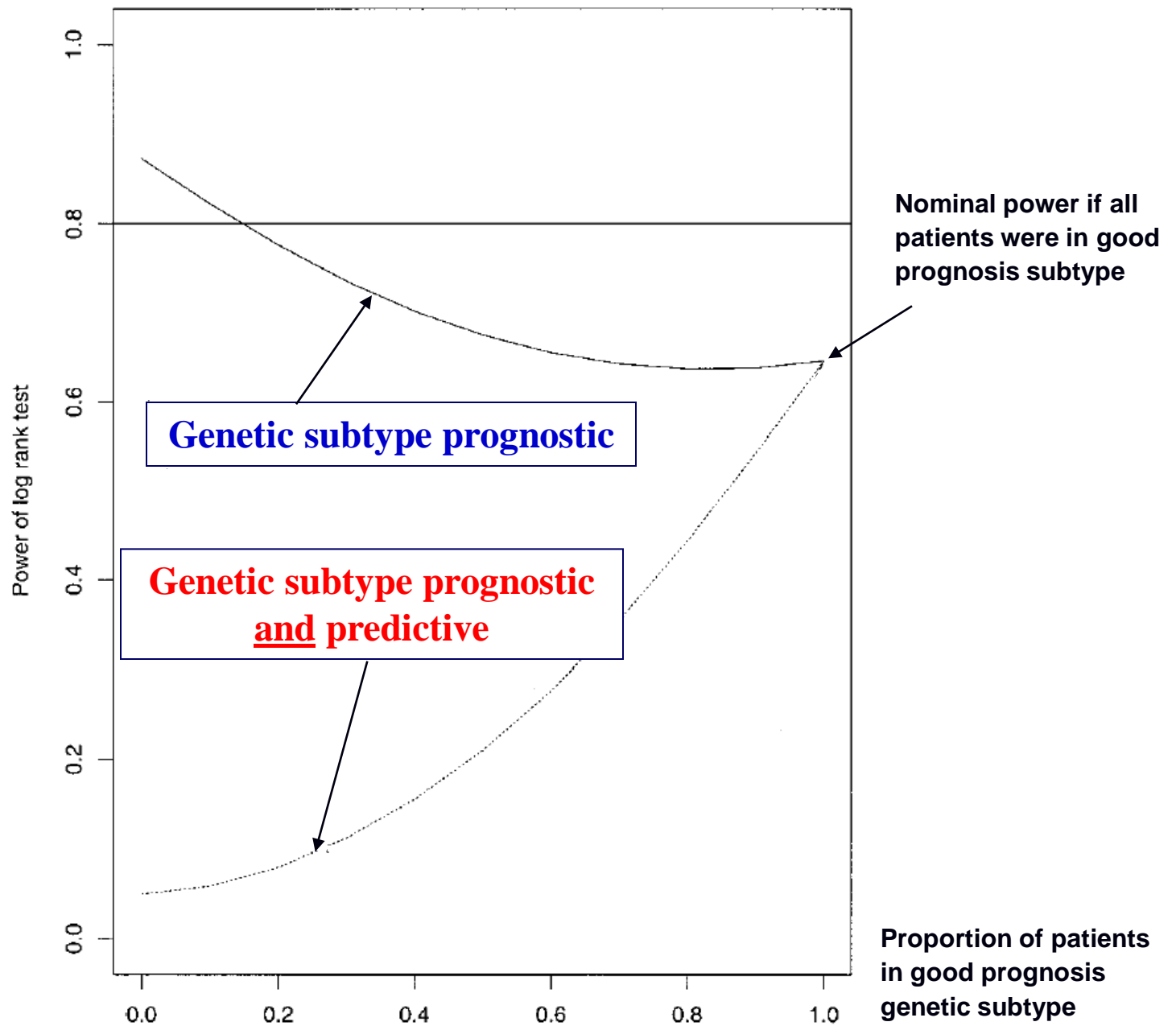


Fig 1. Power of the log-rank test under the three scenarios as a function of true proportion of genetic subtype 1 in the study population.