



Analizy wieloczynnikowe i metaanalizy – realna wiedza czy szum informacyjny?

Wojciech Saucha

Śląskie Centrum Chorób Serca w Zabrze

Oświadczenie autora prezentacji

- Nie jestem statystykiem! Jestem lekarzem. Prezentacja zawiera szereg uproszczeń często skrajnych.
- Prezentowane dane i wyniki statystyczne zawarte prezentacji są w całości wyssane z palca.
- Wszelkie podobieństwa żyjących lub nieżyjących chirurgów i ich wyników pracy do przykładów zawartych w prezentacji są w najwyższym stopniu niezamierzone.

Dowiedziałem się że jestem chory, poważnie chory.
Około 6 miesięcy temu doszło do zakażenia.

Echinocerebrokokoza

- Choroba pasożytnicza.
- Pasożyt umiejscawia się w mózgu i rośnie.
- Leczy się ją operacyjnie !
- W jednym jedynym ośrodku w kraju !
- Operuje tylko 2 chirurgów.
- Śmiertelność jest relatywnie wysoka. ☹️

- **CO MAM ROBIĆ ?**

Hypersuperklinika w Brzęczyszczkiewicach

operacje Echinocerebrokokozy

wykonane w 2012 roku

Operator	Chory Zmarł	Chory Przeżył	Łącznie operowanych
Dr Mortimer	50	50	100
Dr Angelo	20	80	100

Jakie jest prawdopodobieństwo zgonu?

Operator	Chory Zmarł	Chory Przeżył	łącznie operowanych
Dr Mortimer	50	50	100
Dr Angelo	20	80	100

Dr Mortimer $p=50/100=0,5$

Dr Angelo $p=20/100=0.2$

Prawdopodobieństwo p mieści się w przedziale od 0 do 1

JAKIE SĄ MOJE SZANSE?

Szansa =Odds

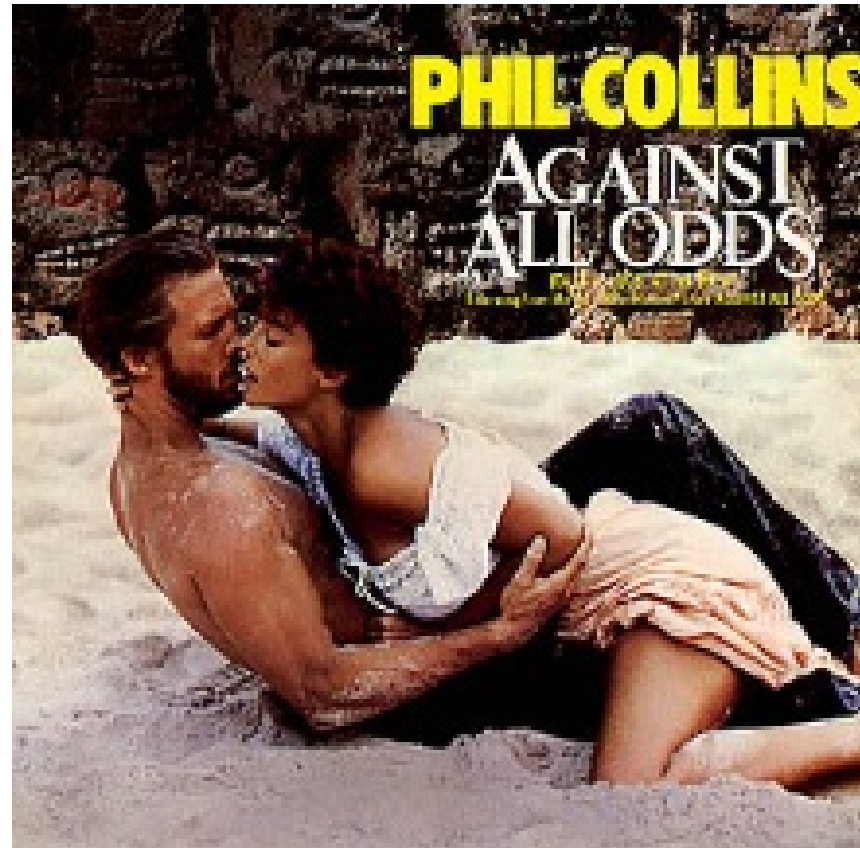
Szansa

Operator	Chory Zmarł	Chory Przeżył	Łącznie operowanych
Dr Mortimer	50	50	100
Dr Angelo	20	80	100

Szansa zgonu po operacji dr Mortimera $O=50/50=1$

Szansa zgonu po operacji dr Angelo $O=20/80=0,25$

Szansa mieści się w przedziale od 0 do ∞ i $O=p/(1-p)$



*„And you coming back to me is against all odds
That's the chance I've got to take”*

Iloraz szans (Odds ratio)

Szansa zgonu po operacji dr Mortimera $O=50/50=1$

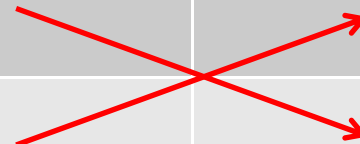
Szansa zgonu po operacji dr Angelo $O=20/80=0,25$

$$\text{OR} = \frac{\text{Szansa zgonu po operacji dr Mortimera}}{\text{Szansa zgonu po operacji dr Angelo}} = \frac{1}{0,25} = 4$$

Iloraz szans (Odds Ratio)

Miara związku pomiędzy narażeniem na czynnik a wynikiem oddziaływania.

Operator	Chory Zmarł	Chory Przeżył
Dr Mortimer	50	50
Dr Angelo	20	80



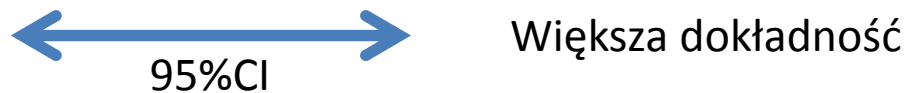
$$OR = (50 \cdot 80) / (20 \cdot 50) = 4000 / 1000 = 4$$

OR	Znaczenie
=1	Brak wpływu
<1	Efekt ochronny
>1	Efekt szkodliwy

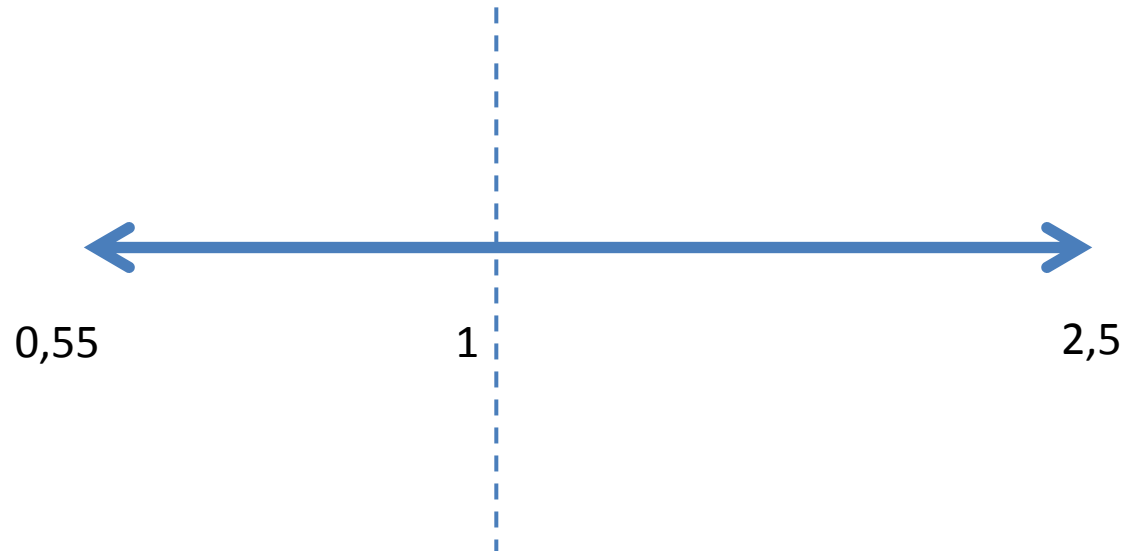
**W JAKI SPOSÓB DANE Z ROKU 2012
ODNIEŚĆ DO POPULACJI OGÓLNEJ?**

Przedział ufności (Confidence interval CI)

- Pomaga oszacować dokładność OR
 - (Najczęściej oznacza się 95%CI)



95%CI dla OR



OR	Znaczenie
=1	Brak wpływu
<1	Efekt ochronny
>1	Efekt szkodliwy

95%CI dla OR

Operator	Chory Zmarł	Chory Przeżył
Dr Mortimer	50	50
Dr Angelo	20	80

$$OR = (50 \cdot 80) / (20 \cdot 50) = 4000 / 1000 = 4$$

95%CI (2.1357 - 7.4917)

OR	Znaczenie
=1	Brak wpływu
<1	Efekt ochronny
>1	Efekt szkodliwy

Dlaczego wyliczamy OR?

- W badaniach typu *case-control* iloraz szans w przybliżeniu jest tożsamy z **ryzykiem względnym** (*relative risk*).
- W związku z tym w badanym okresie czasu **ryzyko zgonu gdy operatorem był dr Mortimer było 4-krotnie większe niż gdy operował dr Angelo.**

OR i 95%CI

Nic nie mówi o znamienności statystycznej!

**CZY ISTNIEJE STATYSTYCZNIE
ZNAMIENNA RÓŻNICA MIĘDZY
WYNIKAMI LECZENIA OPERATORÓW?**

Test Chi²

Operator	Chory Zmarł	Chory Przeżył
Dr Mortimer	50	50
Dr Angelo	20	80

OR = 4, 95%CI (2.1357 - 7.4917); p<0,001

Istnieje statystycznie znamienna różnica pomiędzy wynikiem leczenia chorych operowanych przez dr Mortimera i dr Angelo!

**CZY SĄ TO WYSTARCZAJĄCE DANE BY
ŚWIADOMIE WYBRAĆ LEPSZEGO
OPERATORA?**

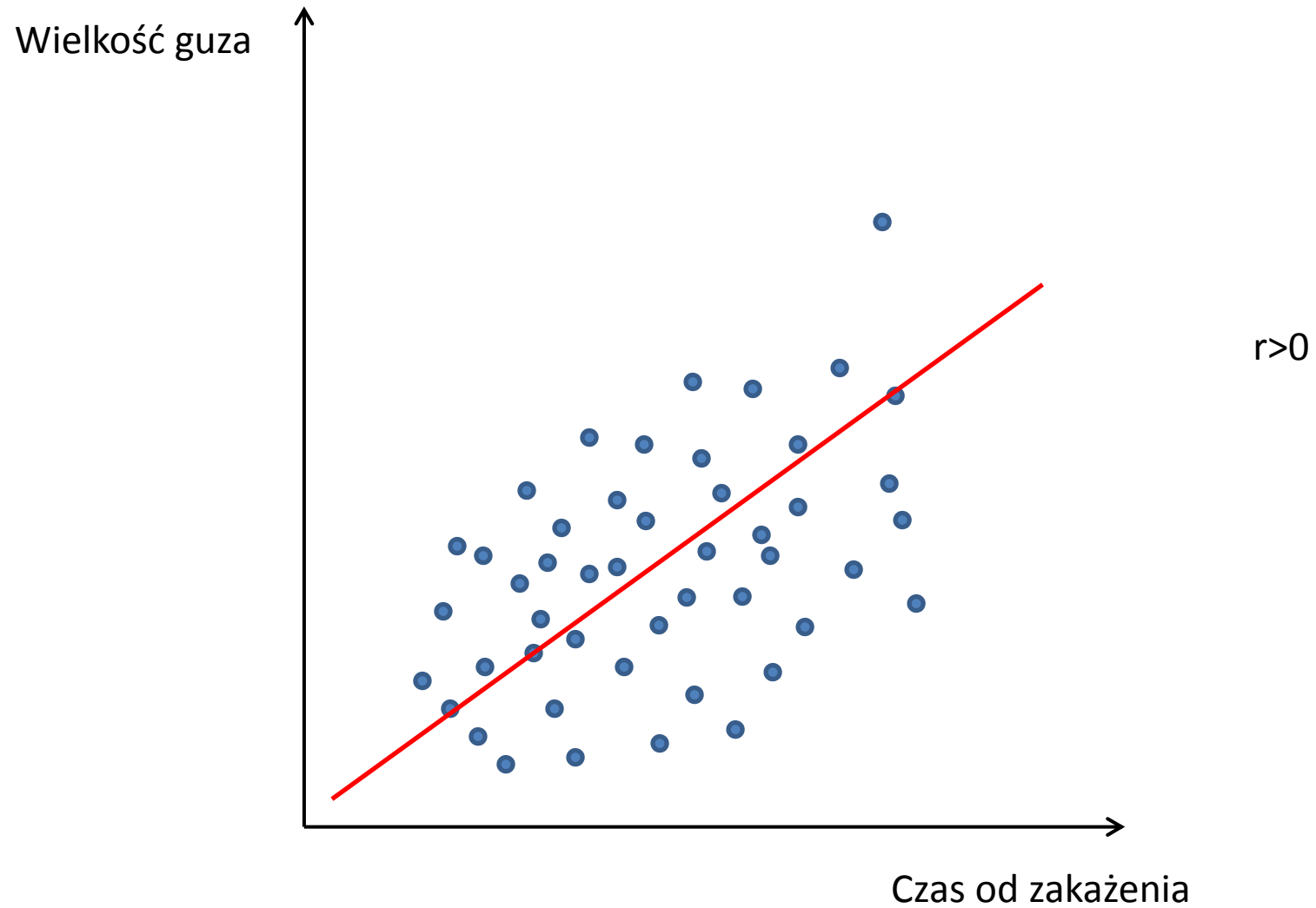
Co jeszcze wiemy o chorobie?

Dane z literatury:

- Śródoperacyjnie zmierzona wielkość guza pasożytniczego jest głównym czynnikiem ryzyka operacyjnego.

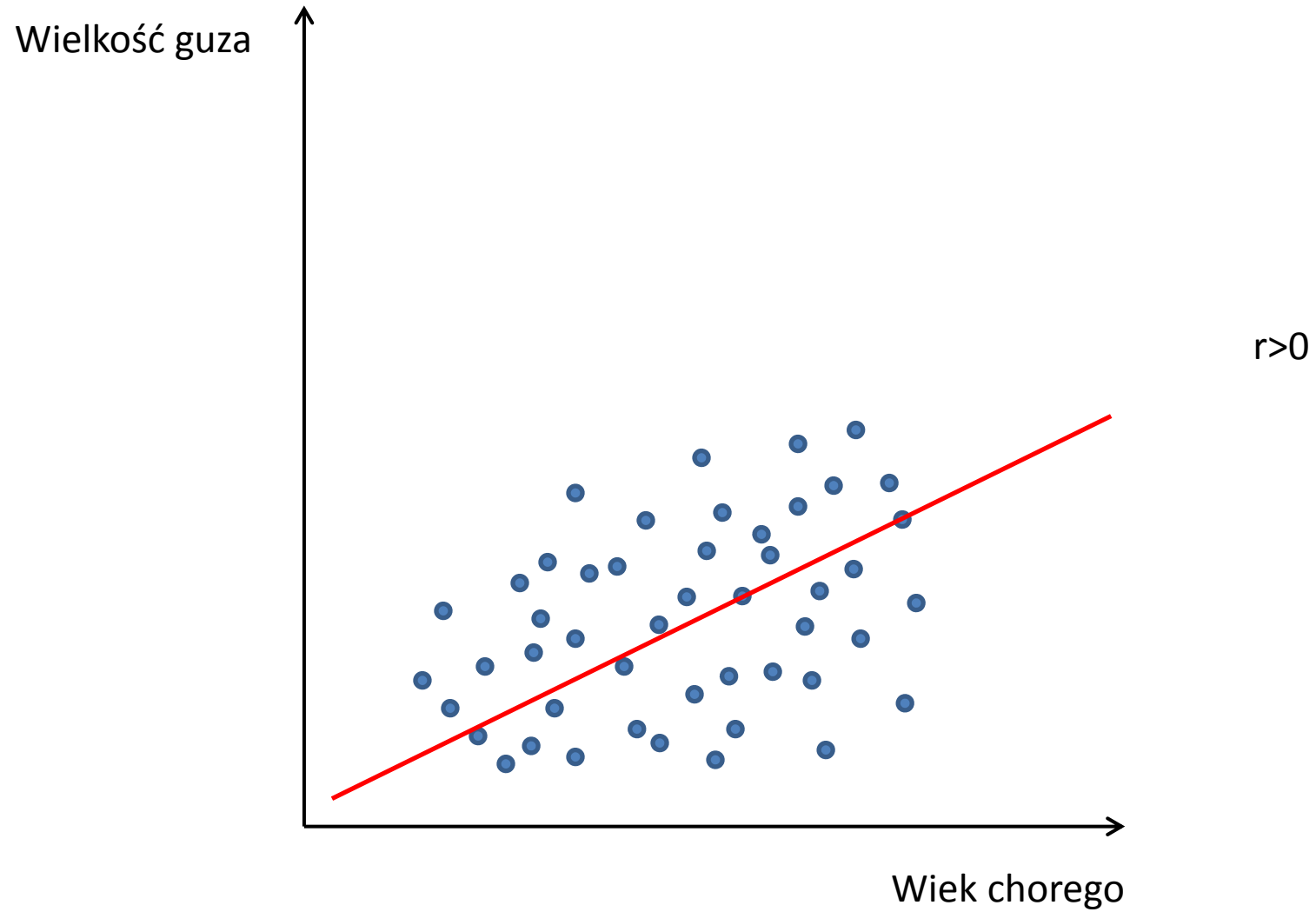
Inne zależności ?

korelacje



Inne zależności ?

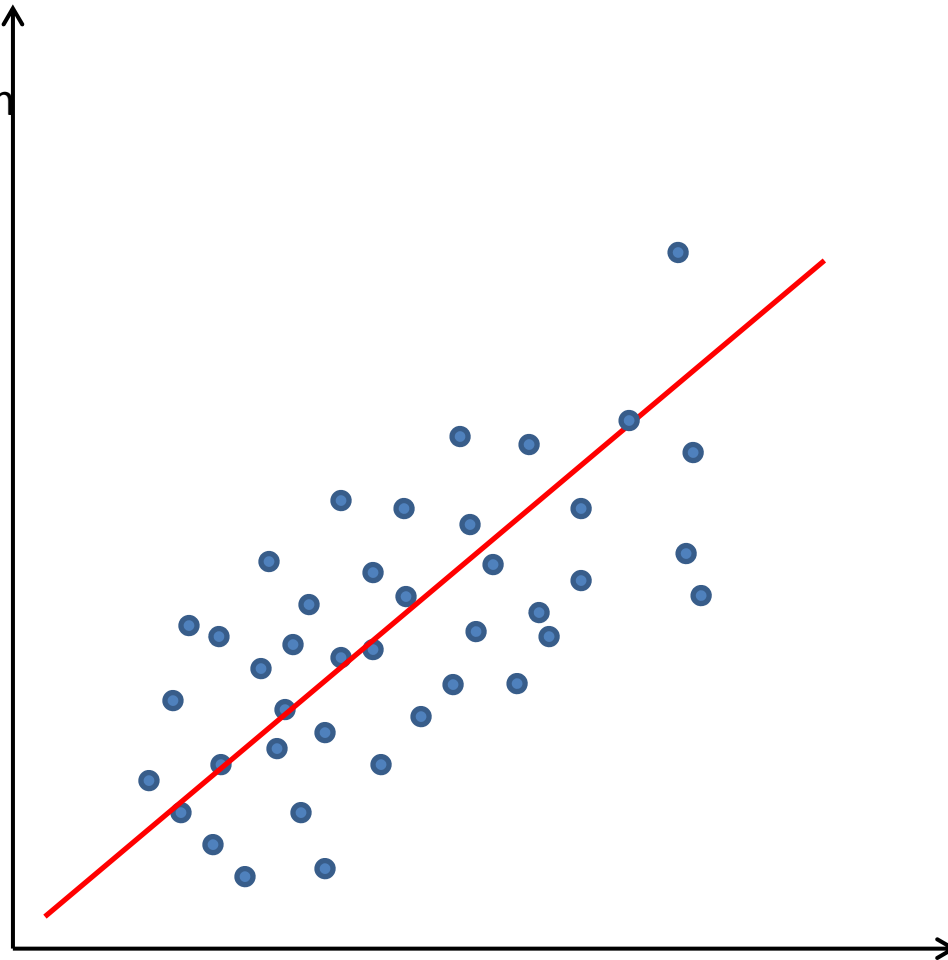
korelacje



Inne zależności ?

korelacje

Wielkość guza
po 12 miesiącach
od zakażenia

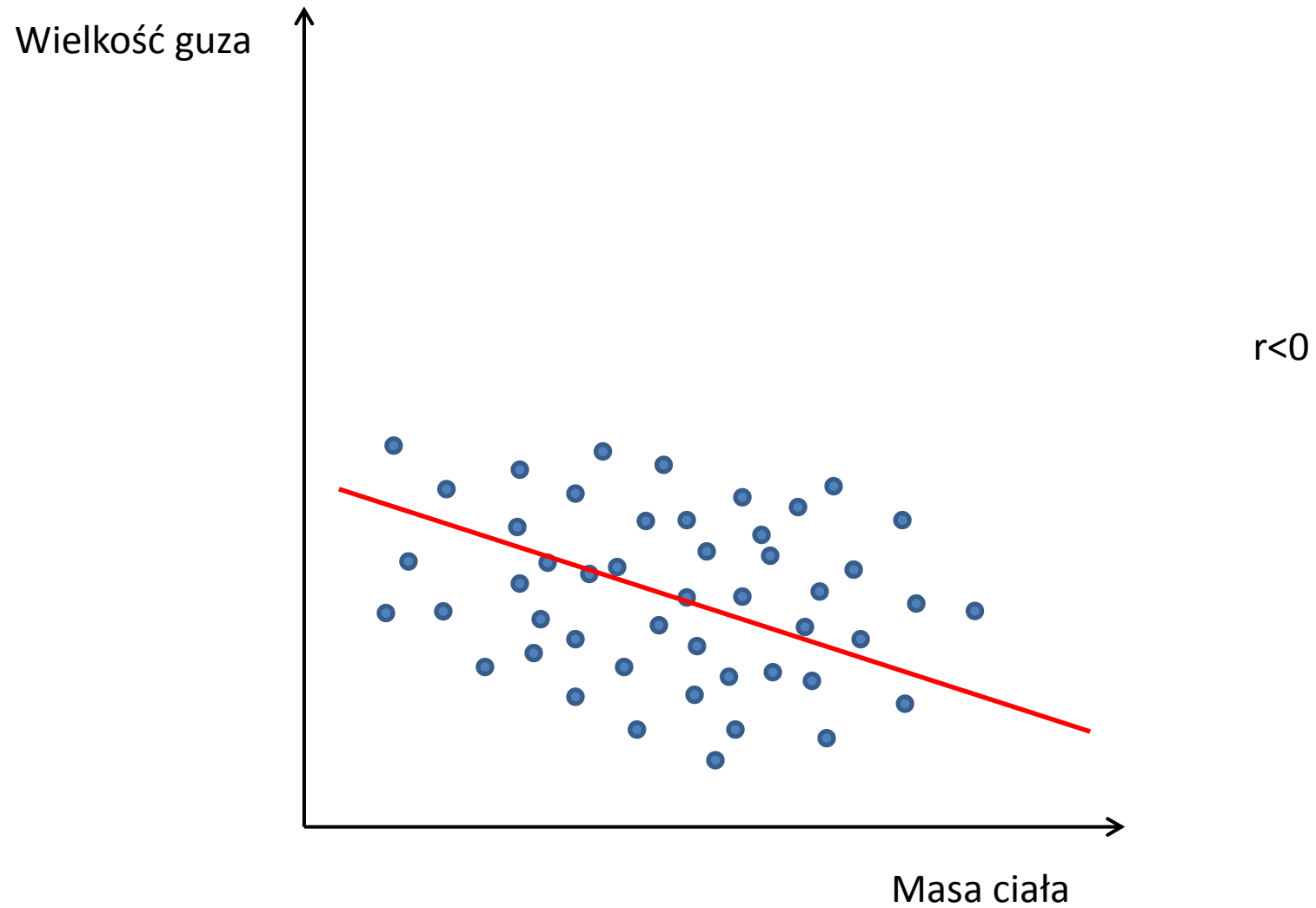


$r > 0$

Wiek chorego

Inne zależności ?

korelacje



Macierz korelacji

	Wielkość guza	Czas od zakażenia	Wiek chorego	Masa ciała
Wielkość guza	1	0,6	0,3	-0.4
Czas od zakażenia	0,6	1	0,2	-0,3
Wiek chorego	0,3	0,2	1	0,4
Masa ciała	-0.4	-0,3	0,4	1

**WIEMY, ŻE ISTNIEJĄ KORELACJE
POMIĘDZY ZMIENNYMI**

**JAK TO MOŻE POMÓC ROZWIĄZAĆ
PROBLEM DR MORTIMERA?**

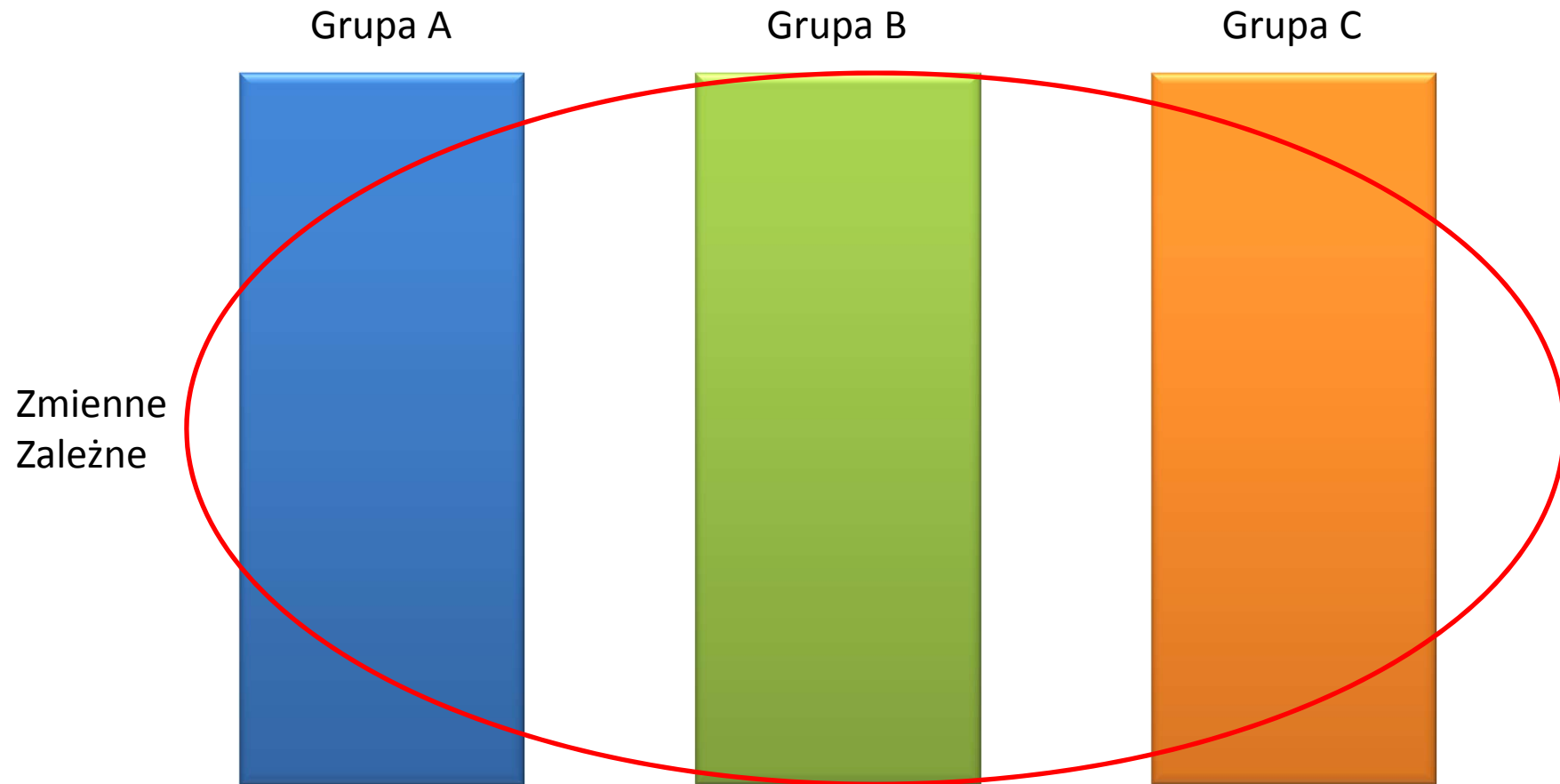
**DANE, DANE, POTRZEBUJĘ
DANYCH!!!!**

Materiał do analizy !

	Operator	Wiek	Wielk.Guz	Czas od Z.	Masa C.	Zgon
Chory 1	Mortimer	55	35	18	81	0
Chory 2	Angelo	35	12	10	55	0
Chory 3	Angelo	34	16	12	61	0
Chory 4	Mortimer	65	31	22	72	1
Chory 5	Mortimer	55	28	19	77	1
Chory 6	Angelo	21	18	10	45	1
....
...
Chory 200	Mortimer	66	30	24	67	0

**MANOVA CZYLI
MULTIVARIATE ANALYSIS OF VARIANCE**

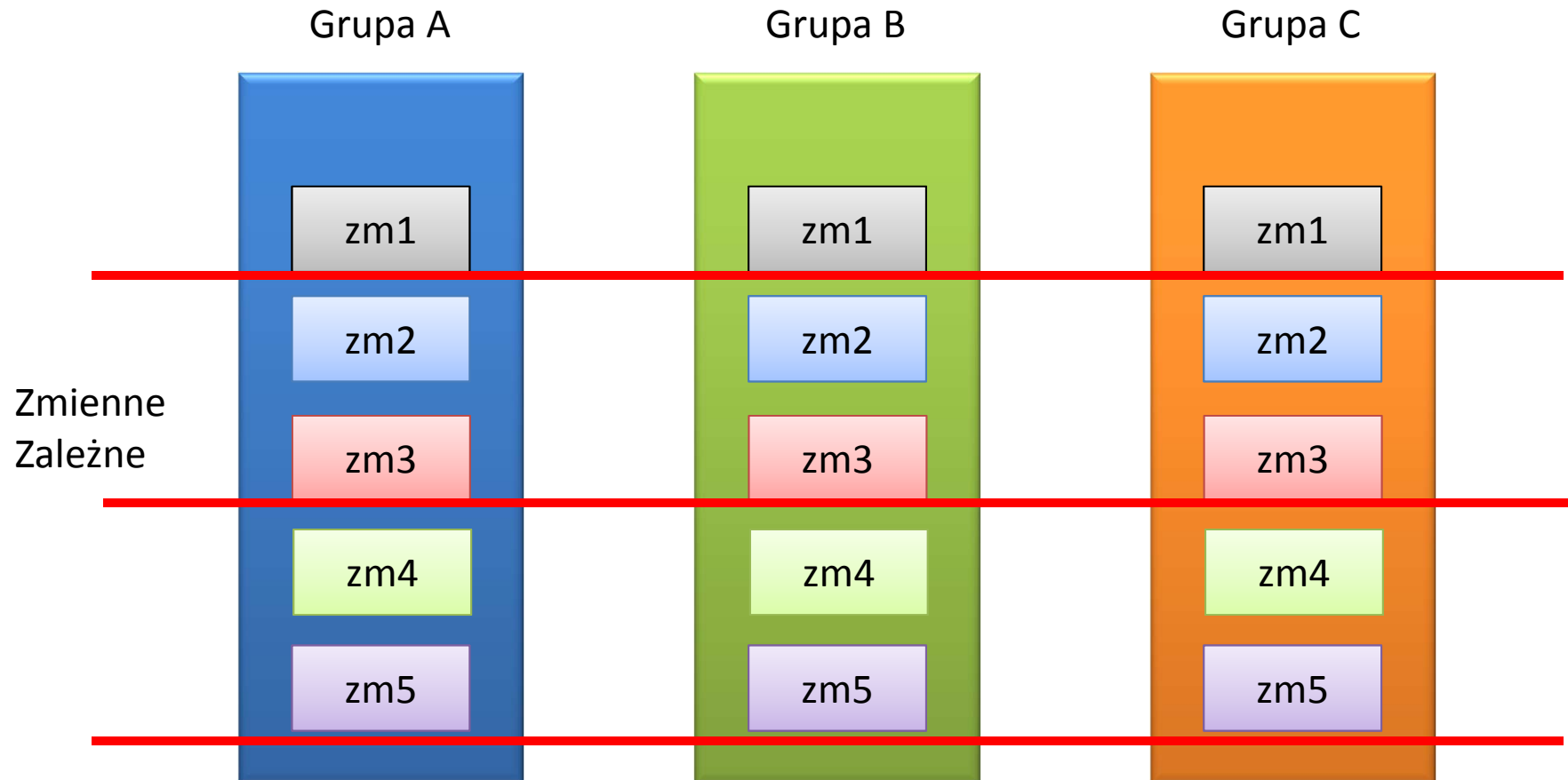
Krok 1



Pytanie 1 - czy istnieje różnica pomiędzy grupami (efekt główny)?

Odpowiedź – TAK lub NIE

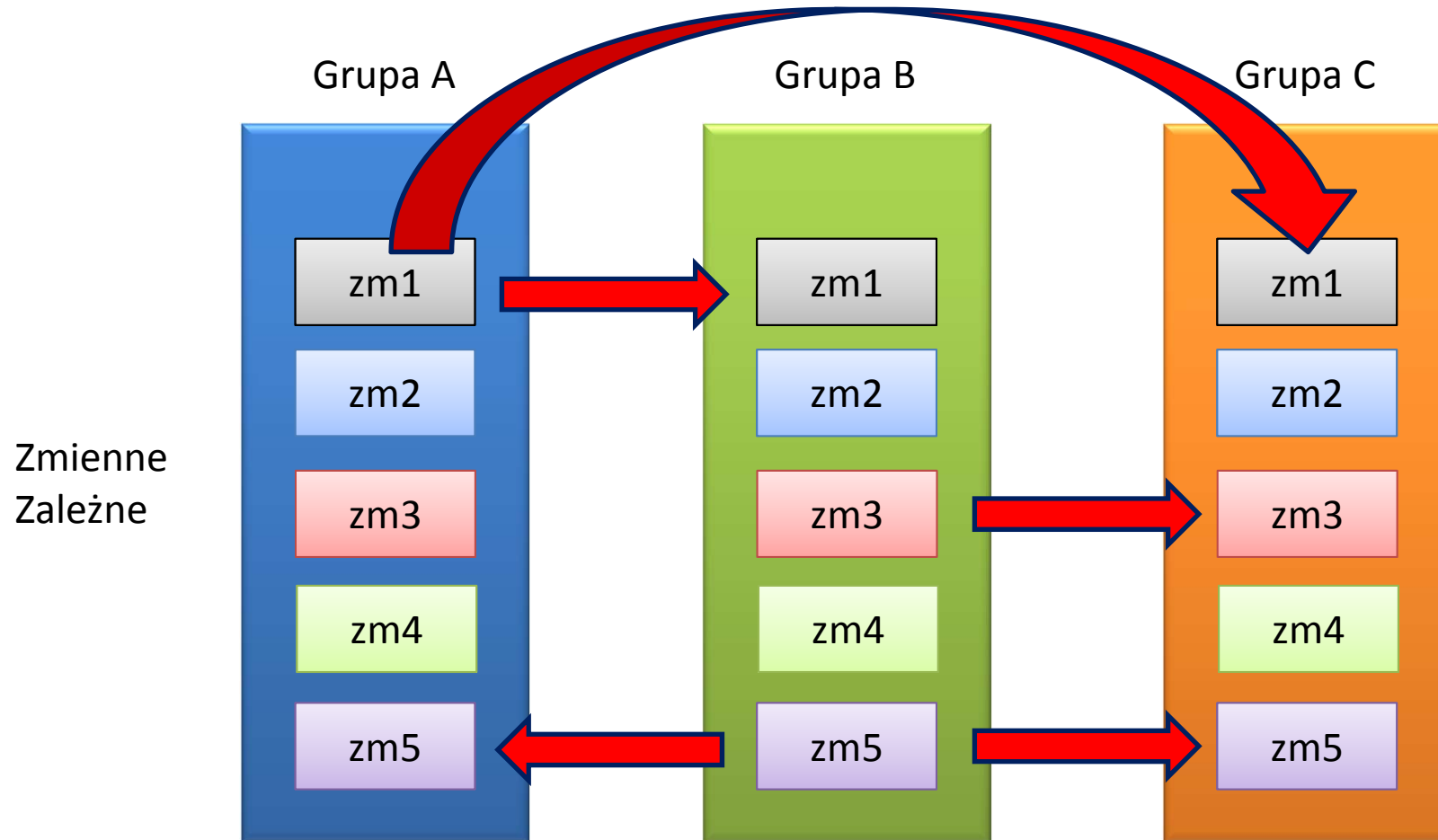
Krok 2



Pytanie 2 – Jakie zmienne różnią się pomiędzy grupami?

Odpowiedź – zm1 TAK,
zm2 NIE,
zm3 TAK.....itd

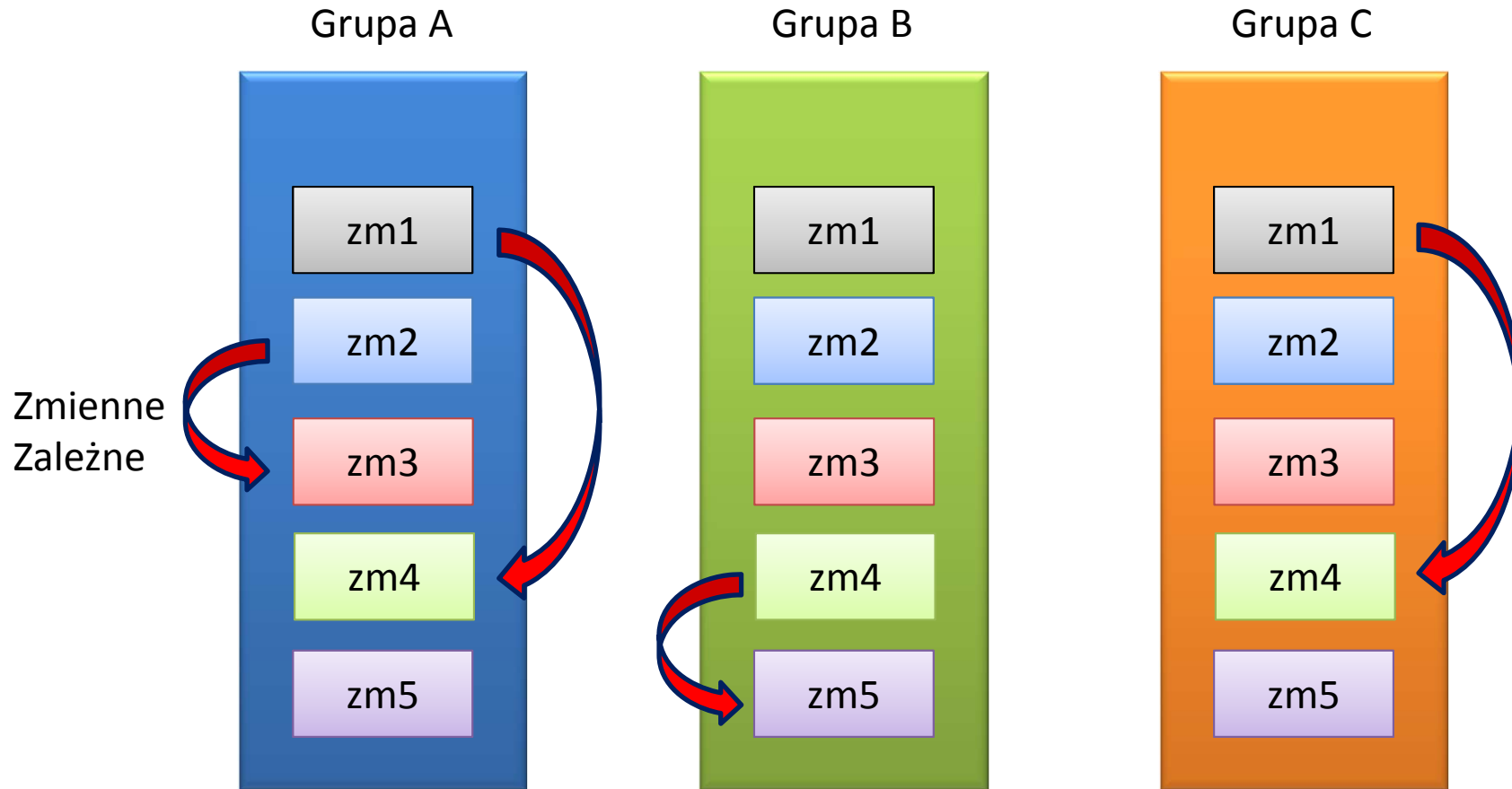
Krok3



Pytanie 3 – Które zmienne różnią się pomiędzy którymi grupami?

Odpowiedź – zm1 Grupa A różni się od Grupy B i Grupy C,
zm3 Grupa B różni się od Grupy Citd
Oraz jaki jest kierunek zmian.

Krok 4



Pytanie 4 – Jak zmienne w obrębie grup wpływają na siebie nawzajem ?

Odpowiedź –

zm1 ze zm4 w Grupie A,

zm2 ze zm3 w Grupie A ,

zm4 ze zm5 w Grupie B ...itd

MANOVA

Do analizy włączono : Grupa – operator
oraz zmienne: wiek, wielk_guz, czas_od_z, masa_c,

<p>Główny efekt wieloczynnikowy. <i>Jaki jest główny efekt pomiędzy grupami?</i></p>	<p>Istnieje znamieny wieloczynnikowy efekt główny pomiędzy operatorami Wilks'λ=0,423, F (6, 92,033)=3,2 , p<0,001 Partial eta sq = 0,211. Siła wykrycia efektu 0.961</p>											
<p><i>Jakie zmienne różnią się pomiędzy grupami?</i></p>	<p>Zaobserwowano znamienne różnice pomiędzy populacjami leczonymi przez operatorów dla Wiek F(2, 21) = 2,45 , p<0,01 Partial eta sq=0,24, Siła =0,79 Wielkości guza F(2, 21), p<0,01 Partial eta sq=0,34, Siła =0,89 Czas od z. F(2, 21), p<0,001 Partial eta sq=0,39, Siła =0,91 Brak zamiennych różnic dla zmiennej masa ciała.</p>											
<p><i>Jakie jest znaczenie zmiennych ?</i></p>	<p>Średnie wartości zmiennych zależnych podano poniżej:</p> <table border="1" data-bbox="815 1139 2029 1289"> <tr> <td data-bbox="815 1139 1102 1214">Dr Mortimer</td> <td data-bbox="1102 1139 1422 1214">Wiek 54</td> <td data-bbox="1422 1139 1727 1214">Wielk_g 32</td> <td data-bbox="1727 1139 2029 1214">Czas_od_z 18</td> </tr> <tr> <td data-bbox="815 1214 1102 1289">Dr Angelo</td> <td data-bbox="1102 1214 1422 1289">Wiek 34</td> <td data-bbox="1422 1214 1727 1289">Wielk_g 18</td> <td data-bbox="1727 1214 2029 1289">Czas_od_z 12</td> </tr> </table>				Dr Mortimer	Wiek 54	Wielk_g 32	Czas_od_z 18	Dr Angelo	Wiek 34	Wielk_g 18	Czas_od_z 12
Dr Mortimer	Wiek 54	Wielk_g 32	Czas_od_z 18									
Dr Angelo	Wiek 34	Wielk_g 18	Czas_od_z 12									
<p><i>Z jaką siłą zmienne w obrębie grup są ze sobą powiązane?</i></p>	<p>Nie wykonywano analizy</p>											

MANOVA - zalety

- Ocenia wszystkie badane zmienne „za jednym zamachem”
- Jest większa szansa wykrycia który czynnik jest tak naprawdę istotny.
- Zabezpiecza przed błędami wynikającymi z porównywania metodą ANOVA zmiennych „każdy z każdym” (kumulacja błędów)
- Zapewnia systematyczność oceny wyników poprzez ściśle określony algorytm.

MANOVA -wady

- Jest skomplikowana.
- Można w ten sposób oceniać tylko pewne dane.
 - Normalność wszystkich rozkładów zmiennych
 - Jednorodność wariancji
 - Itd.
- Zależności pomiędzy zmiennymi w obrębie grup są analizowane tylko liniowo.

MANOVA - wnioski

- Dobra metoda statystyczna ale nie jest uniwersalna.
- Wykonana bez sprawdzenia warunków jej zastosowania może prowadzić do błędnych wniosków.
- Odpowiednie zastosowanie tej metody i zachowanie poprawności metodycznej znacznie podnosi rangę pracy naukowej.
- Należy być krytycznym w stosunku do prac które podają tylko gołe znamienności statystyczne bez podania „kuchni statystycznej” umożliwiającej weryfikację podanych wyników!

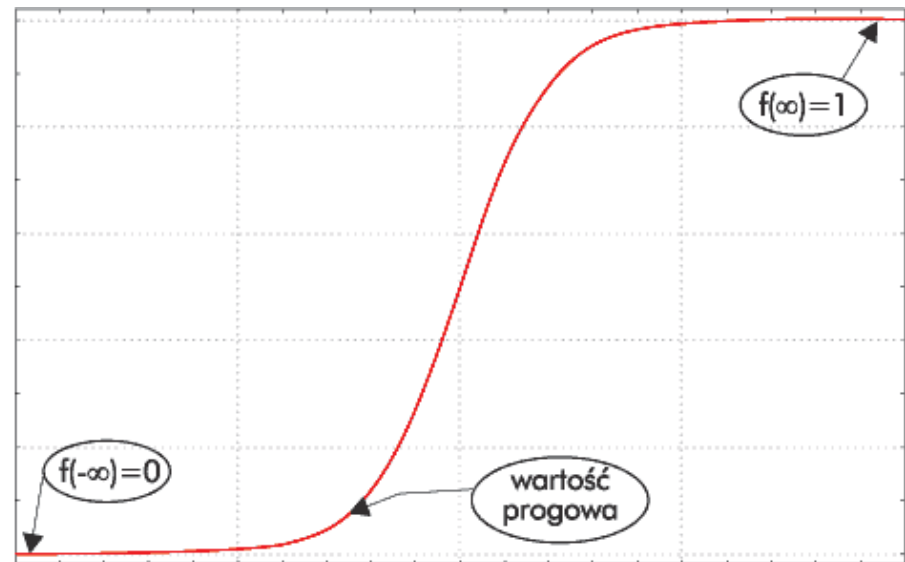
**A JEŚLI CHCIAŁBYM POLICZYĆ JAKIE
JEST MOJE RYZYKO OPERACYJNE?**

Regresja logistyczna

- Nadaje się do oceny zjawisk o charakterze dychotomicznym (np. tak/nie; 0/1; przeżycie/zgon).
- Zmienne opisujące mogą być zarówno ilościowe jak i jakościowe.

Funkcja logit

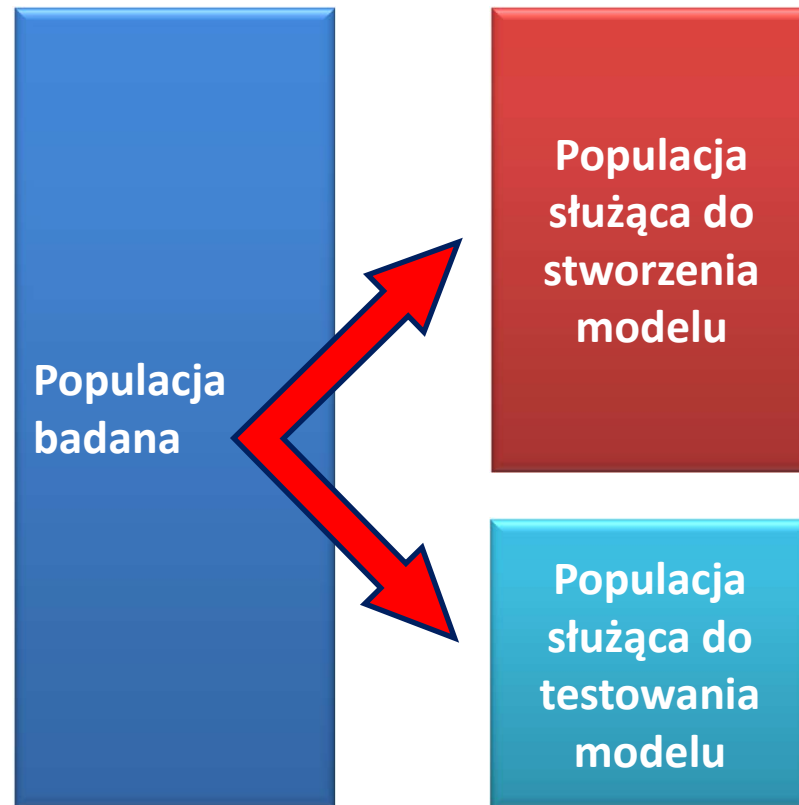
- Przyjmuje wartość od 0 do 1.
- Dobrze odwzorowuje procesy w biologii i medycynie.
- Charakterystyczna jest tzw. wartość progowa.



Regresja logistyczna – jak to się robi

1. Stworzenie modelu matematycznego.
 - Dobór zmiennych do modelu.
 - Metoda Forward
 - Metoda Backward
2. Testowanie modelu poprawności modelu.
3. Prospektywne testowanie modelu.

Regresja logistyczna



Regresja logistyczna - wyniki

- Utworzony model regresji logistycznej „przy okazji” podaje:
 - **Iloraz szans** (odds ratio) dla każdej zmiennej w modelu.
 - **Istotność** każdej zmiennej w modelu podając p.

Wracamy do dr Mortimera i dr Angelo

Zmienna badana/estymowana: Zgon

Zmienne opisujące: Operator; Wiek; Wielk_g; Czas_od_z

	Operator	Wiek	Wielk_g	Czas_od_z
Poziom p	NS	<0.01	<0,001	<0,001
Iloraz szans (OR)	0,9	2,1	5,4	4,8
-95% CL	0,35	1,5	3,4	2,8
+95% CL	1,55	4,2	6,0	5,8

Podstawiając moje dane do modelu

Jeśli będę się operował w Hypersuperklinice w Brzeczyszczykiewiczach to:

- Ryzyko zgonu (logistic) wynosi **15%**.
- Jest niezależne od operatora.

P.S.

Dr Mortimer raczej nie jest faworyzowany przez układającego plan operacyjny.

.....Dowiedziałem się że plan układu dr Angelo.

A MOŻE OPERACJA ZA GRANICĄ?

Fakty

- Badania medyczne prowadzone są często na małych grupach chorych szczególnie w przypadku rzadkich chorób. W związku z tym statystyczna siła tych badań jest niewielka.
- Badania kliniczne są często przeprowadzane na niereprezentatywnych populacjach i ich wyników nie można odnieść do populacji ogólnej.

Fakty cd

- Badania kliniczne przeprowadzane w jednym ośrodku są podatne na zafałszowanie wyników (*bias*) ze względu na specyficzne cechy danego ośrodka takie jak:
 - Nawyki i umiejętności personelu
 - Polityka przyjęć do szpitala
 - Różnice etniczne związane z geograficznym położeniem ośrodka
 - Powtarzalne błędy np. laboratoryjne
 - Itd...

Rozwiązania

- Duże prospektywne badania wielośrodkowe.
- Przeglądy systematyczne (Systematic reviews).
- Metaanalizy.

Metaanalizy

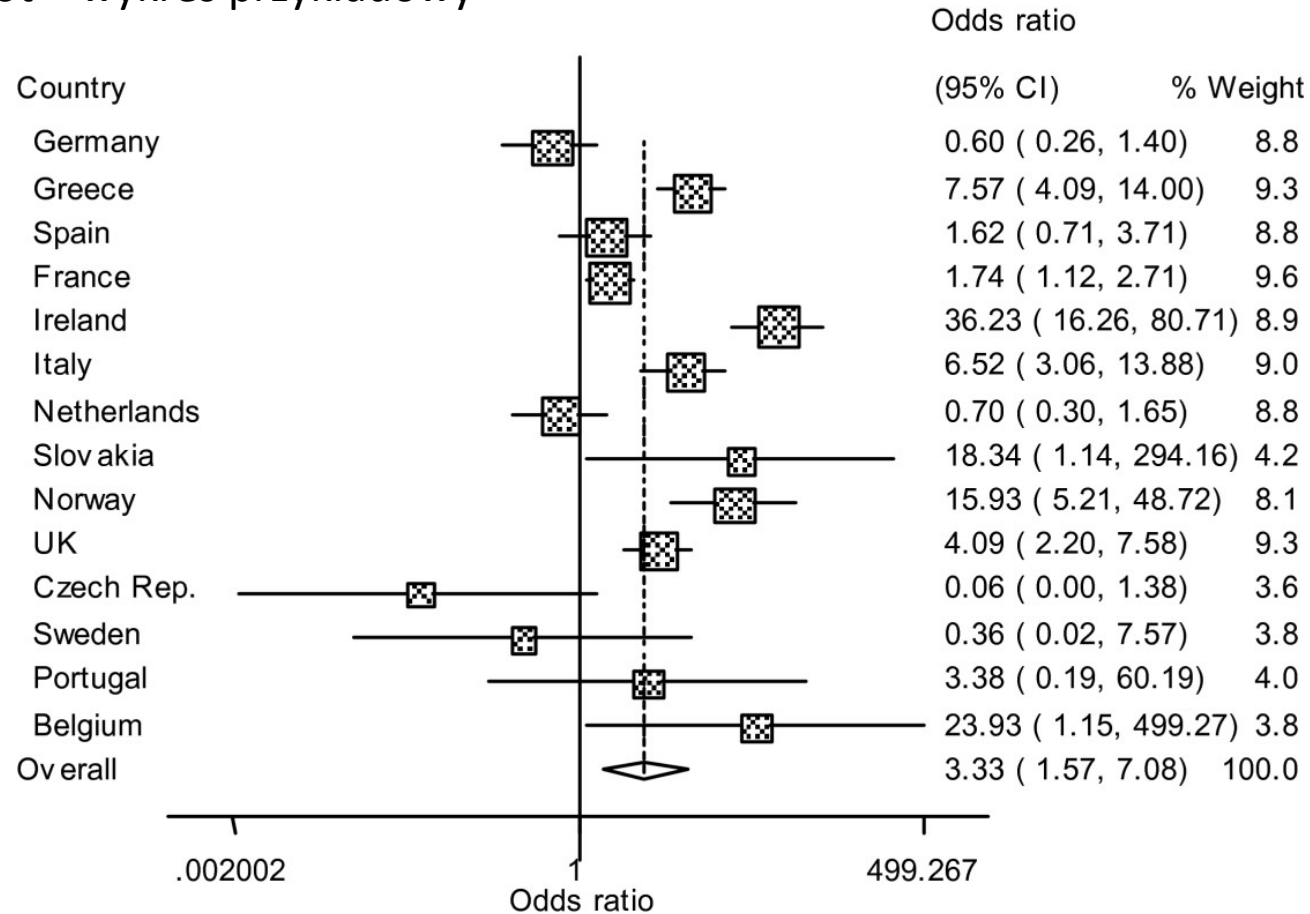
- Metoda statystyczna, która łączy ze sobą wiele niezależnych od siebie podobnych prac tak aby na drodze skomplikowanych obliczeń estymować **jeden wspólny wynik**.
- Poprzez dobór odpowiednich prac do analizy oraz odpowiednich metod statystycznych, zwiększa się siłę.

Metaanalizy

- Opracowywane są według jednolitego wzorca czyli istnieją protokoły według których metaanalizy muszą prezentowane. *Quality of Reports of Meta-analysis [QUOROM] i Observational Studies in Epidemiology [MOOSE]*.
- Praktycznie nie jest możliwe przygotowanie metaanalizy bez udziału statystyka lub/i epidemiologa.

Prezentacja wyników metaanaliz

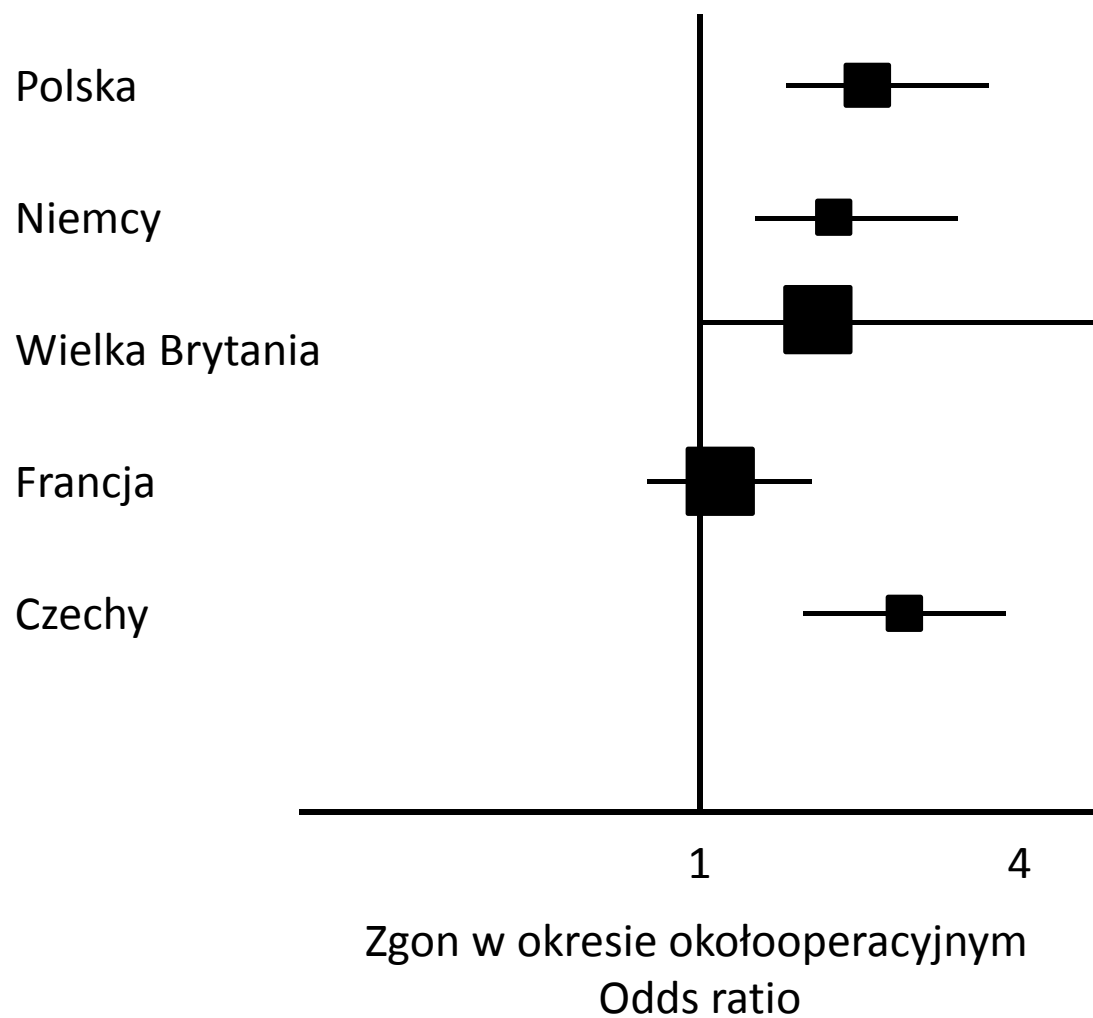
Forest plot – wykres przykładowy



Podsumowanie

- Metaanalizy są źródłem wiedzy popartej racjonalnie silnymi dowodami statystycznymi.
- Często w niektórych dziedzinach medycznych są jedynym źródłem, które systematyzuje dotychczas uzyskane wyniki badań, przynajmniej do czasu wykonania badań wielośrodkowych.
- Koszt wykonania metaanalizy jest nieporównywalny z kosztem wykonania dużego badania wielośrodkowego.

Echinocerebrokokoza w metaanalizie





Merci pour votre attention.

Au revoir...