

XXIX KONFERENCJA

ZASTOSOWANIA
STATYSTYKI
I DATA MINING
W BADANIACH
NAUKOWYCH



Praktyczne i teoretyczne wyzwania badań klinicznych okiem biostatystyka

Adrian Olszewski

2KMM (CRO)

Kilka słów o mnie...



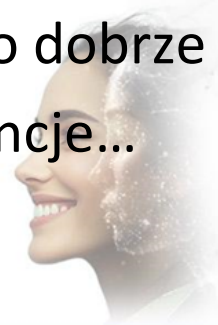
- Biostatystyk w firmie 2KMM CRO
- Obserwacyjne i eksperymentalne badania kliniczne
- Projektowanie i analiza | GNU R
- **Wróż na pełen etat 😊**



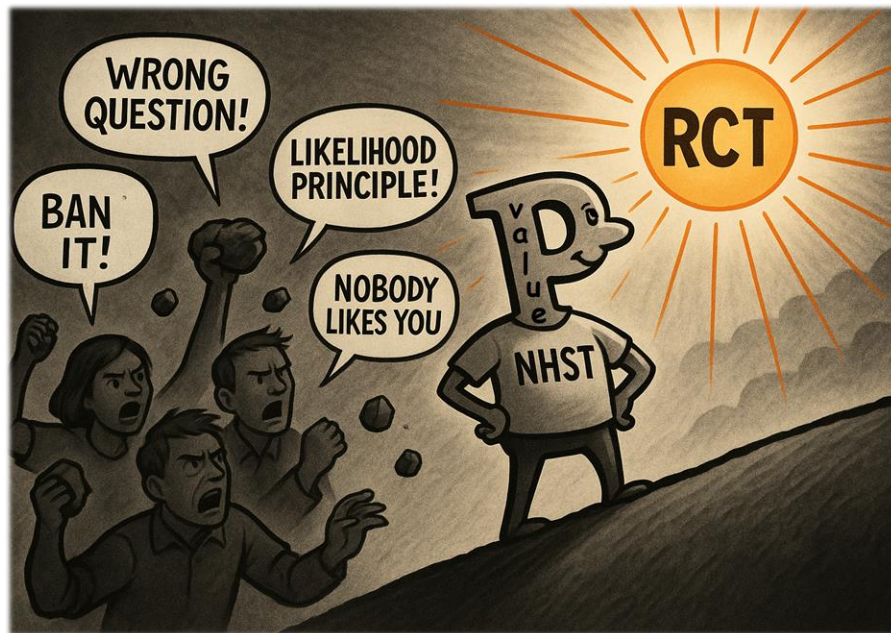
Kilka słów o realiach biostatystyki klinicznej...



- **Trudne** i małoliczne dane
- **Planowanie strategii analizy**
- Plany awaryjne „co jeśli”
- Dziesiątki wytycznych (ICH, FDA, ...)
- Uzasadnienie - na wszystko
- „Pod ostrzałem” – i bardzo dobrze
- „Bio” ma pewne konsekwencje...



Kilka słów o realiach biostatystyki klinicznej...



- P-value odnajduje sens istnienia ☺
- Głównie częstościowo, ale Bayes – OK!
- Istotność dziedzinowa „by design”
- Przyczynowość „by design”
- Badania projektowane dla mocy
- Niedozwolona żonglerka H_0 !
- Analizy czułości – do wszystkiego
- Silna kontrola błędu 1 rodzaju
- Testy 1-stronne – istnieją ☺



Kilka słów o realiach biostatystyki klinicznej...



Agenda

- Krótkie wprowadzenia do rodzajów badań
- Wybrane aspekty projektowania badań
- Plany Awaryjne
- Plan Analiz Statystycznych
- Inne, wybrane wyzwania



Krótkie wprowadzenia do rodzajów badań



- **Interwencyjne**

- Randomizowane i nierandomizowane
- Fazy: 0 (nie przedkliniczna), I, IIa, IIb, III, IV

- **Obserwacyjne**

- Pro- i retrospektywne
- kohortowe, kliniczno-kontrolne, przekrojowe, przeglądy rejestrów

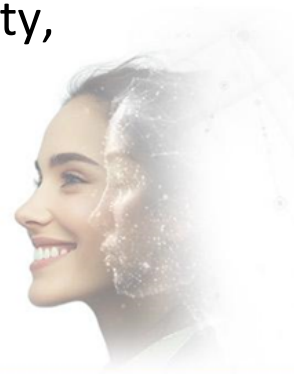
- **Eksploracyjne - rozeznawcze**



Krótkie wprowadzenia do rodzajów badań



- **Konfirmacyjne: *czy?***
 - RCT fazy 3: czy działa i bezpieczny?
 - Pomocnicze, fazy I i II, PK/PD
- **Eksploracyjne: *jak?***
 - Praktyki, potrzeby, schematy,
- **Rozeznawcze: *co/po co?***
 - Co jest warte uwagi? R&D



Wybrane wyzwania w badaniach RCT – od czego zacząć?

Uporządkowanie celów badawczych i przetłumaczenie ich na język hipotez, dobór metod, zaplanowanie prac.

Przeprowadzenie analiz w oparciu w przygotowany scenariusz prac.



Wyzwania w badaniach RCT – projektowanie



- Projektowanie badania:
 - Zrozumienie celu (protokół, research)
 - Uporządkowanie wiedzy - zadanie niezbędnych pytań „technicznych”
 - Opracowanie „kształtu” badania
 - Opracowanie metodyki analizy zgodnie z wytycznymi agencji regulacyjnych, „tradycjami branży” oraz najlepszymi praktykami.

Wyzwania w badaniach RCT – projektowanie: zadanie pytań

- Co będzie celem badania?
określenie skuteczności terapii, eksploracja czynników wpływających na XYZ, poznanie bieżących praktyk medycznych, zmiana wskazania dla leku, farmakokinetyka i farmakodynamika leku, ocena efektów długofalowych, ...
- Czy to badanie eksploracyjne czy confirmacyjne?
- Czy to badanie randomizowane, pragmatyczne, obserwacyjne?
- Jeśli eksploracyjne, to czy retro- czy prospektywne? Przypadek-kontrola? Kohortowe? Post-marketing?
- Skąd pochodzą dane?
eCRF, retrospektywny przegląd rejestrów medycznych, dane gromadzone prospektywnie, agregowane dane z literatury (ramię syntetyczne)



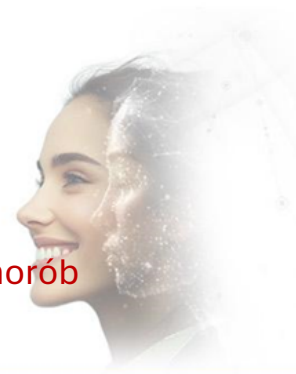
Wyzwania w badaniach RCT – projektowanie: zadanie pytań

- Jaki parametr kliniczny odzwierciedla skuteczność?
wzrost hemoglobiny, spadek LD-CL, spadek ciśnienia skurczowego
- Jaka miara statystyczna najlepiej to opisuje?
średnia, mediana, czas do XYZ, ryzyko, % sukcesów, % przeżycia, RMST, zmiana względna
- Jak będą określone parametry? Kiedy będą mierzone?
w 30 dniu terapii, co 2 godziny od podania, co miesiąc
- Skuteczność będzie określona względem czego?
placebo, aktywnego komparatora, określonej dawki
- Ile ramion ma mieć badanie? Czy ich liczba jest stała?
jednoramienne (farmakokinetyka i –dynamika), dwuramienne, wieloramienne – adaptacyjne



Wyzwania w badaniach RCT – projektowanie: zadanie pytań

- Czy dysponujemy danymi pilotażowymi?
przydatne podczas planowania badania
- Jakie są kryteria włączenia / wyłączenia pacjentów?
jak będzie zdefiniowana grupa badawcza?
- Jak mała zmiana będzie klinicznie istotna?
praktyczna istotność: np. co najmniej 30mg/dL
- Jak będzie użyta kliniczna istotność w badaniu?
non-inferiority, superiority, bio-equivalence, 1- / 2- stronne testowanie
- Czy są znane jakieś czynniki modyfikujące efekt?
wiek, płeć, rasa, palenie, picie alkoholu, zażywanie pewnych leków, występowanie określonych chorób towarzyszących



Wyzwania w badaniach RCT – projektowanie: zadanie pytań

- Co w sytuacji XYZ?
błędy lub brak pomiaru, utrata pacjenta (*lost to follow-up*, naruszone „okno czasowe” dla wizyty, naruszenie protokołu, itd.)
- Czy to jedyny cel badania?
badacze zazwyczaj chcą uzyskać wielowymiarowy obraz działania terapii, by określić pełniejszy profil skuteczności, bezpieczeństwa oraz opłacalności.
- Jak dodatkowo cele będą zorganizowane wg. ważności?
pierwszorzędowe, drugorzędowe, eksploracyjne
- W jakiej kolejności będą weryfikowane (testowane)?
równorzędnie, hierarchicznie - wedle zadanej kolejności
- Jak zapewnimy kompletność danych? Co w razie braków?
dyskusja wpływu czasu trwania badania, środki zaradcze, plany imputacji



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Metoda **randomizacji** (prosta, blokowa, stratyfikowana, adaptacyjna)?
- Jaki będzie **schemat** badania (równoległe, naprzemienne), czy adaptacyjne – co będzie adaptowane, jakie warunki stopu
- Lista **celów** badania – pytań badawczych odpowiadających na pytania: czy lek *działa*, czy lek *jest akceptowalnie bezpieczny*, czy jego stosowanie jest *korzystne dla pacjenta*? + pomocnicze
- **Podział ważności** celów badania: pierwszo- i drugorzędowe, eksploracyjne
- Definicje **punktów końcowych** (endpointów), tj. mierzonych wielkości, scoringu z kwestionariuszy, odsetków zdarzeń, czasu do zdarzenia, itd.
- Określenie **istotności klinicznej** dla wybranych endpointów (*MCID*)
- **Rodzaj badania**: non-equality, non-inferiority, superiority, (bio)equivalence.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- **Non-equality / Superiority** nowa terapia vs. placebo (trywialna różnica: 0/1)

Stosowane dla nowych terapii, nie posiadających jeszcze aktywnego komparatora.

- **Non-inferiority** nowa terapia vs. istniejąca (margines δ)

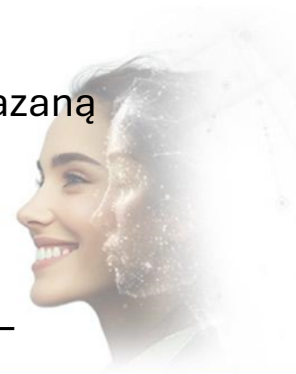
Stosowane, by wykazać „nie-gorszość” względem istniejącego standardu (SoC). Np. gdy nowa terapia jest lepiej tolerowana, tańsza w produkcji, bezpieczniejsza, należy do innej grupy leków (np. NLPZ vs. paracetamol, β -bloker vs. ACE vs. ARB)

- **Clinical superiority** nowa terapia vs. istniejąca (MCID)

Stosowane, by wykazać „lepszość” nowej terapii względem istniejącej o co najmniej wskazaną wartość istotności klinicznej (np. redukcja LDL-C o min. 50mg/dL).

- **(Bio-) equivalence** nowa terapia vs. istniejąca (margines δ)

Stosowane, by wykazać „równoważność biologiczną” nowej terapii względem istniejącej – badanie typowo dla **leków generycznych**.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Określenie metod utrzymania **kompletności** danych dla wybranych endpointów, w tym – jeśli zasadne – **metod imputacji**.
- Imputacja powinna uwzględniać **charakter** endpointu oraz jego **związki** z innymi zmiennymi w badaniu.

Nie są tolerowane metody jednozmiennowe, np. imputacja średnią czy medianą. Warunkowo dopuszczalne są metody typu *OCF/*OCB – szczególnie przez FDA, o ile tylko charakter zmian implikuje konserwatywizm.

Imputacja średnią:

- ⊗ wpływa na wariancję (obniża ją) bez powodu.
- ⊗ zakłada symetryczny rozkład jednomodalny danych (niekoniecznie gaussowski, ale każdy do niego podobny), co automatycznie oznacza, że dane zostały wygenerowane w procesie addytywnym – raczej rzadko o tym wiemy.
- ⊗ jest wrażliwa na wartości odstające, więc imputacja może ją przesunąć w górę (skośność w prawo) lub w dół (skośność w lewo)
- ⊗ w przypadku danych wielowymiarowych całkowicie ignoruje zależność między wymiarami (kontekst)...
- ⊗ ... co oznacza również, że po cichu zakłada wzór MCAR
- ⊗ dla danych kategorycznych, których poziomy generują podgrupy, we wszystkich podgrupach używana będzie ta sama średnia.
- ⊗ imputacja może nie być zgodna z charakterem analizy docelowej (np. poprzez mediany lub inne kwantyle, rangi)
- ⊗ (*uncongeniality*) imputacja jest bardziej restrykcyjna niż model danych kompletnych - więc bardzo możliwy bias

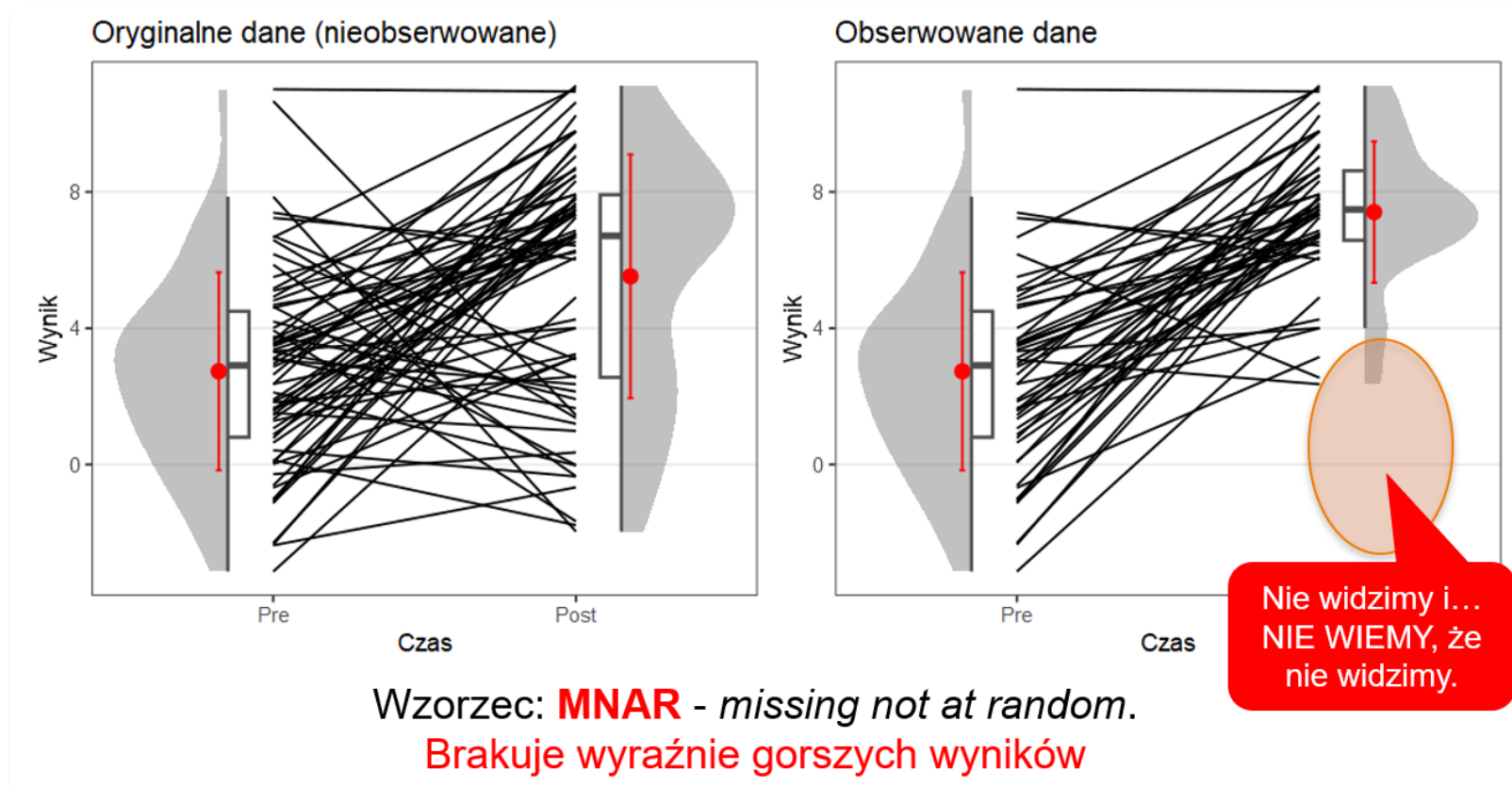


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Zanim jednak cokolwiek zostanie uzupełnione, trzeba spróbować zrozumieć potencjalną naturę braków dla danej zmiennej – tj. **dlaczego może brakować jej wartości?** **To najważniejsze pytanie**, jakie można zadać przy okazji tego zagadnienia!
- Istotne jest również określenie mechanizmu „wypadania” danych:
 - **MCAR** (całkowicie losowe, niezależnie od czegokolwiek) – „najlepsze” 😊 .
Teoretycznie *testowalne* względem MAR (test Little’a, regresja logistyczna, etc).
 - **MAR** (warunkowe dla pewnych kombinacji predyktorów; w ich ramach – całkowicie losowe) – „akceptowalne” dobrze zdefiniowanego (!) modelu regresji lub podwójnie odpornego estymatora.
 - **MNAR** (nielosowe: zależnie od swych wartości lub nieznanymi czynników!).
Przypadek najgorszy, bardzo realny i **praktycznie nie do wykluczenia!**



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- W użyciu są dwa klasyczne standardy:
 - **JM** (*joint modelling*) - określa wielowymiarowy rozkład (dla uproszczenia zwykle normalny) dla wszystkich zmiennych z brakującymi danymi i losuje z niego wartości do wstawienia. **Tylko dane jednego rodzaju.**
 - **FCE** (*fully-conditional specification*), której przykładem jest MICE - imputuje każdą wskazaną zmienną sekwencyjnie, używając jednowymiarowego modelu imputacji (różnymi metodami, np. PMM, regresja logistyczna czy porządkowa, Poissona, itd) zdefiniowanego w oparciu o inne zmienne, jak gdyby w łańcuchu – stąd nazwa „*multiple imputation via chained equations*”. **Dowolny rodzaj danych.**
- oraz ostatnio jeszcze metoda pochodząca z frameworka estymand – **RBMI**: *reference-based multiple imputation*.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- [ODI] will be imputed using the PMM method with the following predictor variables:
Treatment x ([No evidence of motion] + [SF36 PCS] + [ODI baseline] + [ODI pain] + [Age + Smoking + BMI] + Sex)
- [ODI Pain] will be imputed using the PMM method with the following predictor variables:
Treatment x ([ODI Pain baseline] + [No evidence of motion] + [Unintended soft tissue ossification] + [Swelling/edema] + [Age + Smoking + BMI] + Sex)
- [LLQ Pain] will be imputed using the PMM method with the following predictor variables:
Treatment x ([LLQ Pain baseline] + [Unintended soft tissue ossification] + [No evidence of motion] + [Neurological status] + [Age + Smoking + BMI] + Sex)

MICE

LOCF +
NOCB

ID	Timepoints ≤ Month 3	Month 6	Month 12	Month 20
#1	X assigned Absent	Present →	X carried Present	Present
#2	X assigned Absent	X carried Absent	X carried Absent	← Absent
#3	X assigned Absent	X assigned Absent	X assigned Absent	X assigned Absent
X indicates missing information				



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Dodatkowo, dla mechanizmu MNAR wymagane są analizy czułości. Analiza czułości to ocena wpływu zmiany pewnego parametru na stabilność wyników.
- Nie znamy mechanizmu MNAR, ale możemy przetestować kilka typowych mechanizmów → **Pattern-mixture model**
- Np. imputujemy wartości z dodaną (lub przemnożoną) zmienną przebiegającą pewien *rozsądny* zakres wartości. Na takich danych przeprowadzona jest analiza. Sprawdzamy, kiedy jej wynik (p-value, Cohen D, precyzja = $\frac{1}{2}$ CI, znak estymaty) ulegnie zmianie. Nazywa się to „*tipping-point analysis*” (punkt „przejęcia”).
- Odmiana: analiza **best-worst**, daje pogląd dla skrajnych scenariuszy.
- Analiza **non-responder**: brakom danych przypisana jest „porażka”.

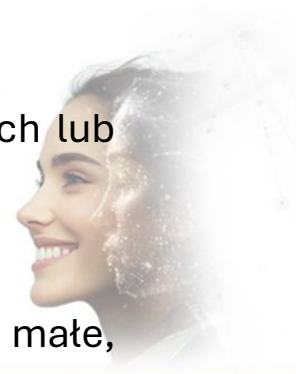


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Braków danych spowodowanych wypadaniem pacjentów z badania **nie sposób uniknąć**, szczególnie w badaniach z dziedziny onkologii, kardiologii, chorób neurodegeneracyjnych, a także w badaniach długotrwałych.

Dzięki systemom **eCRF** zbudowanym wg polityki „*data quality by design*”, z rozbudowanymi mechanizmami walidacji opartymi o zakresy referencyjne, warunki logiczne oraz system tzw. „*data queries*”, można zdecydowanie **ograniczyć braki danych wynikające z niedostatecznej staranności ich gromadzenia**, m.in:

- pominięte (np. zapomniane) informacje - system „flagowania” pól obowiązkowych
- literówki (15.1 vs. 11.5), w tym błędnie postawiony separator dziesiętny (1.23 vs. 12.3),
- niespójność jednostek prowadząca – po ich ujednoliceniu – do znacznie zawyżonych lub zaniżonych wartości (3.25 mmol/L = **125.7 mg/dL** vs. 3.25 mg/dL),
- niespójne „kodowanie” wartości słownikowych (np. ICD9, ICD10, MedDRA, ATC)
- wartości odstające sugerujące pochodzenie „maszynowego” (np. hemoliza krwi)
- wartości budzące jakiegokolwiek wątpliwości, w tym ekstremalnie duże bądź małe, wprowadzone niecelowo (być może *misconduct*) lub – co gorsza – celowo (*fraud*).



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

GoResearch™ live - CT-WD40-2025 20:58 User John Smith English

PATIENTS **SELECTED PATIENT** REPORTS DOCUMENTS IP MANAGEMENT RANDOMIZATION MANAGEMENT SDV MODES CONTACTS

911-003 Randomized

Search

- Observation summary
- Incl./Excl. criteria
- Inclusion/exclusion criteria
- Enrollment visit
- Patient information
ICF date: 2025-02-03
- Laboratory tests and vaccination
- Visit 2
- Visit**
- (S)AE & safety data
- (S)AE & safety data
- (S)AE & safety data
Started: 2025-04-15
- + (S)AE & safety data
- End of observation
- End of observation

Visit date **2025-03-13**

Hemagglutination inhibition geometric mean titers

Examination date **2025-03-15**

A/H3N2 - Test II result [IU/ml] 85.00	A/H3N2 - Test I result [IU/ml] 88.00
A/H1N1 -Test II result [IU/ml] 16.00	A/H1N1 -Test I result [IU/ml] 1888.00
B - Test II result [IU/ml] 21.00	B - Test I result [IU/ml]

Did the subject document any post vaccination side effects in the diary? **Yes** **No**

Did any serious adverse events or non-serious not documented in the diary occur? **Yes** **No**

Last modification by: John Smith, 2025-04-29 12:53:03

DELETE **CANCEL** **SAVE**

Data validation

Search

Visit date (Visit in 21st day)

The date should be later than or equal to examination date.

A/H1N1 -Test I result [IU/ml] (Hemagglutination inhibition geometric mean titers)

Value should be a decimal value greater or equ...

B - Test I result [IU/ml] (Hemagglutination inhibition geometric mean titers)

Field is required.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

GoResearch™ live - CT-WD40-2025 20:58 User John Smith English

PATIENTS **SELECTED PATIENT** STATUS REPORT DOCUMENTS IP MANAGEMENT CONTACTS

911-003 Randomised

Search

- Observation summary
- Incl./Excl. criteria
- Inclusion/exclusion criteria
- Enrolment visit
- Patient information
- Laboratory tests and vaccination
- Visit 2
- Visit
- (S)AE & safety data
- (S)AE & safety data
- (S)AE & safety data
- (S)AE & safety data
- End of observation
- End of observation

Event details

Headache

Event details

Event start date **2025-04-15** full date

Event end date **2025-04-16** ongoing ☐

Severity **Mild** **Moderate** Severe Life threatening Fatal Unknown

Seriousness **Not serious** Serious

Result **Regressed without consequence**

Vaccination related **Suspected** **Not suspected**

Last modification by: John Smith, 2025-04-29 12:53:03

DELETE **CANCEL** **SAVE**

Queries

Search

DM (Elen Clark) 2025-04-29 13:40:11
AE end date

CRA (Jack King) 2025-04-29 13:45:20
Missing laboratory test result

CRA (Jack King) 2025-04-29 14:04:23
Missing PT and aPTT test result

CRA (Jack King) 2025-04-29 14:04:23
There is no PT result documented. Please provide missing information.

Investigator (Jane Kovalsky) 2025-04-29 14:04:48
Information provided according to source documentation.

CRA (Jack King) 2025-04-29 14:06:24
Missing results provided by Investigator.

DM (Elen Clark) 2025-04-29 14:01:26
Ongoing infection

+ ADD QUERY



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- **Przetłumaczenie celów na hipotezy** statystyczne wykorzystując składowe:
 - ✓ endpoint (np. stężenie cholesterolu LDL po 5 miesiącach od rozp. terapii)
 - ✓ miara opisująca (średnia) i miara zmiany (różnica średnich)
 - ✓ wybór wartości względem której zmiana będzie porównana
(0, 0.5, 1; istotność kliniczna – MCID; margines non-inferiority – δ ; minimalna wykrywalna wartość – MDC)
 - ✓ Rodzaj porównania: *non-equality*, *non-inferiority*, (clinical) *superiority*, *equivalence*
np. $H_0: m_{CONTROL} - m_{ACTIVE} \leq -\delta$

Wynika stąd również **natura testowania** (dot. testów i przedziałów ufności):

- **2-stronne** dla badań *equality*, *superiority*
- **1-stronne** dla badań *non-inferiority*, *clinical superiority*, *equivalence* (TOST)
<https://www.onesided.org>

- **Dobór testu statystycznego** bądź **wybór modelu statystycznego** oraz **metody estymacji** (szczególnie dla badań longitudinalnych). Określenie jej **parametrów**.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Ale uwaga na ryzyko „**technicznie poprawnej odpowiedzi na niezadane pytanie**”:

- **wybór endpointu:** anemia: żelazo czy hemoglobina? NSCLC: *overall survival* czy *progression-free survival*?
- **wybór miary:** średnie, mediany, stochastyczna dominacja (rangi), Odds Ratio, Risk Reduction, Hazard Ratio vs. różnica RMST vs. różnica median, itd., itp.
- **wybór statystyki testowej:** zmiana H_0 to zmiana pytania!

np. Mann-Whitney testuje stochastyczną dominację, która nie ma wiele wspólnego z równością median lub średnich, o ile dane nie są IID + symetryczne, zareaguje na przesunięcie, dyspersję, kształt (skośność) – mogąc całkowicie pogmatwać interpretację.

- **transformacje**, np. Boxa-Coxa (interpretacja, średnia-wariancja, bias $\leftarrow$ Jensen)



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



Prawda może mieć wiele prawdziwych obrazów, zależnie od perspektywy.

...ale na coś trzeba się zdecydować i tego trzymać.

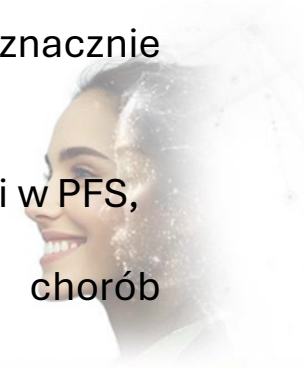


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Rozważmy punkt końcowy **Progression-Free Survival (PFS)**. Jest on często preferowany w stosunku do „złotego standardu” - **Overall Survival (OS)** ze względu na **krótszy okres obserwacji**, co prowadzi do obniżenia kosztów i szybszego podejmowania decyzji.

Ale PFS może słabo korelować z OS, ponieważ:

- badane leczenie może opóźnić postęp choroby (poprawiając PFS), ale nie wydłużyć ogólnego okresu życia ze względu na swoją toksyczność (onkologia!),
- jeśli po progresji choroby pacjenci otrzymują wysoce skuteczne leczenie, może to znacznie wydłużyć OS, zmniejszając w ten sposób wpływ ocenianej terapii ($PFS \ll OS$),
- w przypadku agresywnych nowotworów OS może być mocno skrócony pomimo korzyści w PFS,
- nie wspominając o śmiertelności z przyczyn innych niż nowotwory (z powodu chorób współistniejących) i innych przyczynach dodatkowo dekorelujących PFS i OS...



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

A jest tego znacznie, znacznie więcej!

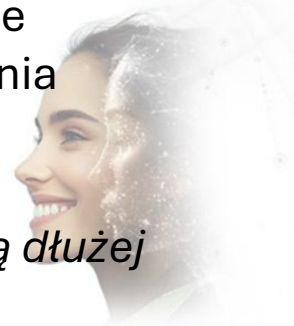
- ✓ Całkowite przeżycie (OS),
- ✓ Przeżycie bez progresji (PFS),
- ✓ Czas do progresji (TTP),
- ✓ Przeżycie bez choroby (DFS),
- ✓ Przeżycie bez zdarzeń (EFS),
- ✓ Czas do niepowodzenia leczenia (TTTF),
- ✓ Czas do następnego leczenia (TTNT),
- ✓ Czas do odpowiedzi (TTR)
- ✓ Czas trwania odpowiedzi,
- ✓ Czas do pierwszego nawrotu (lub remisji),
- ✓ ...



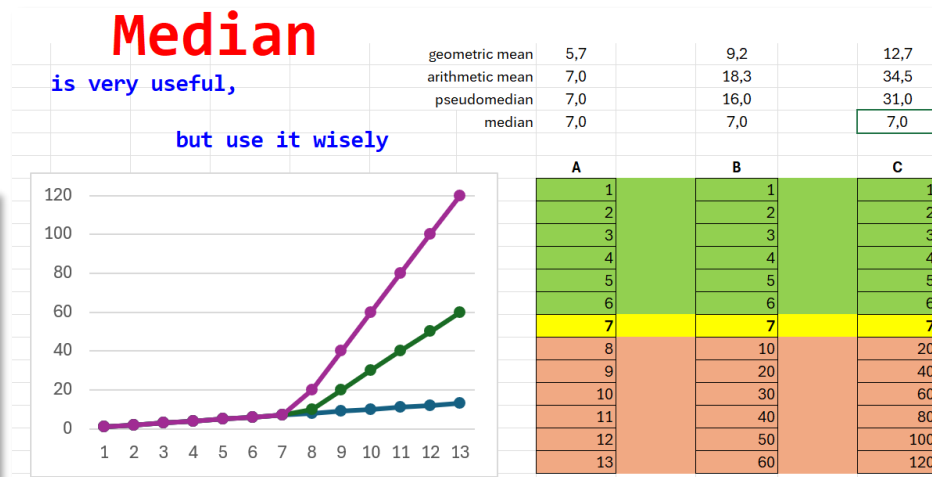
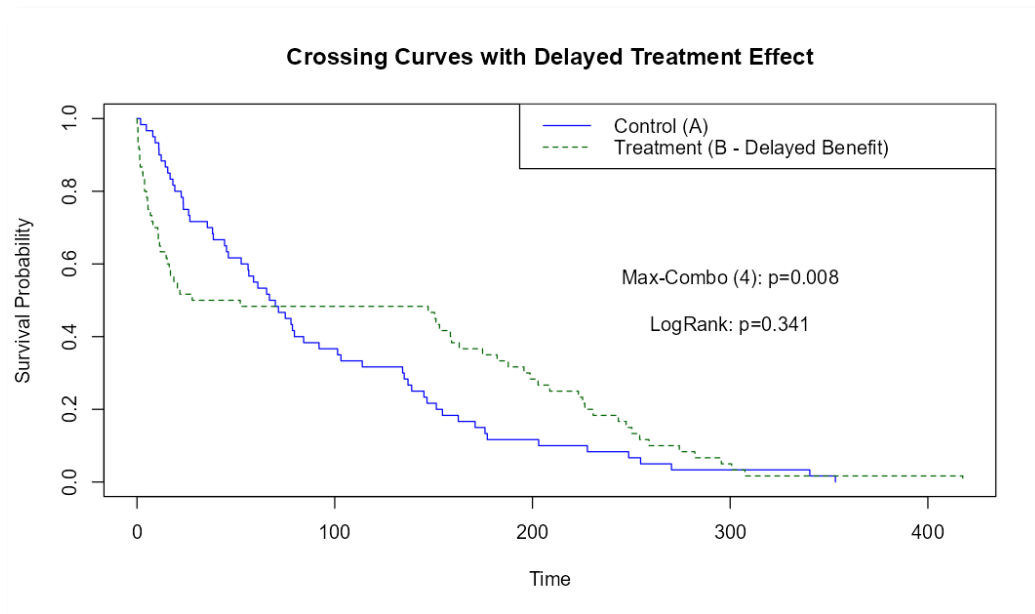
Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Dodajmy zagadnienia takie, jak:

- konkurujące ryzyka (czy cenzurować śmierć i in.),
- obecność zdarzeń terminalnych,
- nawroty i zmiany w leczeniu,
- efekt zmienny w czasie (nieproporcjonalne hazardy)
- „informacyjne cenzurowanie” ($\approx$ MNAR)
- **wybór momentu obserwacji** lub pomiaru $\rightarrow$ pominięcie efektu poprzez **zbyt wczesne** badanie, gdy efekt jeszcze nie wystąpił (np. w onkologii, gdy leczenie może potrzebować trochę czasu, aby się „rozgrzać”), lub **zbyt późne**, gdy interesujące zdarzenie już nastąpiło (np. dynamicznie zmieniające się markery zawału mięśnia sercowego w kardiologii).
- i odpowiedź na pozornie proste pytanie: „Czy leczenie sprawia, że pacjenci żyją dłużej lub lepiej?” staje się jeszcze trudniejsza.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

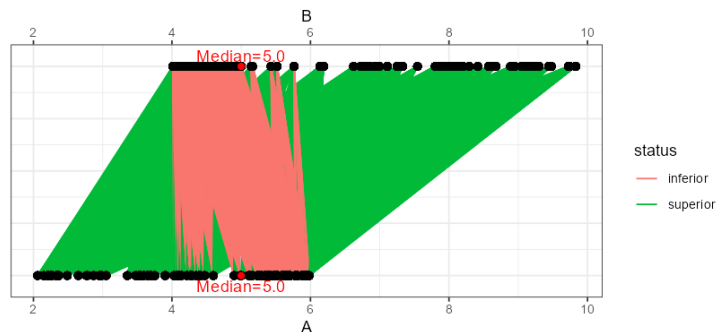


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Stochastic superiority: $P(B>A) = 69.8\%$

Δ medians = 0.000, median(Δ) = 1.461, 2-sample pseudomedian = -1.429

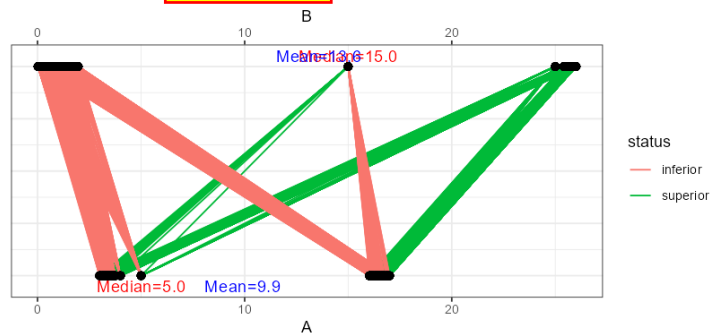
Mann-Whitney: $p=0.0000012$ Quantile regression: $p>0.999$, Mood-Brown: $p>0.999$
mean(Δ) = 1.582, Welch t: $p<0.001$



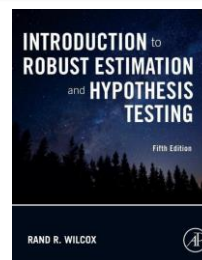
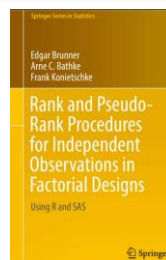
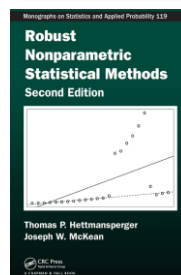
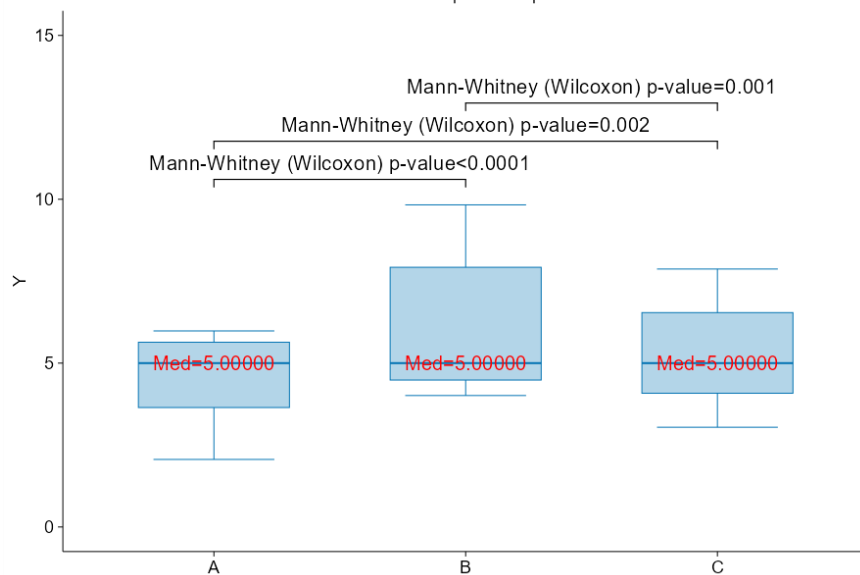
Stochastic superiority: $P(B>A) = 50.0\%$

Δ medians = -10.000, median(Δ) = 8.978, 2-sample pseudomedian = -8.005

Mann-Whitney: $p>0.999$ Quantile regression: $p=0.017$, Mood-Brown: $p=0.888$
mean(Δ) = 3.674, Welch t: $p=0.009$



Kruskal-Wallis p -value = $p<0.001$



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Mamy ponad **850 testów** statystycznych (liczone w 2023) w tym:

- Metody resamplingowe (permutacyjne, bootstrap) - regresja liniowa, n-way ANOVA, Welch / Yuen t
- Metody odporne na outliery (np. „huberyzowane), heteroscedastyczność (znane od lat 70 XX wieku...)
- Metody nieparametryczne dla wielu czynników i powtarzanych pomiarów (ATS, WTS, ART-ANOVA)
- Testowanie oparte o modele statystyczne dla KAŻDEGO rodzaju danych, do testowania dowolnej liczby czynników, adjustacji o zmienne numeryczne, ich interakcji, transformacji – także dla powtarzanych pomiarów i z efektami losowymi – metody Walda, LRT, Rao
- Modele nieparametryczne oparte o rangi (porządkowa regresja logistyczna, regresja rangowa) i kwantyle (regresja kwantylowa), z efektami losowymi
- Semi-parametryczna metoda estymacji dla całej rodziny GLM → GEE (z wieloma rozszerzeniami)
- Nowoczesne metody imputacji (MICE, RBMI, joint modeling),
- Nowoczesne metody adjustacji dla wielokrotnych porównań
- Nowoczesne metody analizy przeżycia: *recurrent events* + *competing risks*: *joint frailty*, *multi-state*, nieparametryczne estymatory hazardu

i wiele innych – każdego miesiąca pojawiają się nowe metody!



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



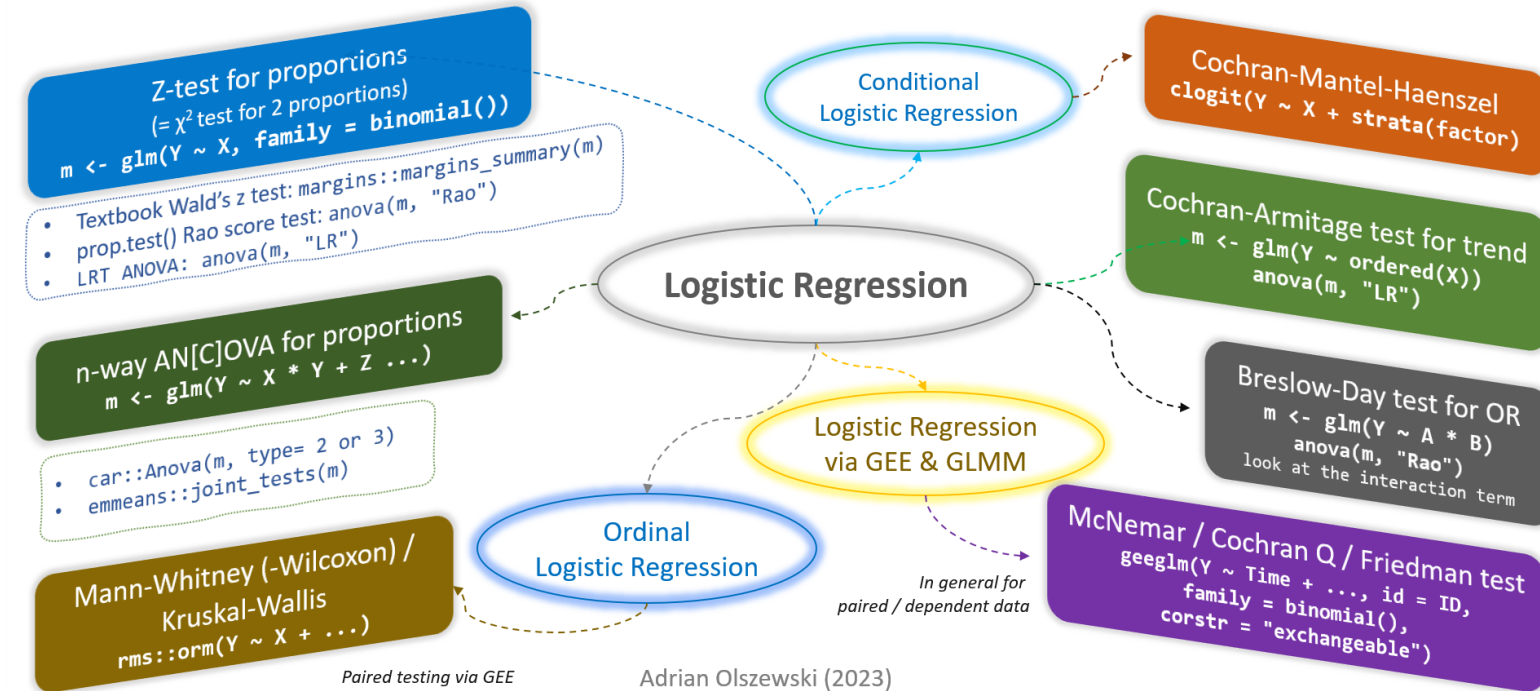
Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- W przypadku testowania hipotez opartego o model statystyczny pozostaje kwestia **zdefiniowania jego postaci**. Typowo, dla badań longitudinalnych jest to model **MMRM** (estymacja GLS lub GEE, szczególnie dla GLM) w postaci:
$$\text{response} \sim \text{arm} * \text{time} + \text{response_baseline} + \text{response_baseline} : \text{time} + \text{time} * (...)$$
- Bardziej złożone modele jak longitudinalne, wymagają ustalenia **dodatkowych parametrów**, np. struktury korelacji resztowej, metody obliczenia stopni swobody (*Kenward-Roger, Satterthwaite*), czy też wyboru estymatorów kowariancji dla małych prób (*Kenward-Roger, Mancl-De Rouen, Pustejovsky*), szczególnie dla estymacji GEE i GLS.
- Dla danych jakościowych (np. binarnych) też są „dodatki” 😊 Np. przy niewielkim spodziewanym % stosuje się dodatkową penalizację Firtha lub regresję dokładną.
- **A to zaledwie kilka aspektów definicji modelu, z których każdy musi zostać nie tylko zaadresowany ale i odpowiednio uzasadniony.**



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Testing hypotheses about binary outcomes (proportions)
with the **Logistic Regression** & selected friends



Adrian Olszewski (2023)



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

https://github.com/adrianolszewski/Logistic-regression-is-regression/blob/main/logistic_reg

Logistic-regression-is-regression / logistic_regression_Rao_Wald_z_test_proportions.md

Preview Code Blame 480 lines (406 loc) · 19.4 KB

Raw Copy Download Edit

$$R = U(\beta_0, \beta_1 = 0) \cdot I(\beta_0, \beta_1 = 0)^{-1} \cdot U(\beta_0, \beta_1 = 0) \quad (40)$$

But since the first element of U is 0, this reduces to just scalar operation: $U^2 / I = U^2 \Sigma$:

$$\begin{aligned} R &= U(\beta_0, \beta_1 = 0) \Sigma(\beta_0, \beta_1 = 0) = \\ &= [n_B(p_B - p)]^2 \cdot \frac{n_A + n_B}{pq n_A n_B} = \frac{[n_B(p_B - p)]^2}{pq \frac{n_A n_B}{n_A + n_B}} \\ &= \frac{n_B^2 \left(p_B - \frac{n_A p_A + n_B p_B}{n_A + n_B} \right)^2}{pq \frac{n_A n_B}{n_A + n_B}} = \frac{n_B^2 \left(\frac{p_B n_A + p_B n_B - n_A p_A - n_B p_B}{n_A + n_B} \right)^2}{pq \frac{n_A n_B}{n_A + n_B}} \\ &= \frac{\frac{n_B^2 (p_B n_A - n_A p_A)^2}{(n_A + n_B)^2}}{pq \frac{n_A n_B}{n_A + n_B}} = \frac{\frac{n_B^2 n_A^2 (p_B - p_A)^2}{(n_A + n_B)^2}}{pq \frac{n_A n_B}{n_A + n_B}} \quad (41) \\ &= \frac{(p_B - p_A)^2}{pq \frac{n_A n_B}{n_A + n_B} \frac{(n_A + n_B)^2}{n_B^2 n_A^2}} = \frac{(p_B - p_A)^2}{pq \frac{n_A + n_B}{n_A n_B}} = \frac{(p_B - p_A)^2}{pq \left(\frac{1}{n_B} + \frac{1}{n_A} \right)} \\ &= \frac{(p_B - p_A)^2}{p(1-p) \left(\frac{1}{n_A} + \frac{1}{n_B} \right)} = z^2 \blacksquare \end{aligned}$$

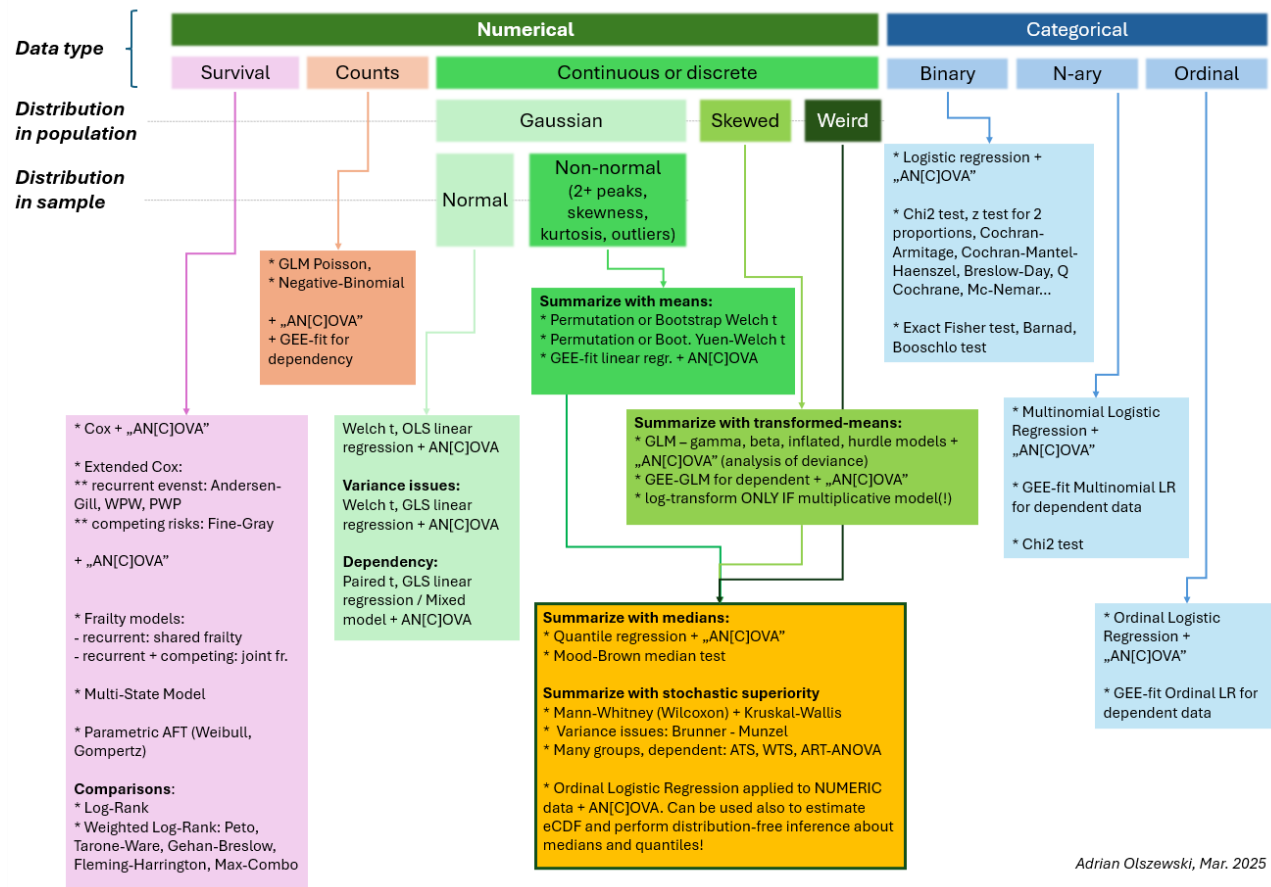
This way I have proven the equivalence of the 2-sample Wald's z-statistic for comparing proportions with pooled variances and the Rao score test over the logistic regression with a single binary predictor distinguishing the compared samples.

```
> wald_z_test_pooled(x1 = 6, n1 = 20, x2 = 10, n2 = 20)
      diff      z      chi2      se p.value p.value_1      LCI      HCI
1 -0.2 -1.290994 1.666667 0.1549193 0.1967056 0.0983528 -0.5036363 0.1036363
```



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Which statistical test to choose?

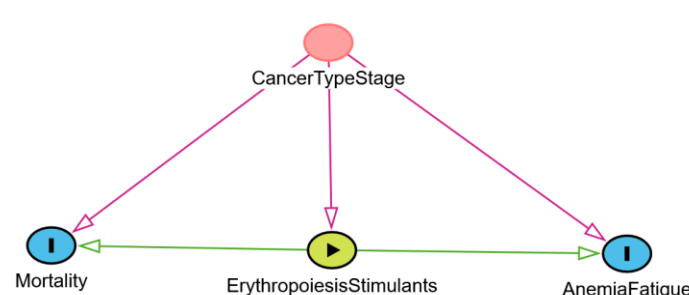


Adrian Olszewski, Mar. 2025

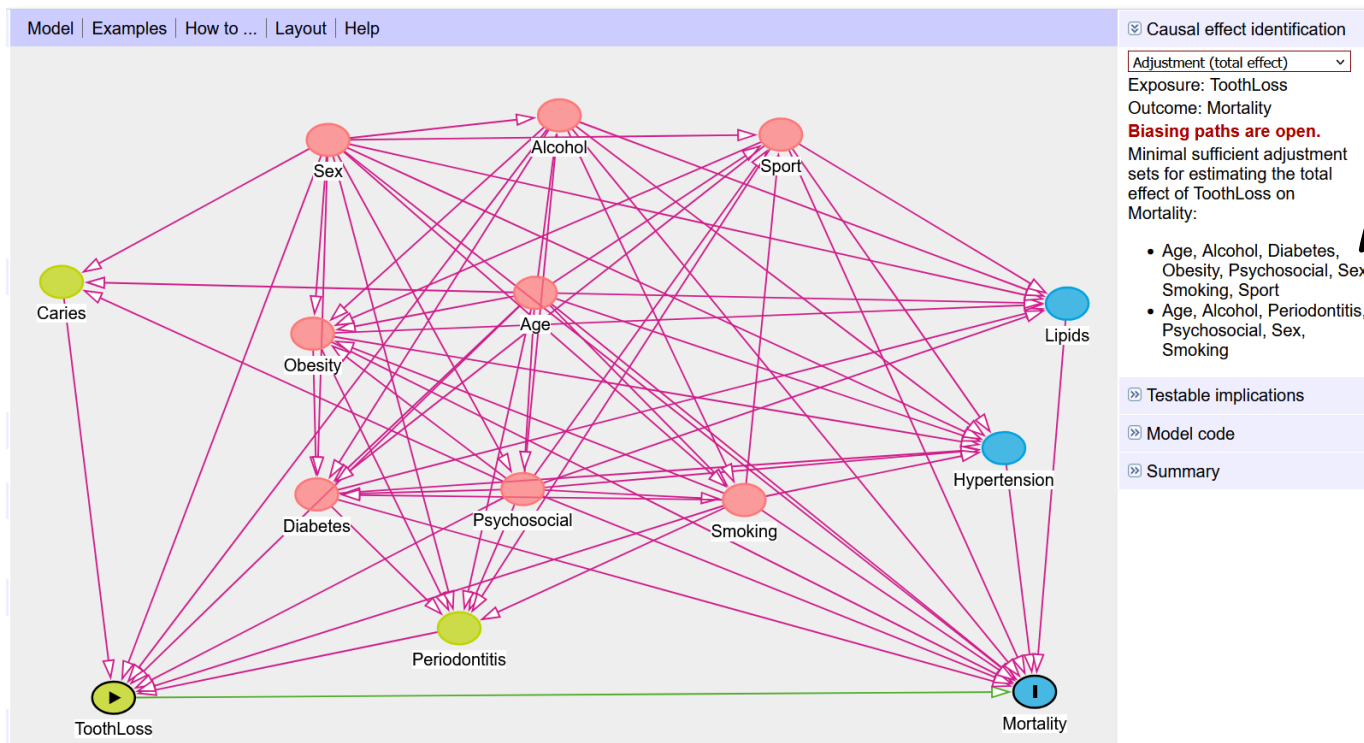


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Często, pomimo randomizacji w badaniach RCT (NIE gwarantuje braku różnic między ramionami przed interwencją, co zwiększa wariancję i obniża moc – stąd potrzeba ANCOVA i MMRM) a także w badaniach nie-randomizowanych zachodzi potrzeba umieszczenia w modelu dodatkowych zmienne (w tym – zmiennych w czasie).
- Trzeba jednak uważać na ich rolę w ujęciu przyczynowości (*causal framework*): *confounder*, *mediator*, *collider*. Inaczej możemy niechcący: **usunąć efekt** (mediator: współliniowość → „przejęcie efektu interwencji”), doprowadzić do **paradoksu Simpsona** (confounder → odwrócenie znaku efektu), lub **Berksona** przez **pozorne korelacje** (collider).
- Niezbędna jest tu współpraca z badaczami i ew. zewnętrznymi specjalistami dziedzinowymi.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



„Podejście wielokrokowe to załatwi!”

Automatycznie dobierze zmienne”

Nie sądzę 😊

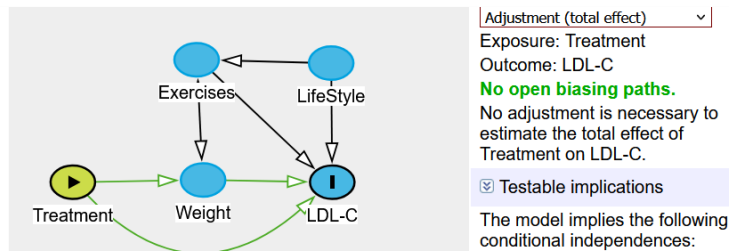
¬_(ツ)_/

Polzer, I., Schwahn, C., Völzke, H. et al. The association of tooth loss with all-cause and circulatory mortality. Is there a benefit of replaced teeth? A systematic review and meta-analysis. Clin Oral Invest 16, 333–351 (2012). <https://doi.org/10.1007/s00784-011-0625-9>

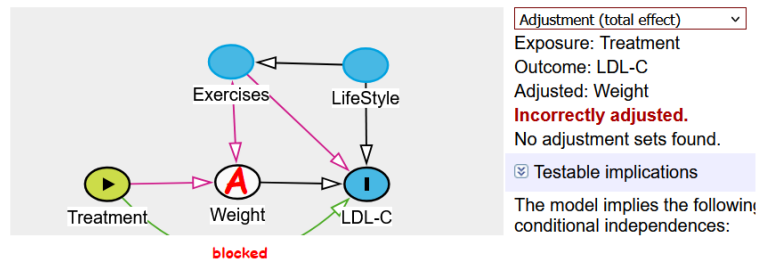


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

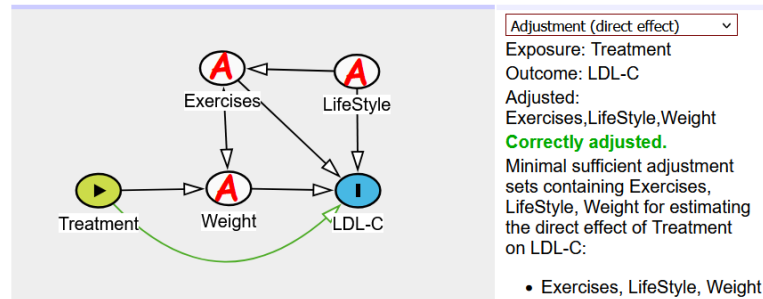
Total effect
of TRT -> LDL



Total effect
of TRT -> LDL

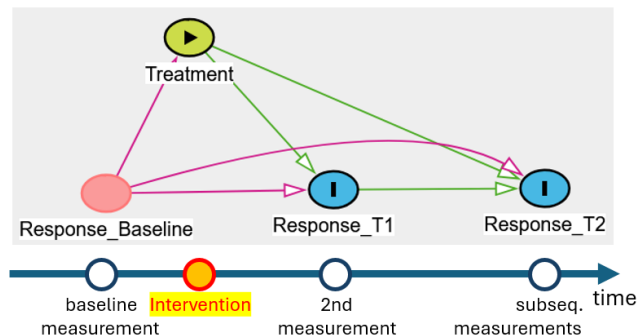


Direct effect
of TRT -> LDL

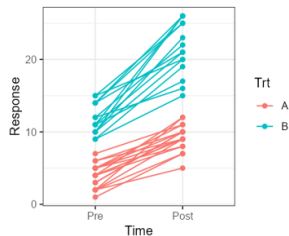


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Non-randomized interventional study



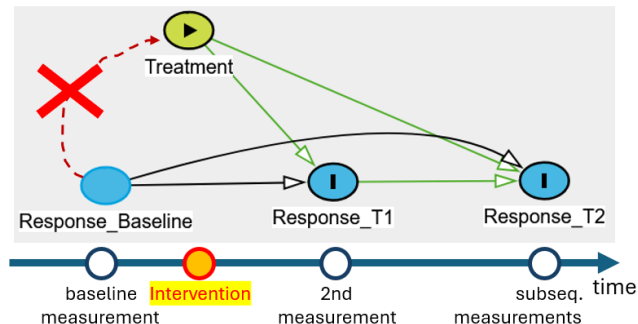
- **Baseline may correlate with subsequent measurements** (e.g. better initial condition → better responses or opposite)
- **Baseline may determine the assignment to treatment** (e.g. worse condition → treatment A).
- **Then it's impossible to attribute the effects solely to treatment**



Adjust for baseline to block the baseline confounding

- **Baseline x Time**
- **Baseline x Treatment** (different rates may occur)

Randomized interventional study



- **Baseline may correlate with subsequent measurements + RTM** (e.g. better initial condition → better responses or opposite)
- **Randomization blocks the Baseline → Treatment path**
- But the remaining post-randomization imbalance of the response will always exist despite sampling from the same, virtual population, which will worsen the power by offsetting groups from apart. This is purely an artifact but it truly “dilutes”, „blurs” the effect. This artifact needs to be addressed in advance (= without any baseline „testing”!)

Adjust for baseline to wipe out the remainder (it's just a technical trick improving the power, not a causal tool)

- **Baseline x Time** – may still happen (is common)
- **Baseline x Treatment** – handled by randomization



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



Jedno jajko dziennie „OBNIŻA ryzyko cukrzycy typu 2”: kontrowersyjne badanie mówi, że promuje kwasy tłuszczowe, które chronią przed chorobą

Mia De Graaf Redaktor zdrowia dla Dailymail.com
21:39 03 stycznia 2019, aktualizacja 22:23 03 stycznia 2019



www.wiocha.pl



Jedzenie tylko JEDNEGO jajka dziennie zwiększa ryzyko cukrzycy o 60 procent, ostrzegają badania

Jonathan Chadwick dla Mailonline
17:47 16 lis 2020, aktualizacja 15:01 19 lis 2020



...a jeśli
zignorujemy
złożoność
zagadnienia

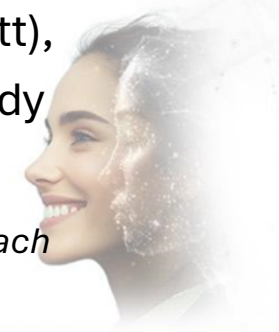


Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Określenie maksymalnego akceptowalnego **poziomu błędu 1-szego rodzaju** oraz **metod jego kontroli** na poziomie badania i „lokalnych” analiz.

(charakterystyka częstościowa może być wymagana także dla badań Bayesowskich, szczególnie, jeśli raportowane do FDA: „*FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*”, rozdział 4.8 *Assessing the operating characteristics of a Bayesian design*, oraz 7.2 *Simulations to Obtain Operating Characteristics*)

- Typowe metody:
 - hipotezy niepowiązane: **Holm** ew. **Hochberg i Hommel**, ale wymagają dodatkowego założenia
 - testowanie Walda oparte o model: **MVT** (→ Tukey-Cramer HSD, Dunnett),
 - hipotezy uporządkowane w hierarchię: **metoda grafowa** (CTP) → metody sekwencyjne: *fixed-sequence, fallback, serial / parallel gatekeeping*
 - rodzina hipotez: *łączone testowanie* (joint testing) – brak informacji o szczegółach
 - pojedyncza hipoteza oparta o *endpoint złożony* (*ale uwaga!*)



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Dla testowania hierarchii hipotez w oparciu o metody sekwencyjne, należy określić **porządek testowania hipotez** w oparciu o przesłanki kliniczne, np.:

non-inferiority @ m-c 20 → superiority @ m-c 20 → non-inferiority @ m-c 10 → ...

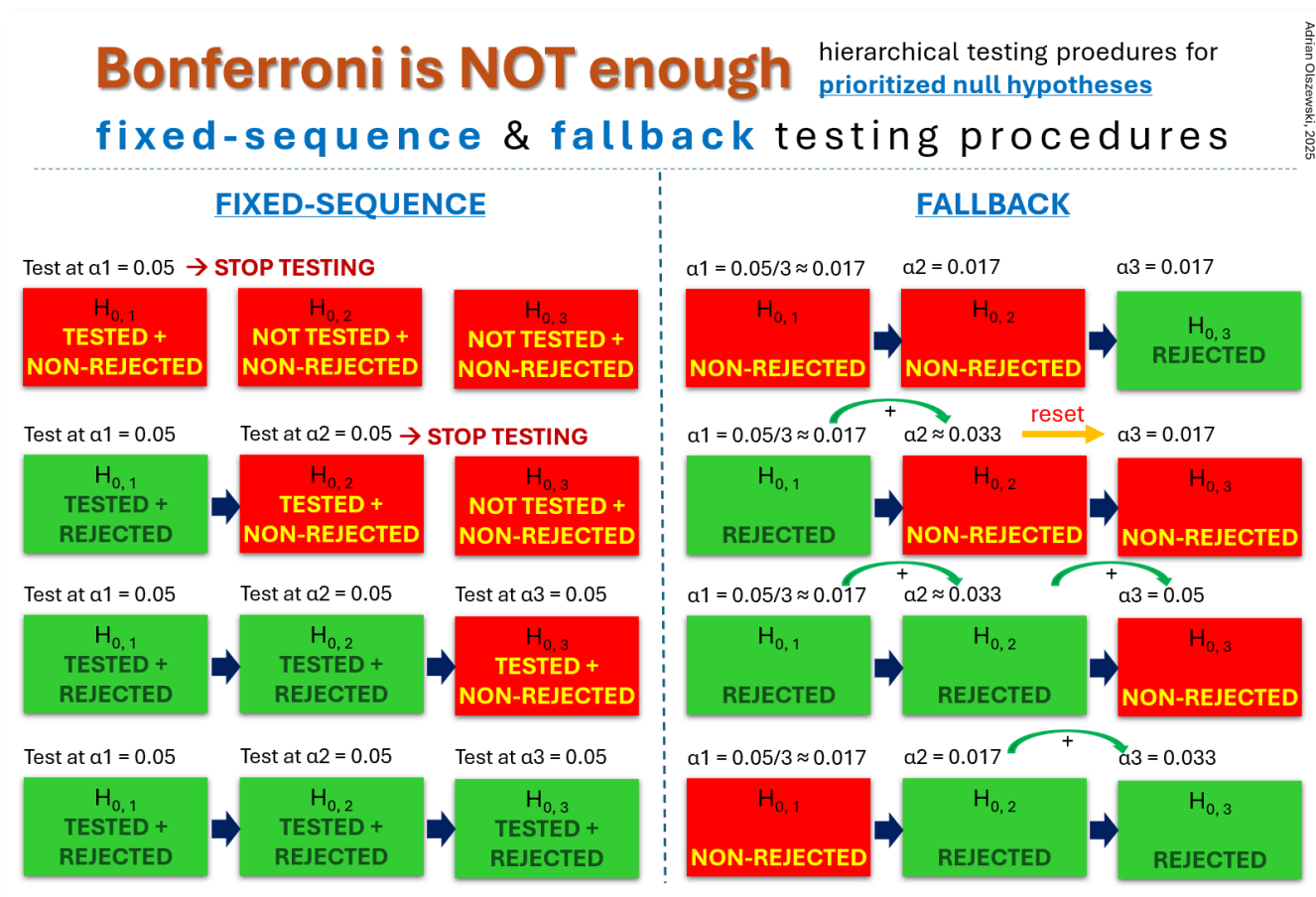
non-inferiority utraty krwi → non-inferiority czasu operacji → superiority % powikłań

👍 Znacznie mniejsza „kara” na poziom istotności (nawet **0!**) dla wielu hipotez i **ogromna elastyczność** (podejście grafowe umożliwia nawet testowanie cykliczne).

👎 Interpretacja w takiej kolejności, w jakiej były testowane. Czasem ma to sens, czasem żadnego. Stosowane bez umiaru może powiększyć kryzys replikacji.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

1.3 Example and diving in

Let's start with a well-known procedure and see how it fits into this graphical approach to weighted multiple test procedures: The Bonferroni-Holm-Procedure from [Holm \[1979\]](#).

Theorem 1.3.1 (Bonferroni-Holm-Procedure). *Let $T_1, \dots, T_m$ be test statistics for $m \in \mathbb{N}$ null hypotheses $H_1, \dots, H_m$ and $p_1, \dots, p_m$ the associated p-values. Then the following test will control the familywise error rate at level $\alpha \in]0, 1[$ in the strong sense:*

Denote the ordered p-values by $p^{(1)} < p^{(2)} < \dots < p^{(m)}$ and the corresponding hypotheses by $H^{(1)}, H^{(2)}, \dots, H^{(m)}$. Reject $H^{(1)}, H^{(2)}, \dots, H^{(j)}$ such that

$$p^{(i)} \leq \frac{\alpha}{n - i + 1} \quad \text{for all } 1 \leq i \leq j.$$

The corresponding graph for the Bonferroni-Holm-Procedure for three hypotheses is given in Figure 1.1. We see a fully connected graph, where each node represents a hypothesis and the nodes and edges have weights.

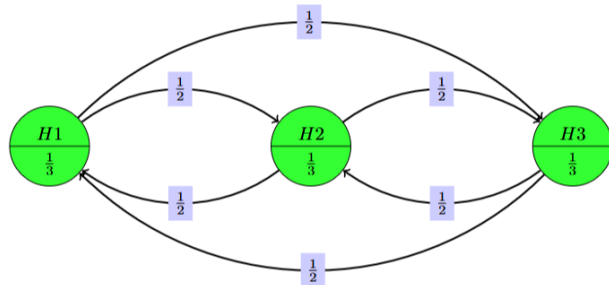


Figure 5 - Graph representing the hierarchy of primary and secondary hypotheses

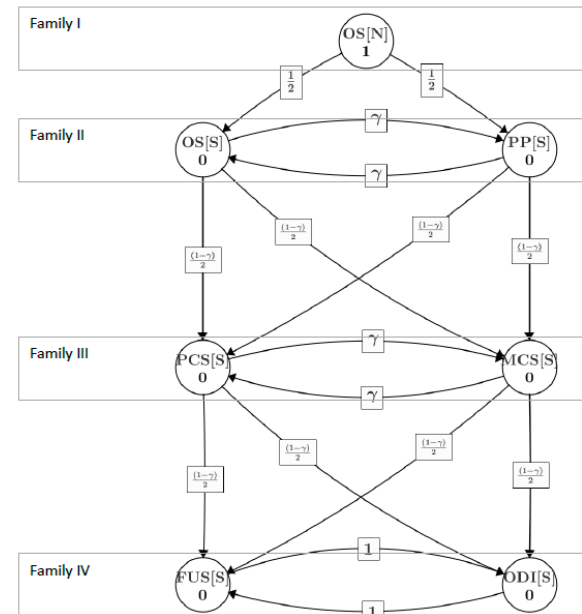


Table 3 - Matrix representing the graph of hypotheses hierarchy in the trial

H ₀	Fam. I OS [N]	Fam. II OS [S]	Fam. II PP [S]	Fam. III PCS [S]	Fam. III MCS [S]	Fam. IV FUS [S]	Fam. IV ODI [S]
OS [N]	-	1/2	1/2	0	0	0	0
OS [S]	0	-	γ	(1-γ)/2	(1-γ)/2	0	0
PP [S]	0	γ	-	(1-γ)/2	(1-γ)/2	0	0
PCS [S]	0	0	0	-	γ	(1-γ)/2	(1-γ)/2
MCS [S]	0	0	0	γ	-	(1-γ)/2	(1-γ)/2
FUS [S]	0	0	0	0	0	-	1



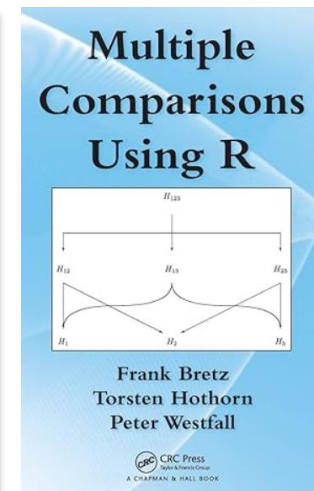
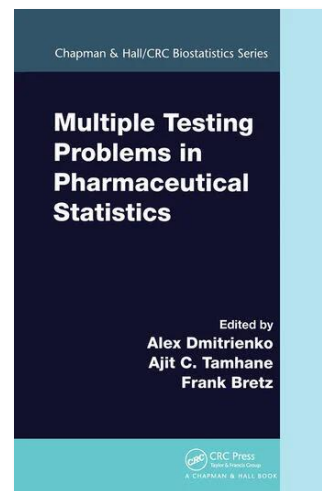
Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Multiple Endpoints in Clinical Trials

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

October 2022
Biostatistics



APPENDIX: STATISTICAL METHODS 18

1. The Bonferroni Method.....	18
2. The Holm Procedure.....	18
3. The Hochberg Procedure.....	19
4. Prospective Alpha Allocation Scheme	20
5. The Fixed-Sequence Method.....	20
6. Resampling-Based, Multiple-Testing Procedures.....	21
7. Gatekeeping Testing Strategies.....	21
8. Graphical Approaches Based on Sequentially Rejective Tests.....	23



● ● ●



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

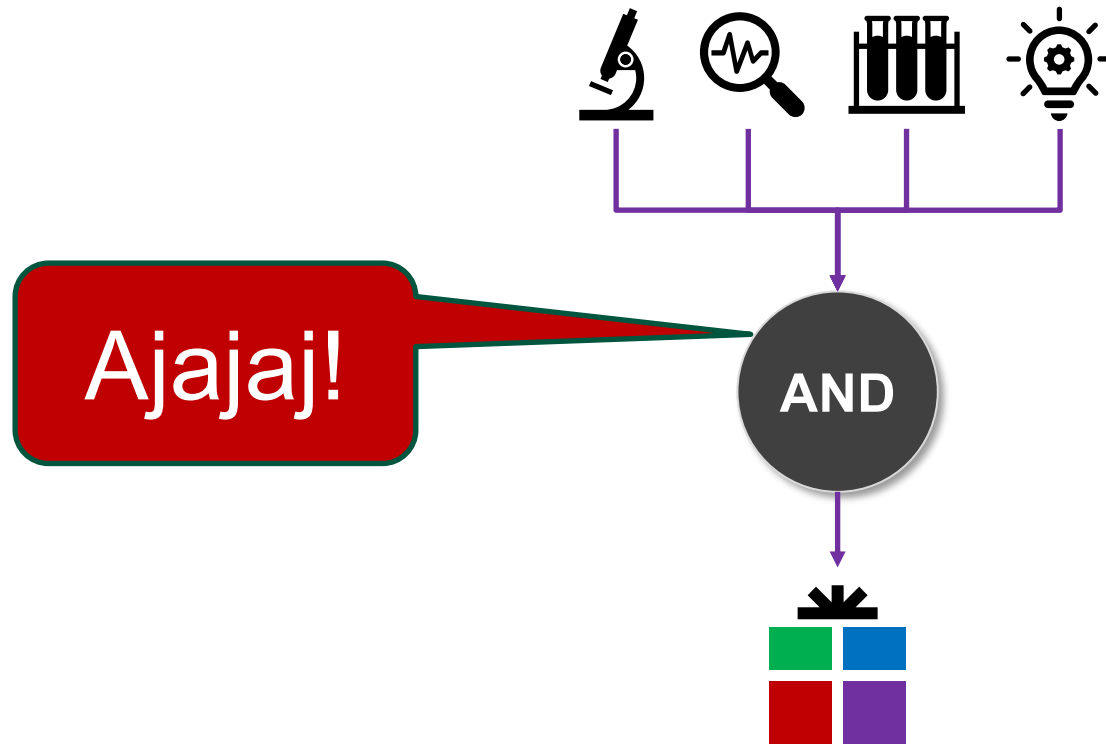
Jedną z metod ograniczenia liczby celów badawczych i punktów końcowych jest ich **powiązanie za pomocą operatora logicznej koniunkcji** („oraz”) w jeden złożony obiekt.

Takie podejście upraszcza testowanie hipotez, ale jest tylko „agregującym” zabiegiem technicznym – trikiem - i wiąże się:

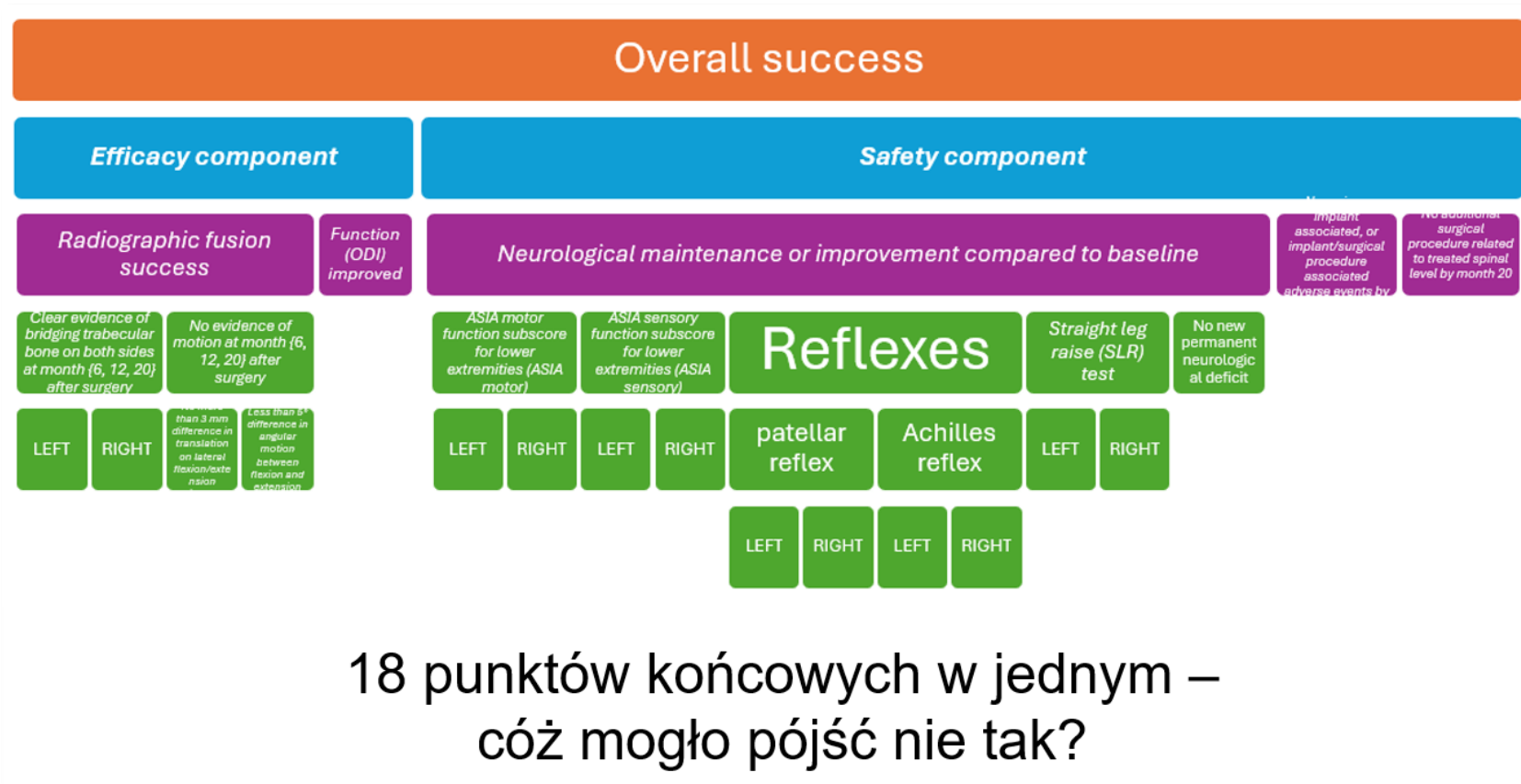
- ☞ z ryzykiem braku wewnętrznej spójności
- ☞ ze znacznym ryzykiem porażki, ponieważ jeśli **jeden tylko komponent w łańcuchu $X_1 \wedge X_2 \wedge \dots \wedge X_n$, będzie logicznym fałszem, to wartość logiczna całego wyrażenia okaże się fałszem.**



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Parameter	Value
Significance level (α)	0.05
Type-2 error (β)	0.10
Power ($1-\beta$)	0.90 (90%)
% success in the reference group (p_1)	0.30 (30%)
% success in the experimental group (p_2)	0.48 (48%)
Absolute difference in % ($\epsilon = p_1 - p_2$)	-0.18 (-18 %points)
Non-inferiority limit (δ)	0.10 (10 %points)
Allocation ratio (κ)	1:1 ($\rightarrow N_1 = N_2$)

Table 14.2.1b Overall clinical success rates [imputed mITT, PP]

Overall success rate / Visit	Dataset	Overall success rate / Visit						Difference (CI*)
		N _{TOT}	N	%*	N _{TOT}	N	%*	
Month 20	mITT-i	55	-	1.9%	56	-	0.2%	0.00 (-0.01, 0.01)
	PP	39	1	2.6%	38	0	0.0%	
Month 12	mITT-i	55	-	2.0%	56	-	3.7%	0.00 (-0.01, 0.01)
Month 6	mITT-i	55	-	0.0%	56	-	1.8%	0.00 (-0.01, 0.01)

* %s, CIs and p-values are pooled in the mITT-i dataset; mITT-i = imputed mITT
CI are 1-sided 95% CI



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- Dla nowych badań konieczne może być zapisanie powyższych ustaleń w nowym podejściu, tzw. „**estimand framework**”. Temat zasługuje na osobny wykład.
- ...co oznacza dodatkowo konieczność opracowania **polityki traktowania różnych zdarzeń występujących w trakcie badania**, tzw. „intercurrent events” i determinuje metody imputacji danych (RBMI).
- *Intercurrent events - czyli co zrobić, gdy pacjent:*
 - *odstawia lek (z dowolnego powodu),*
 - *otrzymuje leczenie ratunkowe,*
 - *wypada z badania ale kontynuuje leczenie,*
 - *wypada z badania i nic o nim nie wiemy (może kontynuować lub nie),*
 - *umiera przed pomiarem*



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

- **Treatment policy estimand** – analizujemy niezależnie od tego, co się stało (np. pacjent przerwał leczenie – jego pomiar nadal się liczy, nawet jeśli nie brał leku). Dodatkowo dostaje takie leczenie, jak zaplanowano, więc najbardziej pasuje do ITT
- **Hypothetical estimand** – pytamy: „jaki byłby efekt, gdyby intercurrent event nie wystąpił?” (pokazanie maksymalnego efektu, „co by było gdyby”) – imputacja „w ramieniu” (np. MI)
- **While on treatment estimand** – interesuje nas efekt tylko dopóki pacjent brał lek. Jeśli potem odstawi lek, nie patrzymy na wyniki po tym zdarzeniu. (nieco jak PP)
- **Composite estimand** – uznajemy np. śmierć lub treatment switching za niepowodzenie, niezależnie od poziomu LDL.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Pojawia się też nowe podejście do imputacji **RBMI**: *Reference-Based Multiple Imputation*:

- **treatment policy** (pozostaje w ramieniu)
- **jump to reference** (i modyfikacja: copy increments in reference) - przyszłe wartości endpointu będą podobne do wartości obserwowanych w grupie referencyjnej
- **copy reference** - wartości pacjenta po ICE będą dokładnie takie same jak średnie wartości w grupie referencyjnej.

...i kilka innych.



Wyzwania w badaniach RCT – projektowanie: sformalizowanie odpowiedzi

Dopiero teraz (!), wiedząc co i w jakiś sposób ma być analizowane, można przystąpić do obliczania **minimalnej liczby pacjentów**, niezbędnej do uzyskania zdolności detekcji minimalnego klinicznie istotnego efektu, zakładając pewną minimalną moc statystyczną (zdolność do wykrywania efektu) – typowo min. 80%, najlepiej 90 – 95%.

/ statistically **significant** → statistically **detectable** lub statistically **discernible** /

Potrzebujemy bowiem:

- ☑ poziomu istotności i mocy
- ☑ statystyki testowej dla prostego testu lub kompletnej definicji modelu (a od liczby zmiennych i innych parametrów zależy liczba potrzebnych obserwacji)
- ☑ rozmiaru efektu w populacji i zmienności populacyjnej
- ☑ marginesów non-inferiority lub MCID (jeśli konieczne)
- ☑ dodatkowych parametrów (np. dla badań longitudinalnych)
- ☑ ale także.....



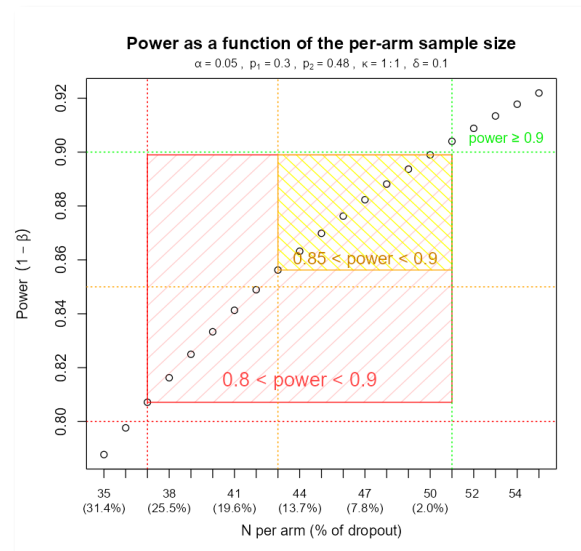
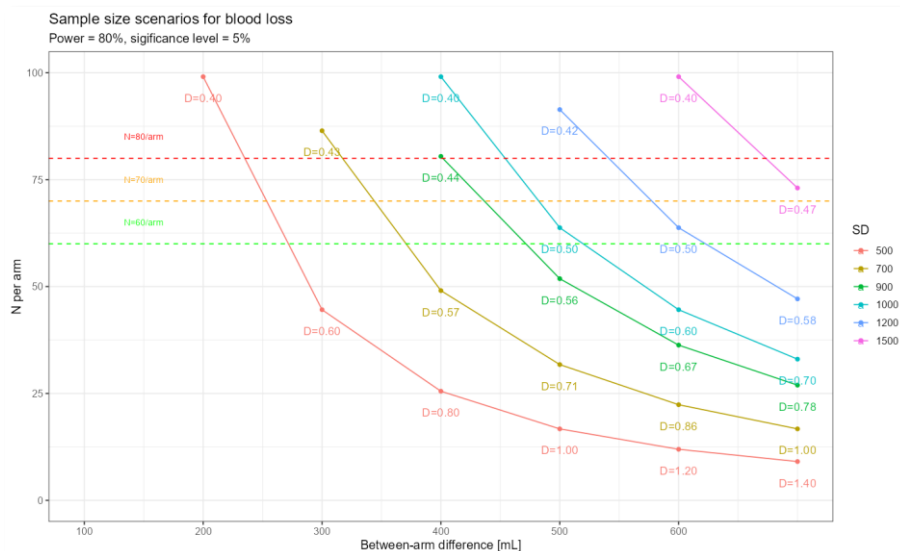
Wyzwania w badaniach RCT – projektowanie



- Projektowanie badania:
 - Określenie ram możliwości Sponsora
 - Wypracowanie kompromisu pomiędzy statystyczną i naukową poprawnością, a możliwościami realizacji i etycznością badania.
 - Wyznaczenie najmniejszej liczby pacjentów do rekrutacji, zaplanowanie analiz pośrednich i warunków wczesnego zatrzymania.

Wyzwania w badaniach RCT – projektowanie: kompromis potrzeby vs możliwości

Znając ograniczenia finansowe i logistyczne sponsora, jego potrzeby naukowe, a także opierając się o wiedzę dziedzinową z danego obszaru terapeutycznego (onkologia, kardiologia) co do powodów i natężenia „wypadania” pacjentów z badania (*dropout*), należy znaleźć **kompromis między tym co jest potrzebne a tym co możliwe**, zachowując etyczność badania.

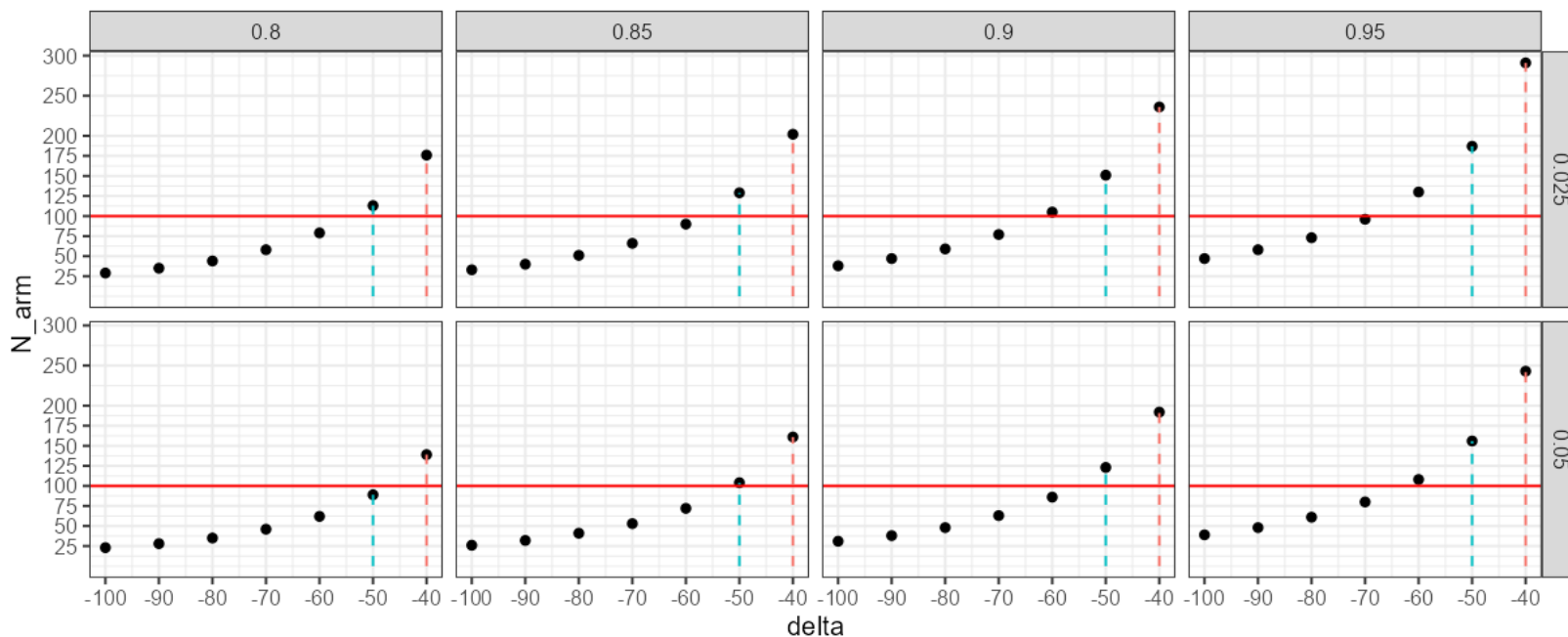


Wyzwania w badaniach RCT – projektowanie: kompromis potrzeby vs możliwości

Number of required patients as a function of significance level, power, margin and dispersion

Blood loss [mL] ($s=133.84$)

Clinically meaningful margin — $\Delta=25\%$ — $\Delta=33\%$

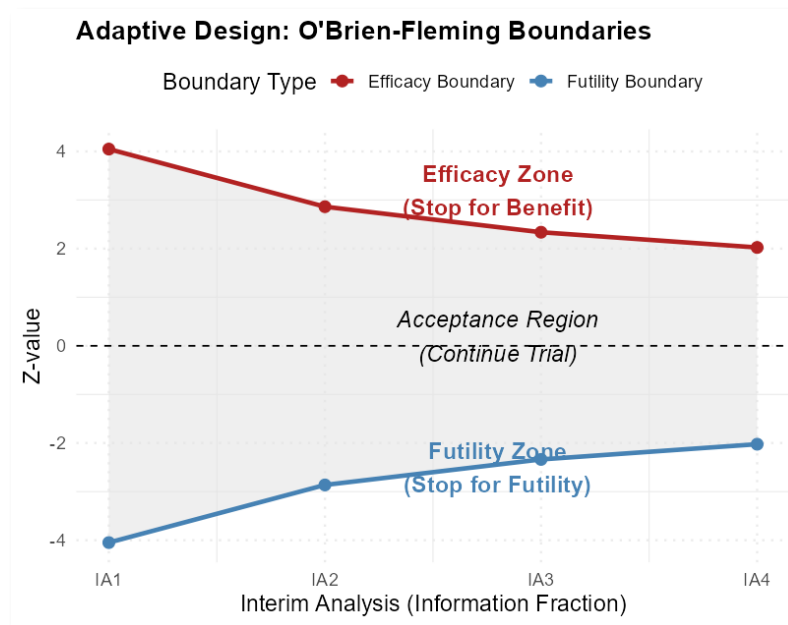


Wyzwania w badaniach RCT – projektowanie: kompromis potrzeby vs możliwości

Dodatkowo, w **badaniach adaptacyjnych** mamy dużą elastyczność.

Podczas analiz **interim** możemy, np.

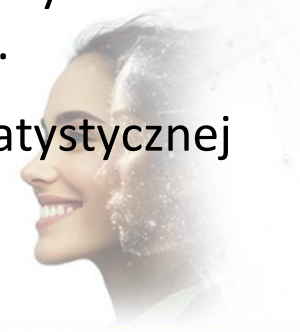
- odrzucać nieefektywne ramiona,
- re-estymować liczebność próby,
- a nawet zatrzymać badanie z powodu wczesnego wykrycia:
 - efektu (**early efficacy** – **ale uwaga!**),
 - problemów bezpieczeństwa (**safety**)
 - braku spodziewanej efektywności (**futility**).



Wyzwania w badaniach RCT – projektowanie



- Projektowanie badania:
 - Przewidywanie potencjalnych „czarnych scenariuszy” i opracowanie metod ratunkowych.
 - Na podstawie listy analiz do wykonania – opracowanie wykazu tabel, wykresów, listingów.
 - Napisanie Planu Analizy Statystycznej



Wyzwania w badaniach RCT – projektowanie: plany ratunkowe

- Podczas planowania analiz dyskutuje się także kwestie weryfikacji ich założeń i zakłada pewne warunki. Niestety, praktyka pokazuje, że niejednokrotnie jest to myślenie życzeniowe bez pokrycia:
 - ✖ Rozkłady empiryczne bywają dalekie od założonych, a założenia – niespełnione
 - ✖ Ekstremalne obserwacje się zdarzają. W biochemii klinicznej są normą...
 - ✖ Wypada wiele obserwacji, zatem tracimy moc. Nierzadko jest to MNAR.
 - ✖ Do tego imputacja (MICE) zawodzi – w końcu to model...
 - ✖ Zaplanowane modele „nie zbiegają” („*failed to converge to a solution*”)
 - ✖ Wyniki są trudne do interpretacji - dziwne (duże/male), niestabilne
 - ✖ Pojawiają się zaskakujące scenariusze z powodu „trudnych danych”
 - ✖ Mediany czasu przeżycia nie istnieją.
 - ✖ Założenia zaplanowanych metod nie są spełnione (hazardy, rozkłady, IID)

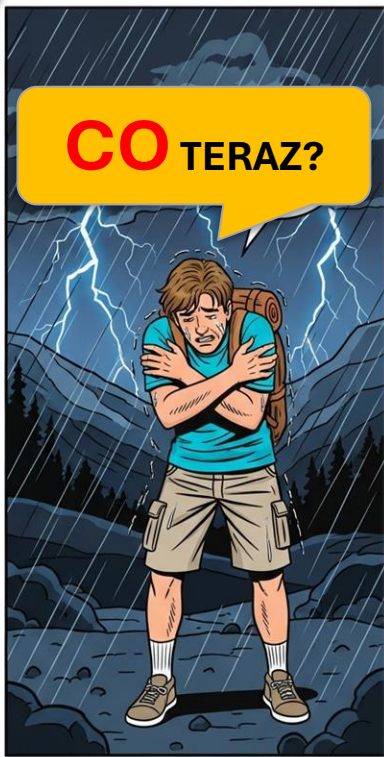


Wyzwania w badaniach RCT – projektowanie: plany ratunkowe

Troszkę się chmurzy, ale
szybciej zejdę szlakiem
bez schroniska.



CO TERAZ?



Co poszło
nie tak...?



Wyzwania w badaniach RCT – projektowanie: plany ratunkowe

- Bez planu awaryjnego analiza zatrzyma się nawet na pierwszym problemie.
- Kosztowne i czasochłonne badanie utknie w martwym punkcie, do czasu wsparcie ze strony agencji regulacyjnej. **A jeśli wsparcia nie będzie?**
- Jeśli mamy plan B, czy nawet C i D, to badanie może być do pewnego stopnia „uratowane” (**choć zawsze jakimś kosztem**).
- Oczywiście nie jesteśmy w stanie przewidzieć wszystkich problemów i zabezpieczyć się na każdy możliwy przypadek, ale warto spróbować stać się na chwilę „czarnowidzem” dla „dobra badania” (i naszego spokoju).
- Ważne, by plan B nie powodował zmiany hipotezy zerowej lub czynił to w stopniu akceptowalnym.
- Metody resamplingowe (permutacyjny test Welcha t), GEE, porządkowa regresja logistyczna



Wyzwania w badaniach RCT – projektowanie: plany ratunkowe

Flexible distribution-free & robust testing common hypotheses about numerical data

Subject of comparisons H0 vs. H1	1 factor				2+ factors + interactions		
	2 levels		k levels		of k levels	+ numerical covariates	+ repeated data
	unpaired	repeated data	unpaired	repeated data			
stochastic equivalence VS. stochastic superiority	- Mann-Whitney (Wilcoxon) - Brunner-Munzel	Signed-rank (Wilcoxon)	1-way: Kruskal-Wallis 2-way: Scheirer-Ray-Hare	- Friedman - Skellings-Mack RM-ANOVA 3-way: ATS 3-way: WTS - n-way ART ANOVA	- ATS ANOVA-Type Statistic [p] WTS permuted Wald-T.S. - ART ANOVA Aligned-Rank Transform - Brunner-Dette-Munk	Ordinal Logistic Regression = Proportional-Odds Model + Wald or LRT inference	GEE-fit Generalized Estimating Equations Ordinal Logistic Regression with appropriate correlation structure
equal means VS. unequal means	- Permutation Welch t - Permutation Yuen-Welch t (outliers)		1-way: Van der Waerden - Permutation ANOVA - GEE-fit General Linear Model	GEE-fit General Linear Model appropriate correlation structure	- GEE-General Linear Model + Wald inference - GEE-fit Ordinal Logistic Regression + extra work to estimate the eCDF - Robust (huberized) Linear Mixed Model + Wald/LRT For known non-Gaussian distributions - Generalized Linear/Additive Mixed Model + Wald/LRT - GEE-fit Generalized Linear Model + Wald		
equal medians VS. unequal medians	- Brown-Mood's - MW(W) under IID samples	Wilcoxon on symmetric sample	- Quantile regression (with random effects) + Wald or bootstrap inference - Khmaladze test for pure location shift				

Adrian Olszewski, 2024



Wyzwania w badaniach RCT – projektowanie: plany ratunkowe

Non-parametric and robust “ANCOVA”-like in R

estimation	R package	Example* (* uses <i>emmeans</i>)
rank-based	Rfit	<pre>> m_rfit <- Rfit::rfit(PD ~ Method + PD_Bas, data=data) > m_rfit_grid <- qdrg(formula = PD ~ Method + PD_Bas, data=data, coef = m_rfit\$coefficients, vcov = Rfit::vcov.rfit(m_rfit)) > pairs(emmeans(m_rfit_grid, specs = ~Method)) contrast estimate SE df z.ratio p.value Control - Active 1.75 0.178 Inf 9.822 <.0001</pre>
ordinal logistic regression	MASS::polr, rms::olr, ordinal:clm obtaining means from empirical CDF →	<pre>> m_por <- ordinal::clm(PD_ord ~ Method+PD_Bas, data=data) > pairs(emmeans(m_por, specs = ~Method)) contrast estimate SE df z.ratio p.value Control - Active 2.07 0.566 Inf 3.661 0.0003 > m_orm <- rms::orm(PD ~ Method+PD_Bas, data=data, x=TRUE, y=TRUE) > mean_fun <- Mean(m_orm) > Predict(m_orm, fun=mean_fun, Method=c("Control", "Active"), PD_Bas = mean(data\$PD_Bas)) Method PD_Bas yhat lower upper 1 Control 6.641667 5.730200 5.285577 6.174823 2 Active 6.641667 4.366212 3.933130 4.799293</pre>
quantile regression	quantreg	<pre>> m_rq <- quantreg::rq(PD ~ Method+PD_Bas, data=data, tau = 0.5) > pairs(emmeans(m_rq, specs = ~Method)) contrast estimate SE df t.ratio p.value Control - Active 2 0.285 57 7.028 <.0001</pre>
M/MM estimators	MASS::rlm, robustbase	<pre>> m_rlm <- MASS::rlm(PD ~ Method+PD_Bas, data=data) > pairs(emmeans(m_rlm, specs = ~Method)) contrast estimate SE df z.ratio p.value Control - Active 1.55 0.307 NA 5.050 <.0001</pre>
permutation	permuco	<pre>> permuco::lmperm(PD ~ Method+PD_Bas, data=data) Estimate Std. Error t value resampled Pr(> t) MethodActive -1.2929 0.3595 -3.597 0.0004</pre>
kernel (smoothing)	sm	<pre>> with(data, sm::sm.ancova(x = PD_Bas, y = PD, group = Method, model = "equal")) Test of equality : h = 0.411028 p-value = 0.0023</pre>

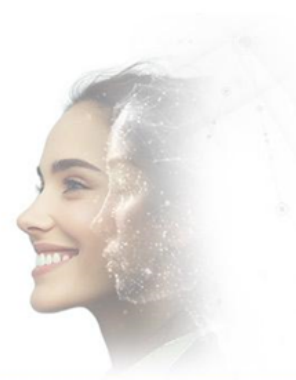


Wyzwania w badaniach RCT – projektowanie: plany ratunkowe

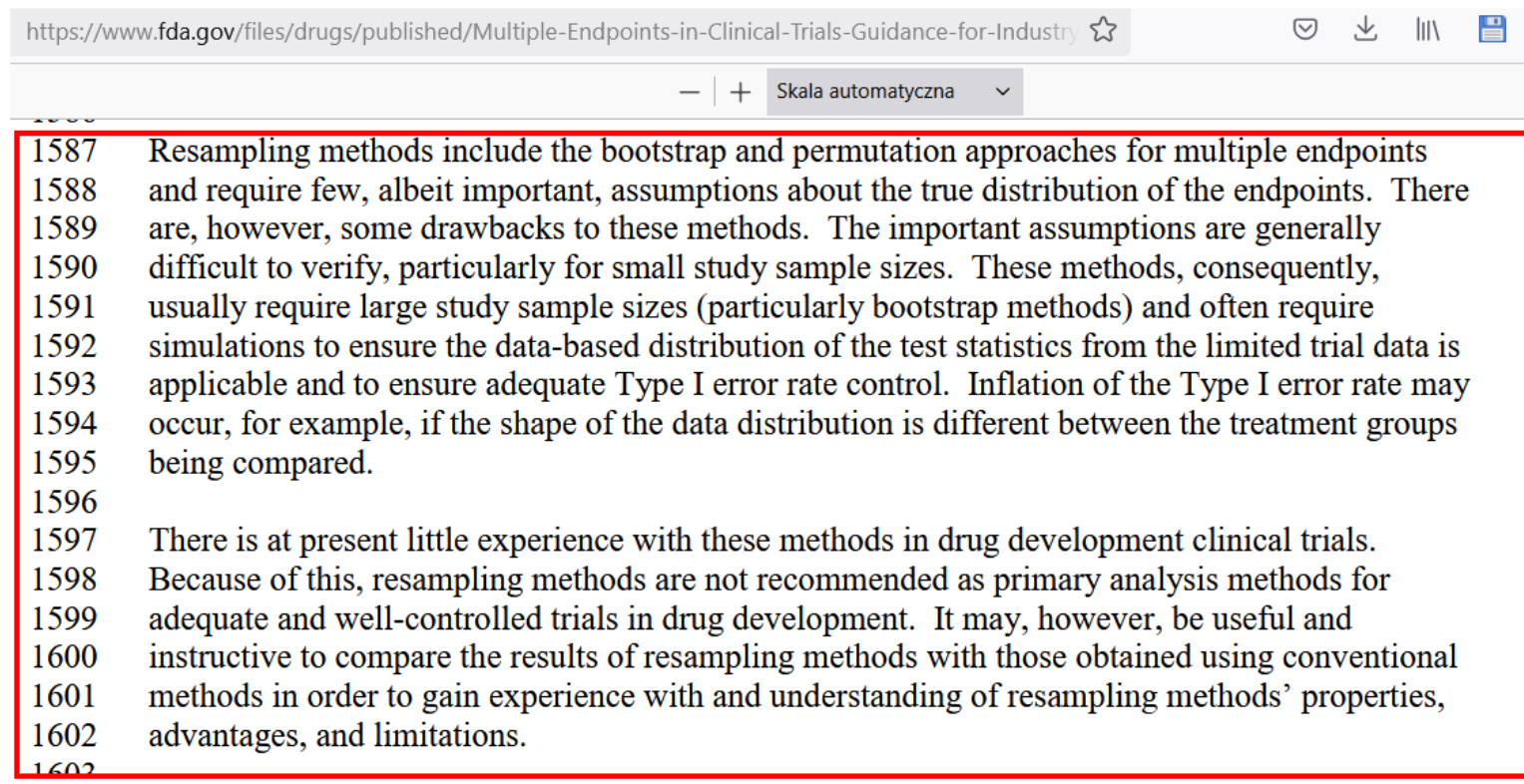


Sytuacja **komplikuje się**, jeśli:

- ☠ Plan B również nie działa, a nie mamy już planu C, nie mówiąc o planie „*last-resort*”
- ☠ Działanie naprawcze – psuje, tj. zmienia hipotezę zerową, **czyli odpowiadamy na pytanie inne, niż planowaliśmy.**
- ☠ Zapomnimy o analizie czułości dla różnych „ustawień” wybranej metody, **a recenzent użyje innych, dochodząc do innych wniosków...**



Wyzwania w badaniach RCT – projektowanie: plany ratunkowe



The screenshot shows a web browser window displaying a document from the FDA website. The address bar shows the URL: <https://www.fda.gov/files/drugs/published/Multiple-Endpoints-in-Clinical-Trials-Guidance-for-Industry>. The browser interface includes a search bar, a star icon, and a download icon. The document content is displayed in a table with line numbers on the left. The text discusses resampling methods, their assumptions, and their limitations in drug development clinical trials.

1587	Resampling methods include the bootstrap and permutation approaches for multiple endpoints
1588	and require few, albeit important, assumptions about the true distribution of the endpoints. There
1589	are, however, some drawbacks to these methods. The important assumptions are generally
1590	difficult to verify, particularly for small study sample sizes. These methods, consequently,
1591	usually require large study sample sizes (particularly bootstrap methods) and often require
1592	simulations to ensure the data-based distribution of the test statistics from the limited trial data is
1593	applicable and to ensure adequate Type I error rate control. Inflation of the Type I error rate may
1594	occur, for example, if the shape of the data distribution is different between the treatment groups
1595	being compared.
1596	
1597	There is at present little experience with these methods in drug development clinical trials.
1598	Because of this, resampling methods are not recommended as primary analysis methods for
1599	adequate and well-controlled trials in drug development. It may, however, be useful and
1600	instructive to compare the results of resampling methods with those obtained using conventional
1601	methods in order to gain experience with and understanding of resampling methods' properties,
1602	advantages, and limitations.
1602	



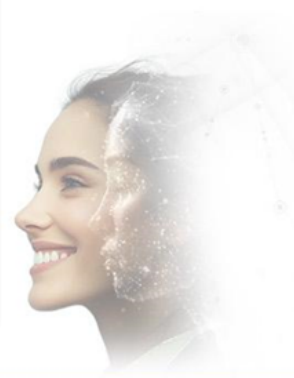
Wyzwania w badaniach RCT – projektowanie: plany ratunkowe

Statistical analysis

The assessment of bioequivalence is based upon 90% confidence intervals for the ratio of the population geometric means (test/reference) for the parameters under consideration. This method is equivalent to two one-sided tests with the null hypothesis of bioinequivalence at the 5% significance level.

The pharmacokinetic parameters under consideration should be analysed using ANOVA. The data should be transformed prior to analysis using a logarithmic transformation. A confidence interval for the difference between formulations on the log-transformed scale is obtained from the ANOVA model. This confidence interval is then back-transformed to obtain the desired confidence interval for the ratio on the original scale. A non-parametric analysis is not acceptable.

The precise model to be used for the analysis should be pre-specified in the protocol. The statistical analysis should take into account sources of variation that can be reasonably assumed to have an effect on the response variable. The terms to be used in the ANOVA model are usually sequence, subject within sequence, period and formulation. Fixed effects, rather than random effects, should be used for all terms.



Wyzwania w badaniach RCT – projektowanie: Tables / Figures / Listings

Table 39: SF36 PCS CMI by site [mITT]

SF36 PCS CMI / Time / Site	Population	SoC			Arm X			Difference (CI*)	Between-site comparisons**	
		N _{TOT}	N	%*	N _{TOT}	N	%*		Site XXX	Site XXX
Month 20 (N=99)	i. mITT	-	-	99.9 %	-	-	99.9 %	99.9 (99.9 ; 99.9)	-	-
Site XXX (N=99)	i. mITT	-	-	99.9 %	-	-	99.9 %	99.9 (99.9 ; 99.9)	p=0.999	p=0.999
Site XXX (N=99)	i. mITT	-	-	99.9 %	-	-	99.9 %	99.9 (99.9 ; 99.9)	-	p=0.999
Site XXX (N=99)	i. mITT	-	-	99.9 %	-	-	99.9 %	99.9 (99.9 ; 99.9)	-	-
Month 20 (N=99)	o. mITT	99	99	99.9 %	99	99	99.9 %	99.9 (99.9 ; 99.9)	-	-
Site XXX (N=99)	o. mITT	99	99	99.9 %	99	99	99.9 %	99.9 (99.9 ; 99.9)	p=0.999	p=0.999
Site XXX (N=99)	o. mITT	99	99	99.9 %	99	99	99.9 %	99.9 (99.9 ; 99.9)	-	p=0.999
Site XXX (N=99)	o. mITT	99	99	99.9 %	99	99	99.9 %	99.9 (99.9 ; 99.9)	-	-

Table 11: Baseline characteristics – medical history: body weight, height, and BMI [SAF]

Dataset / Variable	Measure	SoC		Arm 1		Arm 2
Safety	N	99		99		99
Body weight	N _{valid}	999		999		999
	mean (SD)	99.9 (99.9)		99.9 (99.9)		99.9 (99.9)
	median (Q ₁ ; Q ₃)	99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)
	min ; max	99 ; 99		99 ; 99		99 ; 99
Body height	N _{valid}	999		999		999
	mean (SD)	99.9 (99.9)		99.9 (99.9)		99.9 (99.9)
	median (Q ₁ ; Q ₃)	99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)
	min ; max	99 ; 99		99 ; 99		99 ; 99
BMI	N _{valid}	999		999		999
	mean (SD)	99.9 (99.9)		99.9 (99.9)		99.9 (99.9)
	median (Q ₁ ; Q ₃)	99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)
	min ; max	99 ; 99		99 ; 99		99 ; 99
Observed mITT	N	99		99		99
Body weight	N _{valid}	999		999		999
	mean (SD)	99.9 (99.9)		99.9 (99.9)		99.9 (99.9)
	median (Q ₁ ; Q ₃)	99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)		99.9 (99.9 ; 99.9)
	min ; max	99 ; 99		99 ; 99		99 ; 99

Table 75: Laboratory examinations [blood chemistry] - shift between discharge and baseline results, using the reference ranges [SAF]

Laboratory test / Shift from ↓ to →	Measure	SoC (N=99)			Arm 1 (N=99)			Arm 2 (N=99)		
		High / Abnormal*	Normal	Low	High / Abnormal*	Normal	Low	High / Abnormal*	Normal	Low
Complete Blood Count										
Example – numerical, e.g. <u>Erythrocytes [T/L]</u>										
	N _{pairs}	999			999			999		
High	N (%)*	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)
Normal	N (%)*	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)
Low	N (%)*	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)	99 (99.9%)
Example – categorical, e.g. <u>Nitrites</u>										
	N _{pairs}	999			999			999		
Abnormal	N (%)*	99 (99.9%)	99 (99.9%)	-	99 (99.9%)	99 (99.9%)	-	99 (99.9%)	99 (99.9%)	-
Normal	N (%)*	99 (99.9%)	99 (99.9%)	-	99 (99.9%)	99 (99.9%)	-	99 (99.9%)	99 (99.9%)	-
.....										
* The column combines contextual information to save space: "High" for numerical examinations and "Abnormal" for categorical ones.										

* The column combines contextual information to save space: "High" for numerical examinations and "Abnormal" for categorical ones.



Wyzwania w badaniach RCT – projektowanie: Tables / Figures / Listings

Table	Sub-group			
	(1)	(2)	(3)	(4)
Table 1 - Recruitment summary	-	-	-	-
Table 2 - Patients availability at each visit	-	-	-	-
Table 3 - Excluded patients – reasons	-	-	-	-
Table 4 - Site characteristics	-	-	-	-
Table 5 - Enrolment in sites	-	-	-	-
Table 19 - Past and baseline OCS treatment – unified daily doses translated to Prednisone-equivalents	-	-	-	-
Table 20 - Comorbidities at baseline	-	-	-	-
Table 21 – [PRIMARY] Overall <u>ACQ5</u> score changes from baseline after 8 weeks	•	•	•	•
Table 22 - Overall score changes from baseline after 1, 2, 4, 8, 16 and 56 weeks	•	-	-	•
Table 23 - Proportion of patients with a total score improvement of ≥ 0.5 (MCID) in <u>ACQ-5</u> after 1, 2, 4, 8, 16 and 56 weeks	-	•	•	•
Table 24 - Proportion of patients with well controlled, partially controlled and not well controlled asthma after 1, 2, 4, 8, 16 and 56 weeks.	-	•	-	•
Table 25 - Status of the improvement in <u>ACQ-5</u> after 1, 2, 4, 8, 16 and 56 weeks	•	•	-	-
Table 26 - OCS Dose Reduction after 4, 8, 16, 56 weeks vs. baseline	-	-	-	-
Table 27 - OCS dose (in prednisone equivalent) at each visit and change from the baseline	•	•	-	-
Table 28 - <u>PGI-C</u> over time – descriptive summary	-	-	-	-
Table 29 - <u>PGI-C</u> over time – summary by scale items	•	•		



Wyzwania w badaniach RCT – projektowanie: Tables / Figures / Listings

No.	Title	T/F/L	Complexity	FAS	N2	Overall survival LFUP censored	Overall survival LFUP accounted
22	Treatment – supportive therapies – erythropoiesis-stimulating agents (ESAs) at the index treatment	Table unique	medium	X			
23	Treatment – supportive therapies – granulocyte-colony stimulating factor (GCSF) at the index treatment	Table unique	high	X			
24	Toxicity – types by CTCAE grade (events) at the index treatment	Table unique	high	X			
25	Toxicity – types by CTCAE grade (patients) at the index treatment	Table unique	medium	X			
26	Toxicity – duration at the index treatment	Table unique	high	X			
27	Resource utilization at the index treatment	Table unique	high	X			
28	Recurrence table at the index treatment	Table unique	high	X			
29	Subsequent treatments (summary of all lines after index pooled, including metastatic treatment)	Table unique	medium	X			
30	Detailed subsequent treatment details for stage III NSCLC at each line	Table unique	medium	X			
31	Detailed subsequent treatment details in metastatic NSCLC	Table unique	high	X			
32	Baseline characteristics at the index date [N2 subgroup]	Table repeat	high		X		
33	Therapeutic modalities at the index treatment [N2 subgroup]	Table repeat	high		X		
34	Treatment for locally advanced stage III NSCLC at the index date [N2 subgroup]	Table repeat	high		X		
35	Treatment for locally advanced stage III NSCLC – Surgery at the index treatment [N2 subgroup]	Table repeat	high		X		
36	Treatment for locally advanced stage III NSCLC – Radiotherapy at the index treatment in the N2 group	Table repeat	high		X		
37	Chemotherapy agents (alone and combination therapies) at the index treatment in the N2 group	Table repeat	high		X		
38	Chemotherapy agents – combination therapies and cycle durations at the index treatment at the index treatment in the N2 group	Table repeat	high		X		
39	Chemotherapy agents – combination therapies along with number and durations of cycles at the index treatment in the N2 group	Table repeat	high		X		
40	Overall survival - summary [LFUP patients censored]	Table unique	medium			X	
41	Overall survival – supplementary summary [LFUP patients censored]	Table unique	high			X	
42	Overall survival – measures of the effect [LFUP patients censored]	Table unique	medium			X	
43	Overall survival – summary [LFUP patients counted]	Table repeat	medium				X
44	Overall survival – measures of the effect [LFUP patients counted]	Table repeat	medium				X
45	Progression-free survival – summary [LFUP patients censored]	Table repeat	medium				
46	Progression-free survival – measures of the effect [LFUP patients censored]	Table repeat	medium				
47	Progression-free survival – summary [LFUP patients counted]	Table repeat	medium				
48	Progression-free survival – measures of the effect [LFUP patients counted]	Table repeat	medium				
49	Overall survival - Summary of the Cox regression model [LFUP patients censored]	Table unique	low			X	
50	Overall survival - Summary of the Cox regression model [LFUP patients counted]	Table repeat	low				X
51	Progression-free survival - Summary of the Cox regression model [LFUP patients censored]	Table repeat	low				
52	Progression-free survival - Summary of the Cox regression model [LFUP patients counted]	Table repeat	low				



Plan Analiz Statystycznych – złota księga badania

- ✓ approvals obtained,
- ✓ consents signed,
- ✓ study launched,
- ✓ patients recruited
- ✓ SAP finalised...



...SAP
finalised...

ingflip.com



Plan analiz statystycznych

- ☞ Zmusza do przemyśleń
co? jak? dlaczego? co jeśli?
- ☞ Porządkuje informacje o celach badania i metodach analizy
- ☞ Nakreśla plan działania i „*plany B*”
co zrobić, kiedy, w jaki sposób i dlaczego?
- ☞ Pomaga szacować zasoby, harmonogramy i koszty
- ☞ Transparentność + wiarygodność



Wyzwania w badaniach RCT



Wyzwania w badaniach RCT – inne wyzwania – badania „underpowered”

p-value = 0.007 😊

Cohen's D = 1.36 😊

Średnia różnica: -18.5 > MCID – osiągnięta kliniczna istotność!

I to już dla niewielkiej próby $N = 10 + 10 = 20$ obserwacji!



A może to tylko artefakt – błąd typu M (a nawet S)?

/ Błąd 1 rodzaju o sile zwielokrotnionej M razy /



Wyzwania w badaniach RCT – inne wyzwania – badania „underpowered”

What NOT to do with NON-“null” results Part III: Underpowered study, but significant results

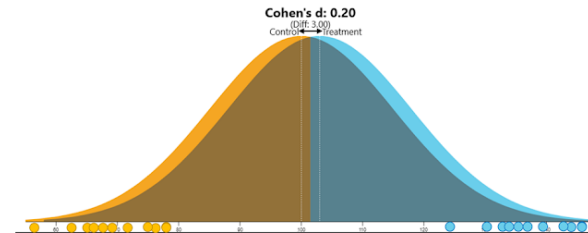
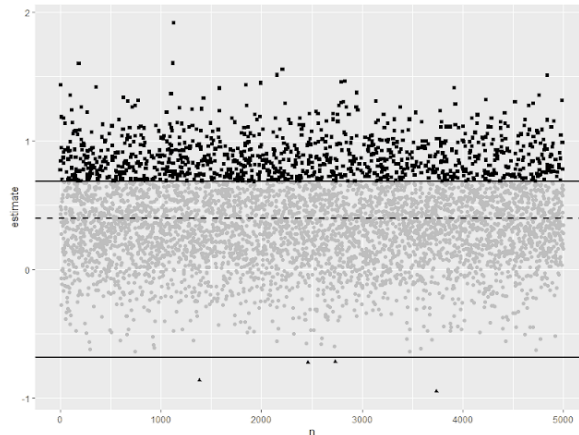
Dr Mircea Zloteanu, Feb 09, 2024

<https://mzloteanu.substack.com/p/what-not-to-do-with-non-null-results>

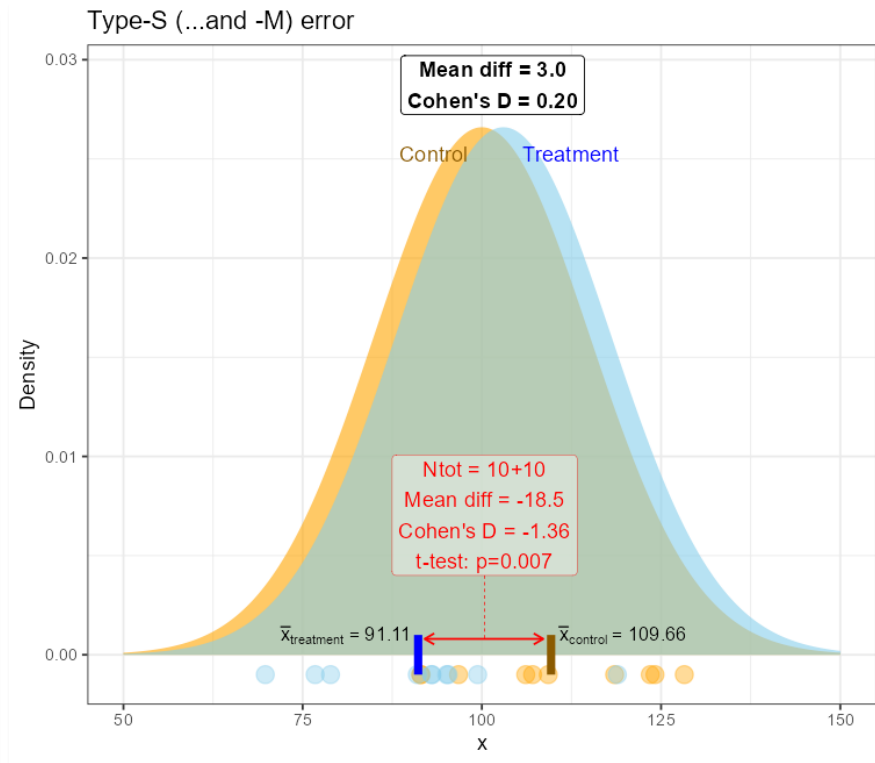
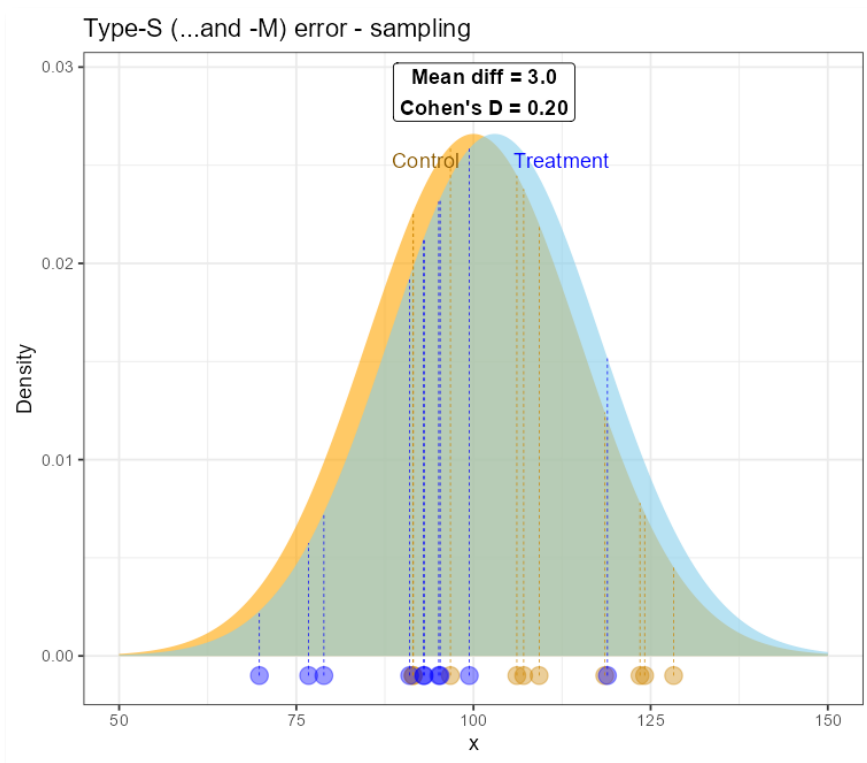


So, what are these errors?

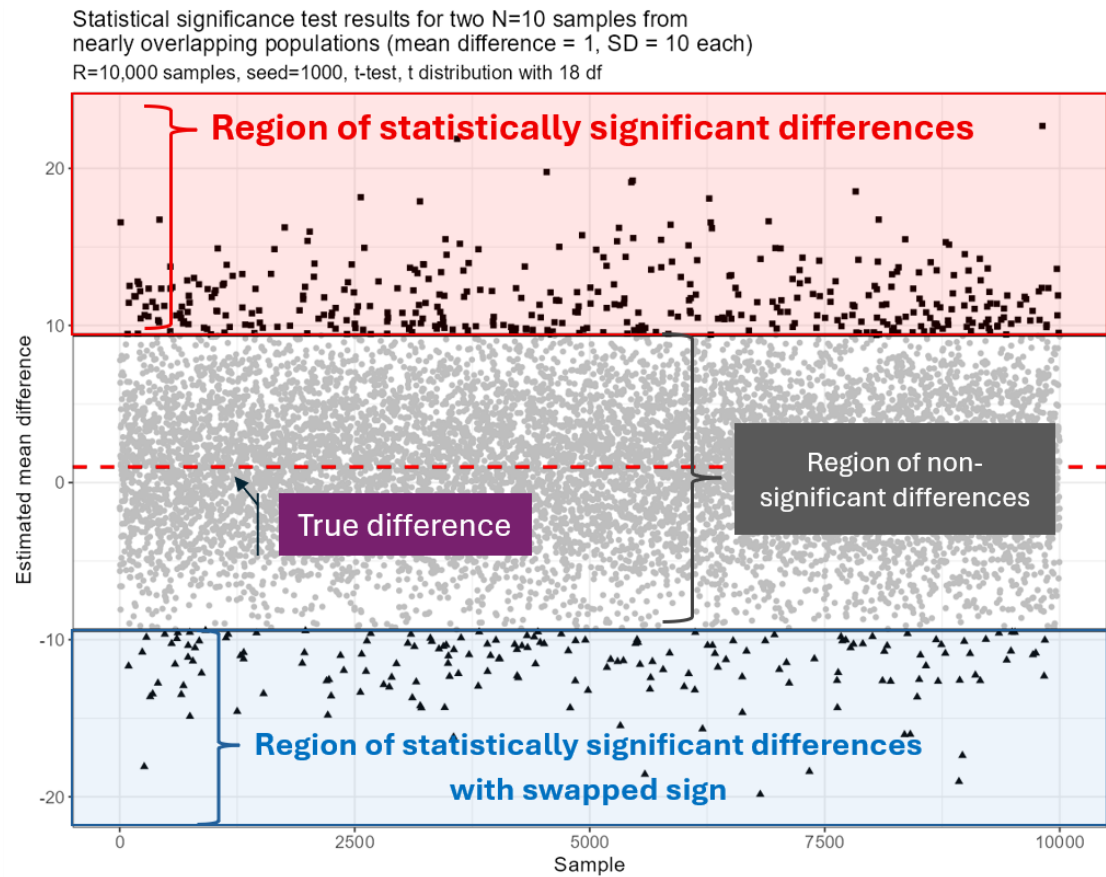
- **Type M error:** effects found in small sample studies are likely to be inflated (error of magnitude) (aka. the winner's curse; van Zwet & Cator, 2021), given that the test is statistically significant.
- **Type S error:** effects found in small sample studies have a higher chance of being in the wrong direction (error of sign), given that the test is statistically significant.



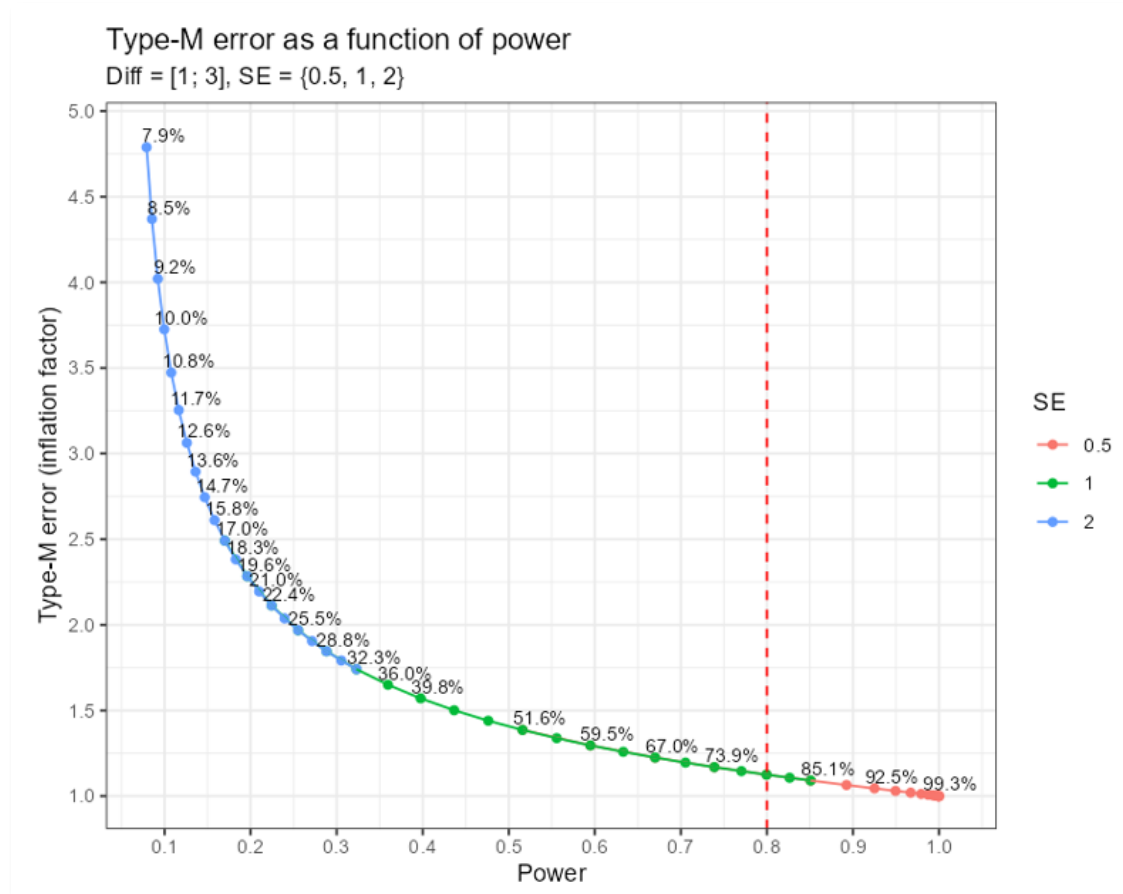
Wyzwania w badaniach RCT – inne wyzwania – badania „underpowered”



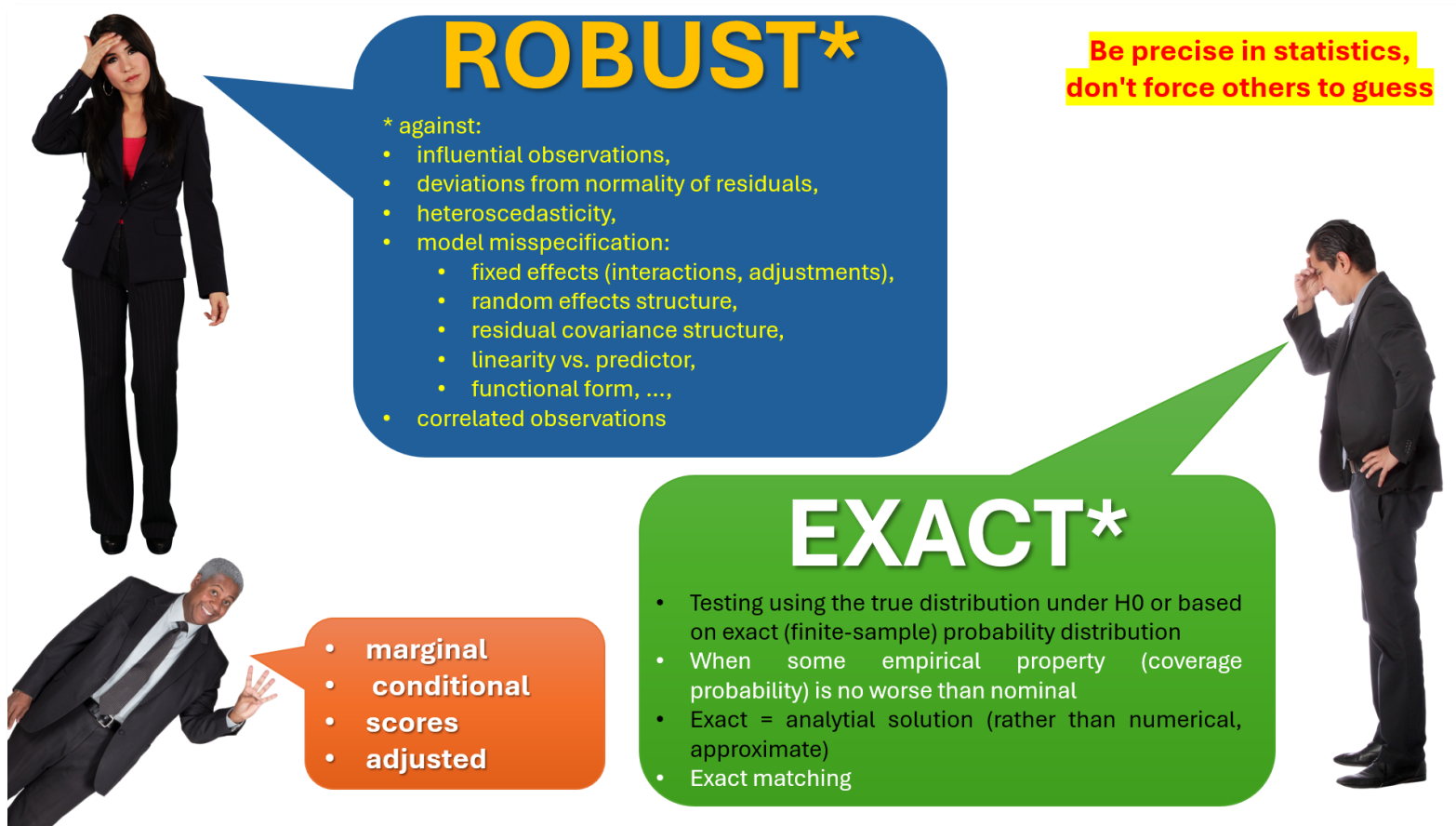
Wyzwania w badaniach RCT – inne wyzwania – badania „underpowered”



Wyzwania w badaniach RCT – inne wyzwania – badania „underpowered”



Wyzwania w badaniach RCT – inne wyzwania – komunikacja!



Wyzwania w badaniach RCT – inne wyzwania – badania nierandomizowane

Przy braku randomizacji zachodzi ryzyko, że obserwowany efekt nie pochodzi wyłącznie od zastosowanej interwencji, ale także o różnych czynników, w tym „naturalnych” różnic między porównywanymi grupami.

Istnieją metody **częściowego** (tylko dla obserwowanych predyktorów!) *naśladowania* efektów randomizacji i **zrównoważenie** porównywanych grup:

- **Estymacja G** (*kontrfaktyczny marginalny efekt brzegowy*) - model potencjalnych (każdy pacjent otrzymuje obie interwencje) wyników „przecięty” z rzeczywistymi danymi,
- **Ważenie IPTW** – ważenie odwrotnością prawdopodobieństwa otrzymania interwencji (bądź należenia do danej grupy) bazując na charakterystyce pacjentów
- **Propensity-Score Matching** – łączenie pacjentów o podobnej charakterystyce



Wyzwania w badaniach RCT – inne wyzwania – p-values



P-values are valid statistical measures that provide convenient conventions for communicating the uncertainty inherent in quantitative results. Indeed, *p*-values and significance tests are among the most studied and best understood statistical procedures in the statistics literature. They are important tools that have advanced science through their proper application.

In summary, *p*-values and significance tests, when properly applied and interpreted, increase the rigor of the conclusions drawn from data. Analyzing data and summarizing results are often more complex than is sometimes popularly conveyed. Although all scientific methods have limitations, the proper application of statistical methods is essential for interpreting the results of data analyses and enhancing the replicability of scientific results.



Wyzwania w badaniach RCT – inne wyzwania – p-values



Statistical Significance, Practical Importance, Planned Analyses (SAP), Study Design and Regulatory Guidelines support the King of Experimental Trials

The role of the frequentist framework, p-values, statistical (NHST) and practical significance (MCID) in randomised controlled (clinical) trials



Wyzwania w badaniach RCT – inne wyzwania – wytyczne

- ICH E3: Structure and Content of Clinical Study Reports,
- ICH E4: Dose-Response Information to Support Drug Registration,
- ICH E5 (R1): Ethnic Factors in the Acceptability of Foreign Clinical Data,
- ICH E7: Studies in Support of Special Populations,
- ICH E8(R1): General Considerations for Clinical Studies,
- ICH E9: Statistical Principles for Clinical Trials,
- ICH E9(R1): Addendum on Estimands and Sensitivity Analysis,
- ICH E10: Choice of Control Group in Clinical Trials,
- Guideline on Adjustment for Baseline Covariates in Clinical Trials (CHMP/295050/2013),
- Guideline on Missing Data in Confirmatory Clinical Trials (CPMP/EWP/1776/99),
- Points to Consider on Multiplicity Issues in Clinical Trials (CPMP/EWP/908/99),
- Guideline on the Investigation of Subgroups in Confirmatory Clinical Trials (CHMP/539146/2013),
- Points to Consider on Switching Between Superiority and Non-Inferiority (CPMP/EWP/482/99),
- CPMP/EWP/2158/99: Choice of a non-inferiority,
- Application with 1. Meta-analyses; 2. One Pivotal Study (CPMP/EWP/2330/99),
- Extrapolation of Efficacy and Safety in Medicine Development (EMA/129698/2012),
- Guideline on Clinical Trials in Small Populations- (CHMP/EWP/83561/2005),
- Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned With an Adaptive Design (CHMP/EWP/2459/02),
- Reflection Paper on the Use of Statistical Methodology for the Comparative Assessment of Quality Attributes in Drug Development (EMA/CHMP/138502/2017),
- FDA (FDA-2016-D-4460): Multiple Endpoints in Clinical Trials: Guidance for Industry,
- FDA: Non-Inferiority Clinical Trials to Establish Effectiveness: Guidance for Industry,
- FDA: Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry,
- FDA: Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products: Guidance for Industry,
- FDA: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials,
- Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drug and Biological Products (Level 1 Guidance; FDA-2019-D-1263),
- Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products (Draft; FDA-2022-D-2983),
- FDA: Use of Complex Innovative Designs for Drug and Biological Products (FDA-2019-D-3679),
- SPIRIT 2013 Statement: Defining standard protocol items for clinical trials,
- RECIST (1.1) for oncological trials,
- CONSORT 2010 Statement,
- NCBI: The Prevention and Treatment of Missing Data in Clinical Trials,





Jak najwcześniejsze zaangażowania statystyka w badaniu pozwoli uniknąć sytuacji, gdy problemy ujawniają się dopiero w momencie, gdy na efektywne działanie jest już za późno i **jedynie, co można będzie zrobić to stwierdzić że** (i ew. dlaczego) **badanie upadło.**

To consult the statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination.

Sir Ronald Aylmer Fisher



XXIX KONFERENCJA

ZASTOSOWANIA
STATYSTYKI
I DATA MINING
W BADANIACH
NAUKOWYCH



Adrian Olszewski

aolszewski@2kmm.pl

 aolszewski.devstat@gmail.com

Sesja pytań i odpowiedzi