

TOMASZ RZEPIŃSKI\*

## PROBLEMY EPISTEMOLOGICZNE ANALIZY PODGRUP RANDOMIZOWANYCH BADAŃ KLINICZNYCH

### Abstract

EPISTEMOLOGICAL CONTROVERSIES OVER RANDOMIZED CONTROLLED TRIALS SUBGROUP ANALYSIS

Subgroup analysis is one of the methods used in contemporary methodology of medicine for assessing the effects of undertaken therapy. It is often claimed that researchers who use the method overestimate its epistemic value. The main purpose of this paper is to analyze the credibility of conclusions formulated on the basis of subgroup analysis. Its findings will help to better understand the epistemological limitations of this method as well as to clarify its function in clinical trials.

*Keywords:* clinical trials, epistemology of medicine, statistical relevance, subgroup analysis

---

Głównym celem analiz prowadzonych we współczesnej metodologii medycyny jest opracowanie metod uzasadniania reguł decyzyjnych na potrzeby praktyki medycznej. Reguły te można podzielić na dwie grupy. Pierwsza obejmuje te, które mają usprawnić proces podejmowania decyzji diagnostycznych. Druga – reguły, których zadaniem jest racjonalizacja procesu podejmowania decyzji terapeutycznych.

Rozwój metodologii medycyny doprowadził do opracowania wielu w istotny sposób różniących się między sobą metod oceny skuteczności terapii. Różnice te dotyczą zarówno przebiegu procedur badawczych i ograniczeń etycznych związanych z ich zastosowaniem, struktury prowadzonych wnioskowań, jak i założeń dotyczących sposobów ustalania związków przyczynowych. Interesującą konsekwencją różnic między metodami oceny skuteczności terapii jest przypisywanie im różnej wiarygodności. Dane uzyskiwane z wykorzystaniem niektórych metod traktowane są z dużą dozą nieufności, podczas gdy inne

---

\* Zakład Logiki i Metodologii Nauk, Instytut Filozofii, Uniwersytet im. A. Mickiewicza, ul. Szamarzewskiego 89c, 60-568 Poznań, rzepinski@wp.pl.

rozpoznawane są jako bardzo wiarygodne. Zagadnienie wiarygodności danych pozyskiwanych z badań biomedycznych jest coraz częściej przedmiotem analiz prowadzonych przez epistemologów i filozofów nauki (Worrall 2002, Cartwright 2010).

W artykule przedstawię spór dotyczący uzasadniania reguł decyzyjnych z wykorzystaniem analizy podgrup dokonanej na podstawie wyników randomizowanych badań klinicznych (RCT). Analiza podgrup jest jedną z nielicznych metod, które umożliwiają indywidualizację podejmowanych oddziaływań terapeutycznych z uwagi na wstępną charakterystykę stanu pacjenta<sup>1</sup>. Tym samym jest to metoda, której wykorzystanie w ocenie skuteczności terapii pozwala oddalić zarzuty przeciwników analiz statystycznych o dehumanizację współczesnej medycyny (Sun i in. 2012: 2). Ponadto indywidualizacja oddziaływań terapeutycznych przynosi korzyści poznawcze, przybliżając nas do realizacji idei uzasadniania w medycynie reguł decyzyjnych o charakterze deterministycznym. Mimo wymienionych zalet można jednak wskazać badaczy kwestionujących wiarygodność wniosków formułowanych na podstawie analizy podgrup.

Zasadniczym celem artykułu jest przedstawienie tych argumentów krytycznych wobec analizy podgrup, które są szczególnie interesujące z perspektywy analiz epistemologicznych dotyczących wiarygodności konkluzji formułowanych na podstawie wnioskowań statystycznych. Pozwoli to lepiej zrozumieć ograniczenia analizy podgrup w praktyce badań biomedycznych oraz umożliwi sprecyzowanie funkcji poznawczej pełnionej przez nią w tych badaniach.

## 1. OCENA WIARYGODNOŚCI BADAŃ BIOMEDYCZNYCH

Badania biomedyczne produktów leczniczych i wyrobów medycznych to badania z udziałem ludzi przeprowadzane w celu uzyskania informacji o skuteczności i bezpieczeństwie podejmowanych działań terapeutycznych. Badania kliniczne są ich podzbiorem właściwym (Brodniewicz 2015: 31). Najogólniej można powiedzieć, że badania kliniczne różnią się tym od innych typów badań biomedycznych, że są związane z podejmowaniem interwencji, albo farmakologicznej, albo innego typu<sup>2</sup>. Zagadnienie wiarygodności badań biomedycznych

---

<sup>1</sup> Alternatywę stanowi tu analiza, w której wykorzystuje się teorię zbiorów przybliżonych (por. Słowiński 1992, Rzepiński 2012, 2014).

<sup>2</sup> Przyjmuję wprowadzone przez Parlament Europejski w 2014 r. ustalenia precyzujące pojęcie badań klinicznych. Jego wyczerpująca charakterystyka jest znacznie bardziej złożona od tej, którą podaję i której będę potrzebował w tym artykule (Brodniewicz 2015: 31).

jako potencjalnych źródeł danych o skuteczności terapii zostało szczegółowo omówione w ramach dyskusji nad medycyną faktów (*Evidence-Based Medicine*, EBM, por. Worrall 2002, Concato, Shah, Horwitz 2000, Djubegovic 2000). EBM, dążąc do ustalenia, które badania biomedyczne dostarczają najbardziej wiarygodnych informacji o skuteczności i bezpieczeństwie terapii, określa jednocześnie, jakie dane powinny być brane pod uwagę przez lekarzy podejmujących decyzje medyczne.

Badania biomedyczne pozwalają ustalać zależności statystyczne między terapią a późniejszym stanem zdrowia osobników w danej populacji. Wiadomo, że wnioski statystyczne są wnioskowaniami zawodnymi. Ich ocena była podejmowana przez filozofów nauki w kontekście zagadnienia poprawności wyjaśnień statystycznych (por. Hempel 1965, Salmon 2006). Zgodnie z koncepcją istotności statystycznej Wesleya Salmona wyjaśnienie statystyczne jest poprawne wówczas, gdy eksplanans identyfikuje zdarzenia, które są statystycznie istotne dla wystąpienia zjawiska opisywanego w eksplanandum. Są to zatem zdarzenia, których wystąpienie zwiększa prawdopodobieństwo zajścia danego zjawiska. Akceptując warunek istotności statystycznej ( $S_R$ ), porównujemy dwie wartości prawdopodobieństwa: wstępnego prawdopodobieństwa zajścia zjawiska opisanego w eksplanandum i prawdopodobieństwa warunkowego zajścia tego zjawiska z uwagi na wystąpienie zdarzenia opisanego w eksplanansie, czyli:

$$(S_R) P(r|t) > P(r)$$

gdzie  $P(r)$  to prawdopodobieństwo wstępne wystąpienia zdarzenia  $r$  opisanego w eksplanandum, a  $P(r|t)$  to prawdopodobieństwo warunkowe wystąpienia zdarzenia  $r$  z uwagi na wystąpienie zdarzenia  $t$ .

Załóżmy, że  $P(r)$  to prawdopodobieństwo poprawy stanu zdrowia u osób chorych na zapalenie płuc, natomiast  $P(r|t)$  to prawdopodobieństwo poprawy stanu zdrowia u tych osób z zapaleniem płuc ( $r$ ), które poddane były antybiotykoterapii ( $t$ ). Zgodnie z koncepcją Salmona wyjaśnienie statystyczne będzie poprawne wówczas, gdy w opisanym przypadku spełniony zostanie warunek  $S_R$ , czyli prawdopodobieństwo poprawy stanu zdrowia po antybiotykoterapii będzie większe niż prawdopodobieństwo poprawy stanu zdrowia bez tej terapii. Warunek  $S_R$  określa jednak tylko zależności statystyczne między zdarzeniami. Występowanie korelacji statystycznych nie może być podstawą ustalania zależności przyczynowych — to truizm, ale problem polega na tym, że przeprowadzane na podstawie badań biomedycznych wnioski statystyczne

---

W dalszej części tekstu termin „badania biomedyczne” jest konsekwentnie traktowany jako zakresowo szerszy od terminu „badania kliniczne”.

styczne są wykorzystywane w medycynie, by identyfikować zależności przyczynowe. Co więcej, uznaje się, że niektóre z badań biomedycznych lepiej nadają się do ustalania związków przyczynowych niż inne, mimo że we wszystkich z nich korzysta się z wnioskowań statystycznych. Dodać należy, że idea identyfikowania związków przyczynowych na podstawie stwierdzanych zależności statystycznych nie jest również obca filozofom nauki rozwijającym różne ujęcia w ramach tzw. probabilistycznych koncepcji przyczynowości. Przykładowo, Papineau (1994: 440) pisze:

jeżeli w badaniu z randomizacją stwierdzamy, że  $P(r|t) > P(r)$ , to możemy wnosić, że  $t$  jest przyczyną zdarzenia  $r$ <sup>3</sup>.

Scharakteryzowany przez Salmona warunek  $S_R$  nie jest więc traktowany w medycynie jako warunek oceny poprawności wyjaśnień statystycznych, lecz stanowi w istocie punkt wyjścia ustalania związków przyczynowych. O tym, czy spełnienie warunku  $S_R$  może stanowić podstawę wnioskowań identyfikujących zależności przyczynowe, decydują dodatkowe własności badań biomedycznych. Różnice w tym zakresie sprawiają, że niektóre z badań biomedycznych mimo spełnienia warunku  $S_R$  są uznawane za mało wiarygodne w procesie ustalania związków przyczynowych. Przykładem jest tu właśnie analiza podgrup. Jak piszą autorzy artykułu o znamienym tytule *Credibility of Claims of Subgroup Effects in Randomized Controlled Trials*, akurat w tej metodzie badawczej szczególnie istotna jest ocena poprawności wniosków identyfikujących zależności zachodzące między zdarzeniami (Sun i in. 2012: 2).

Pragnąc wskazać cechy, które zwiększają wiarygodność wnioskowań prowadzonych w badaniach biomedycznych, należy przywołać cztery ważne podziały w obrębie tych badań.

- (1) Podział badań biomedycznych na badania eksperymentalne i badania obserwacyjne

W badaniach eksperymentalnych interwencje terapeutyczne są stosowane wobec pacjentów w sposób celowy. Badacz decyduje, którzy pacjenci otrzymają aktywną lub nową terapię, a którzy będą leczeni metodą standardową lub wobec których stosowane będzie placebo. Z kolei w badaniach obserwacyjnych różnice w stosowanych interwencjach terapeutycznych są związane z pewnym zastanym stanem będącym wynikiem zwyczajowego stosowania

---

<sup>3</sup> Podobnie ujmują to Cartwright (1989, 2010). Polemikę z tymi ujęciami podejmują Worrall (2007) i Urbach (1985). W polskiej literaturze przedmiotu najbardziej wyczerpująco omówienie zagadnienia identyfikowania zależności przyczynowych w badaniach nieeksperymentalnych w medycynie można znaleźć w książce Kawalca (2006).

przez różnych lekarzy różnych środków terapeutycznych lub też wcześniejszego wpływu czynników szkodliwych na pacjentów<sup>4</sup>.

(2) Podział badań z uwagi na czas pozyskiwania danych o ekspozycji

Z uwagi na czas pozyskiwania danych o ekspozycji wyróżnia się wśród badań biomedycznych badania prospektywne i retrospektywne. Pierwsze z nich są zapoczątkowane w teraźniejszości i kontynuowane w przyszłości. W drugich ekspozycja miała miejsce w przeszłości.

(3) Podział badań biomedycznych z uwagi na kierunek prowadzonego wnioskowania

W *progresywnych* badaniach biomedycznych wnioskowanie prowadzone jest od przyczyny do skutku, natomiast w *regresywnych* przebiega od skutku do przyczyny. Podział ten nie funkcjonuje w literaturze z zakresu epidemiologii, ale wydaje się, że z uwagi na liczne nieporozumienia związane ze zbyt szerokim stosowaniem terminów „prospektywność” i „retrospektywność” warto go wprowadzić<sup>5</sup>.

(4) Podział badań z uwagi na uwzględnienie grupy kontrolnej

Wśród badań biomedycznych wyróżnia się badania z grupą kontrolną i bez niej. W pierwszych dokonuje się porównania grupy pacjentów eksponowanych na działanie pewnego czynnika z grupą pacjentów, którzy działaniu tego czynnika nie podlegali. Badania te mogą mieć charakter eksperymentalny lub obserwacyjny. Grupę kontrolną stanowią osoby przyjmujące imitację leku (placebo) lub poddawane standardowej terapii. Brak grupy kontrolnej w badaniach jest błędem metodyki badania<sup>6</sup>. Zauważmy, że dopiero zastosowanie grupy kontrolnej w badaniu biomedycznym umożliwia spełnienie warunku  $S_R$ .

<sup>4</sup> W epidemiologii stosuje się terminy „narażenie” i „ekspozycja”. Narażenie na działanie czynników szkodliwych może być związane z wykonywaną pracą (np. kontaktem ze środkami toksycznymi) lub trybem życia.

<sup>5</sup> Wprowadzenie terminów „progresywność” i „regresywność” zostało zaproponowane w dyskusji przez Włodzimierza Galewicza. Warto pomyśleć ten rozpow szechnić, chociażby z uwagi na możliwość uniknięcia niepotrzebnych nieporozumień językowych związanych z dwuznacznością terminów „prospektywność” i „retrospektywność”. Na problem niejednoznacznego użycia tych terminów zwracają uwagę Beaglehole, Bonita i Klejström (2002: 53).

<sup>6</sup> Wyjątkiem są badania fazy zero oraz większość badań fazy I. Badania te mają jednak realizować inne cele poznawcze niż badania późniejszych faz (Maselbas, Zastawny 2015: 79-85). Dlatego też mimo braku grupy kontrolnej wiarygodność badań wcześniejszych faz może być wysoko oceniana.

Należy zwrócić uwagę na dwie kwestie. Po pierwsze, wyznaczenie grupy kontrolnej różniącej się od grupy badanej wyłącznie czynnikiem ekspozycji umożliwia prowadzenie wnioskowania na podstawie kanonu jedynej różnicy. Po drugie, zachodzi ścisły związek między kierunkiem wnioskowania przeprowadzanego w badaniu biomedycznym a sposobem wyznaczania grupy kontrolnej. W badaniach, w których wnioskuje się regresywnie (z następstwa), grupa kontrolna wyznaczona jest na podstawie różnic w obserwowanych skutkach. Natomiast w badaniach, w których wnioskuje się progresywnie (od racji), grupa kontrolna wyznaczona jest na podstawie różnic w oddziaływaniu czynnika przyczynowego.

Wyróżnione podziały dotyczą najistotniejszych ogólnych własności<sup>7</sup> badań biomedycznych wpływających na ocenę wiarygodności danych pozyskiwanych za ich pomocą. Formułowany w ramach EBM postulat hierarchizacji danych z uwagi na ich wiarygodność jest ściśle związany z wyróżnionymi cechami (Worrall 2002, Concato, Shah, Horwitz 2000). Badania eksperymentalne są oceniane jako bardziej wiarygodne od obserwacyjnych. Z kolei wśród badań obserwacyjnych bardziej cenione są te, w których wnioskowanie ma charakter progresywny, a same badania są badaniami prospektywnymi. Mniejszą wiarygodność mają natomiast dane uzyskiwane z wykorzystaniem badań retrospektywnych, w których wnioskowanie ma charakter regresywny.

Na potrzeby dalszej analizy przybliżę dwa typy badań: badania kliniczno-kontrolne (CC) oraz randomizowane badania kliniczne (RCT).

## 2. CHARAKTERYSTYKA BADAŃ KLINICZNO-KONTROLNYCH I RANDOMIZOWANYCH BADAŃ KLINICZNYCH

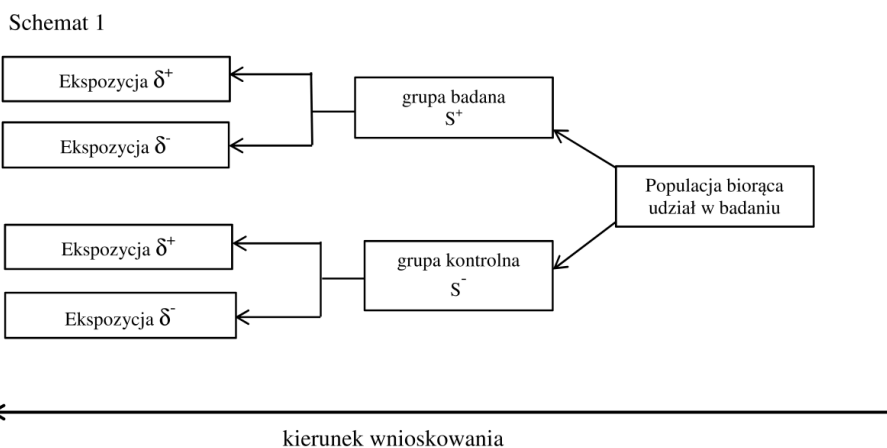
### 2.1. BADANIA KLINICZNO-KONTROLNE<sup>8</sup>

Badanie kliniczno-kontrolne przebiega w następujący sposób: w populacji zakwalifikowanej do badania wyodrębnia się grupę osób, u których stwierdza się wystąpienie określonego stanu, np. S<sup>+</sup>, a następnie sprawdza się, ile procent osób z tym stanem było narażonych na działanie przypuszczalnego czyn-

<sup>7</sup> Własności ogólnych, ponieważ wiarygodność badań biomedycznych danego typu może podlegać dodatkowej ocenie zgodnie z bardziej szczegółowymi kryteriami, np. takimi jak prawidłowość przeprowadzonej randomizacji, zastosowanie czytelnych kryteriów włączenia i wykluczenia pacjentów z badania itp.

<sup>8</sup> Aby uniknąć nieporozumień, należy podkreślić, że badania kliniczno-kontrolne są badaniami biomedycznymi, nie są jednak badaniami klinicznymi. Termin „kliniczno-kontrolne” jest terminem powszechnie używanym w epidemiologii.

nika przyczynowego  $\delta^+$ . W celu zwiększenia wiarygodności badania ustala się jednocześnie grupę kontrolną osób, u których nie występuje stan  $S^+$  i również analizuje się, ile procent osób z tej grupy było narażonych na działanie potencjalnego czynnika przyczynowego  $\delta^+$  (por. Schemat 1).



Schemat 1. Przebieg badania kliniczno-kontrolnego (Beaglehole, Bonita, Klejström 2002: 53)

Wyniki uzyskiwane w badaniu kliniczno-kontrolnym są traktowane jako dane o niskiej wiarygodności, ponieważ:

- (i) badania CC są badaniami obserwacyjnymi. Brak czynnika eksperymentalnego zmniejsza wiarygodność badania<sup>9</sup>;
- (ii) mają charakter retrospektywny. Są zatem oparte na danych, które były przedmiotem pomiarów w przeszłości. Tymczasem z uwagi na niedostateczny rozwój technologii diagnostycznych pomiar dokonany w przeszłości mógł nie obejmować istotnych parametrów;
- (iii) wnioskowania prowadzone w badaniach CC mają charakter regresywny. Wybór zdarzenia z historii przyczynowej pacjentów, które rozpoznawane jest jako potencjalna przyczyna ich późniejszego stanu, ma charakter arbitralny.

Zaletą badania CC jest łatwość pozyskiwania danych oraz niewielkie koszty. Do przeprowadzenia tego typu badań wystarczą dane przechowywane w rejestrach szpitalnych. Dodatkową zaletą jest to, że umożliwiają ocenę wpływu niekorzystnych czynników na zdrowie pacjenta. Pozwalają zatem ocenić

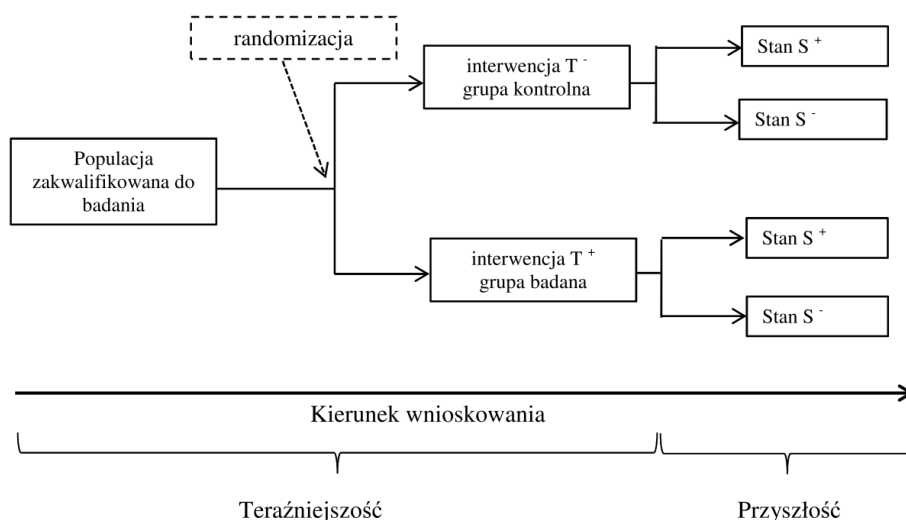
<sup>9</sup> Interwencja nie jest wprowadzona w sposób celowy. Dlatego właśnie nie mamy w tym wypadku do czynienia z badaniami klinicznymi.

wpływ ekspozycji, którego nie można ustalić z wykorzystaniem badań eksperymentalnych, niestosowalnych ze względów etycznych.

## 2.2. RANDOMIZOWANE BADANIA KLINICZNE

Randomizowane badania kliniczne mają następujący przebieg: w populacji osób zakwalifikowanych do badania wyodrębnia się grupy (kohorty) osób poddanych interwencji oraz grupę kontrolną. Zarówno grupa kontrolna, jak i grupy badane obejmują pacjentów z określonym rozpoznaniem jednostki chorobowej. W grupach badanych testuje się nowe rodzaje interwencji. Czynniki lub czynniki interwencyjne są celowo wprowadzane przez badacza. Ważną własnością badania RCT, poza jego eksperymentalnym charakterem, jest losowy przydział pacjentów do poszczególnych kohort. Przebieg badania RCT ilustruje poniższy schemat.

Schemat 2



Schemat 2. Przebieg randomizowanego badania klinicznego (wg Beaglehole, Bonita, Klejström 2002: 55). T<sup>+</sup> – zastosowanie interwencji, T<sup>-</sup> – niezastosowanie interwencji, S<sup>-</sup> – brak oczekiwanego efektu, S<sup>+</sup> – oczekiwany efekt

W dalszej części artykułu stan S<sup>-</sup> będzie określany mianem „zdarzenia niekorzystnego”.

W badaniu RCT przeprowadza się wnioskowanie progresywne. Samo badanie ma zaś charakter prospektywny. Badanie RCT jest traktowane jako badanie



o znacznie większej wiarygodności niż badanie CC. Wyniki badania RCT przedstawia się najczęściej pod postacią tzw. tabeli 2x2:

	Zdarzenie korzystne	Zdarzenie niekorzystne
Interwencja (+)	A	C
Interwencja (-)	B	D

Tabela 1. A, B, C, D – zbiory pacjentów

Dysponując wynikami ujętymi w takiej tabeli, dokonuje się ich analizy, określając procent osób w grupie poddanej interwencji, które odniosły korzyść z zastosowanej ingerencji terapeutycznej oraz procent osób z grupy kontrolnej, u których – mimo braku interwencji terapeutycznej – również wystąpił korzystny efekt. W celu oszacowania korzyści związanych z podjętymi oddziaływaniami terapeutycznymi wprowadza się pojęcia ryzyka, ryzyka względnego i względnej redukcji ryzyka (Jaeschke, Cook, Guyatt 1998).

Przyjmijmy, że  $n$  oznacza licznosc zbioru. Ryzyko wystąpienia zdarzeń niekorzystnych (czyli braku oczekiwanego efektu) w grupie interwencji, czyli  $R(T^+)$ , można wówczas określić jako:

$$R(T^+) = \frac{n(C)}{n(A+C)}$$

Natomiast ryzyko zdarzeń niekorzystnych w grupie bez interwencji jako:

$$R(T^-) = \frac{n(D)}{n(B+D)}$$

Znając  $R(T^+)$  oraz  $R(T^-)$ , można obliczyć, jaką część ryzyka wystąpienia zdarzenia niekorzystnego w grupie bez interwencji stanowi ryzyko wystąpienia zdarzenia niekorzystnego w grupie z interwencją. Jest to tzw. *ryzyko względne*  $RR(T^+)$ :

$$RR(T^+) = \frac{R(T^+)}{R(T^-)}$$

Znając ryzyko względne, można z kolei określić, w jakim stopniu zastosowane oddziaływanie terapeutyczne zmniejszyło ryzyko wystąpienia zdarzenia niekorzystnego w grupie z interwencją. Jest to *względna redukcja ryzyka*  $RRR(T^+)$ :

$$RRR(T^+) = 1 - RR(T^+)$$

Pojęcie względnej redukcji ryzyka jest zatem stosowane, aby wyrazić korzyści związane z podjętą w badaniu RCT interwencją terapeutyczną. Jeżeli różnice między efektami w grupie interwencji ( $T^+$ ) osiągną pewien arbitralnie określony poziom istotności, to wnioskuje się, że zastosowana interwencja jest czynnikiem przyczynowo istotnym dla stanu zdrowia pacjenta. Innymi słowy, o zależnościach przyczynowych wnioskuje się wówczas, gdy spełniony jest warunek:  $P(r|t) > P(r)$ .

W analogiczny sposób, również na podstawie różnicy istotności statystycznej, można ocenić skutki ekspozycji w badaniu kliniczno-kontrolnym<sup>10</sup>.

### 3. ANALIZA PODGRUP W RANDOMIZOWANYM BADANIU KLINICZNYM

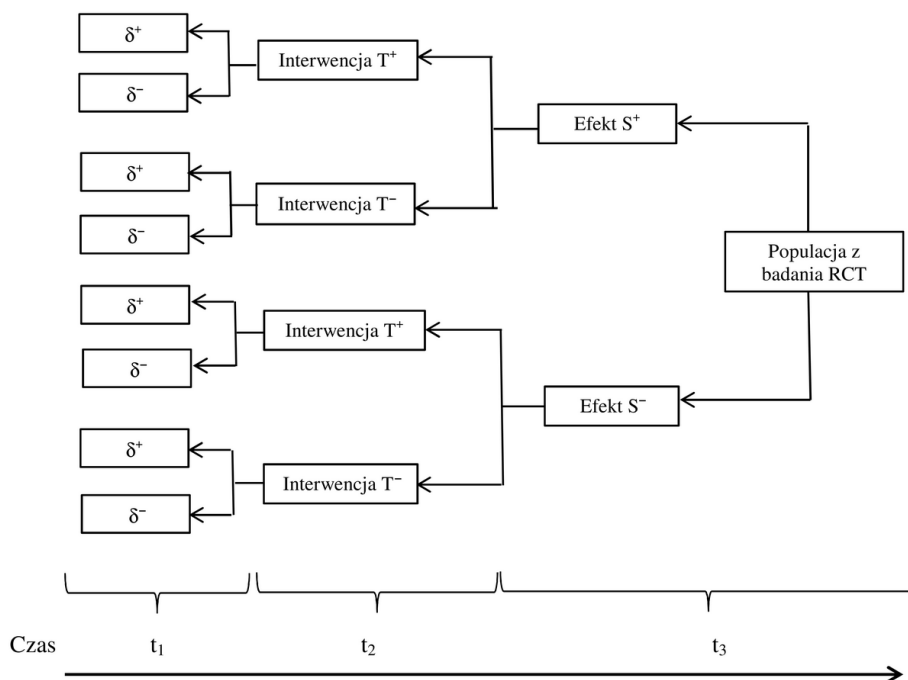
Zauważmy, że dysponując tylko głównymi wynikami badania RCT, lekarz nie jest w stanie stwierdzić, czy zastosowana w tym badaniu interwencja będzie skuteczna w jego praktyce medycznej. Osoby z rozpoznaną chorobą zakwalifikowane do badania RCT mogą różnić się między sobą pod wieloma względami, które wpływają na efekty terapii. Znaczenie może mieć tu wiek, płeć, współwystępowanie innych chorób lub określone wartości wskaźników diagnostycznych identyfikujących różne stopnie zaawansowania choroby (Oxman, Guyatt 1992: 78). Wpływ tych cech w badaniu RCT zostaje uśredniony. Efekt terapeutyczny stanowiący przedmiot pomiaru jest wówczas wypadkową wielu cech charakteryzujących osoby biorące udział w badaniu. Badanie RCT dostarcza wyniku tworzącego uśredniony obraz nieistniejącego pacjenta, będącego zarówno kobietą, jak i mężczyzną, zażywającego dodatkowe środki farmakologiczne, jak i jednocześnie niezżywającego ich itd. Tymczasem lekarz w swojej praktyce ma do czynienia z konkretnym pacjentem o konkretnej charakterystyce. W tym miejscu przydatna może być właśnie analiza podgrup (Oxman, Guyatt 1992: 79).

Przyjmijmy, że w badaniu RCT uzyskano pewne wyniki reprezentowane za pomocą Tabeli 1. Badacz przeprowadzający badanie RCT zastanawia się następnie, czy pewna cecha  $\delta^+$  jest istotna dla efektów terapii. Aby to ustalić, sprawdza, ilu pacjentów z interwencją  $T^+$  ma cechę  $\delta^+$ , a ilu cechę  $\delta^-$ . Te same ustalenia czyni się w wypadku pacjentów z grupy kontrolnej bez interwencji. Następnie oblicza się ryzyka wystąpienia zdarzeń niekorzystnych w wyznaczonych podzbiorach (por. Schemat 3).

---

<sup>10</sup> Analogiczny, lecz nie ten sam. W analizie danych badania CC wykorzystuje się pojęcie *ilorazu szans* (Jędrychowski 2002: 140-142).

Schemat 3



Schemat 3. Analiza podgrup

Celem poznawczym analizy podgrup jest zidentyfikowanie takich czynników występujących w historii przyczynowej pacjentów w czasie  $t_1$ , które wpłynęły na przebieg interwencji podjętej w czasie  $t_2$ , wywołując efekty w czasie  $t_3$  (Oxman, Guyatt 1992: 79). Podkreślić należy, że analiza podgrup nie jest wykonywana przez lekarza na podstawie dostępnych wyników badania RCT, lecz przez badacza przeprowadzającego to badanie, a wnioski formułowane na jej podstawie są publikowane łącznie z głównymi wynikami badania RCT.

Wydawałoby się, że analiza podgrup umożliwia realizację wskazanego wcześniej celu poznawczego. Pozwala mianowicie precyzyjniej niż badanie RCT identyfikować związki między interwencją podjętą wobec określonej podgrupy pacjentów a efektem tej interwencji. Uzyskane w tym badaniu wyniki stanowią podstawę indywidualizacji wskazań terapeutycznych z uwagi na wstępną charakterystykę pacjentów (Parker, Naylor 2000: 952, Burke i in. 2015: 1). W literaturze przedmiotu pokazuje się jednak, że identyfikacja związków przyczynowych na podstawie analizy podgrup może budzić wiele wątpliwości.

#### 4. ANALIZA PODGRUP — PRZYKŁADY ZAGROŻEŃ

Z perspektywy praktyki medycznej analiza podgrup może prowadzić do dwóch niebezpiecznych konsekwencji. Pierwszą jest formułowanie reguł decyzyjnych ustalających zaniechanie działań wobec pacjentów nieposiadających cech, które w tej analizie zostały rozpoznane jako korzystne dla efektów terapii. Drugą jest zidentyfikowanie pewnych cech pacjentów jako korzystnych dla efektów terapii, podczas gdy cechy te w istocie nie mają wpływu na stan końcowy pacjentów. Rozpatrzmy odpowiednie przykłady.

##### 4.1. WYNIKI ANALIZY PODGRUP W BADANIACH GISSI I CCSG

W 1986 r. wykonane zostało badanie RCT mające na celu ocenę skuteczności terapii streptokinazą osób po przebytym zawale serca (badanie GISSI)<sup>11</sup>. Uzyskane wyniki stały się podstawą do przeprowadzenia analizy podgrup. Jej celem było ustalenie wpływu dodatkowych dwóch czynników, które — jak sądzono — mogą oddziaływać na przebieg terapii. Pierwszym był czas, który upłynął u pacjentów biorących udział w badaniu GISSI od początku zawału serca. Drugim — wiek pacjentów. Przeprowadzona analiza pozwoliła stwierdzić, że zastosowanie streptokinazy jest korzystniejsze w grupie osób poniżej 65. roku życia lub wówczas, gdy terapia została podjęta w czasie krótszym niż 6 godzin od pierwszych symptomów zawału (GISSI 1986, Lee i in. 1989). Wnioski te mogły zatem stanowić podstawę do sprecyzowania algorytmów decyzyjnych uzasadniających podjęcie działań terapeutycznych wobec określonych grup pacjentów. Przeprowadzone później badania nie potwierdziły jednak tych wniosków: nie wykazały żadnych statystycznie istotnych różnic między efektami leczenia pacjentów różniących się czasem podania leku — do 24 godzin — oraz wiekiem (ISSIS-2 1988).

Podobny przykład przytaczany jest w kontekście skuteczności stosowania aspiryny w celu zmniejszenia ryzyka zawału serca. Dokonana początkowo analiza podgrup pozwoliła stwierdzić, że aspiryna wykazuje prewencyjne działanie wyłącznie wśród mężczyzn (CCSG 1978). Wnioski te zostały jednak również zakwestionowane (ATC 1988).

Oba przykłady są często przywoływane, ponieważ wskazują na poważne niebezpieczeństwo przy podejmowaniu decyzji medycznych. Może się okazać, że wyniki analizy podgrup, faworyzując podejmowanie działań wobec pacjentów o określonej charakterystyce, prowadzą jednocześnie do zaniechania

---

<sup>11</sup> Streptokinaza to jeden z leków mających na celu „rozpuszczenie” skrzepliny zamykającej światło tętnicy wieńcowej w trakcie zawału.

działań wobec pacjentów, którzy nie spełniają ustalonych w analizie podgrup warunków wstępnych. W odniesieniu do pierwszego przykładu mogłoby to skutkować błędną decyzją o zaniechaniu stosowania streptokinazy u osób, u których zawał serca nastąpił ponad 6 godzin od podjęcia terapii, lub u osób powyżej 65. roku życia (Oxman, Guyatt 1992: 78). W drugim, wyniki analizy podgrup prowadziłyby do zaniechania prewencyjnego stosowania aspiryny u kobiet (Oxman, Guyatt 1992: 80, Hernandez i in. 2006: 260)<sup>12</sup>.

#### 4.2. ANALIZA PODGRUP W BADANIU ISSIS-2

O świadomości zagrożeń wynikających z zastosowania analizy podgrup mogą świadczyć wnioski sformułowane przez autorów badania ISSIS-2. Jego celem była ocena skuteczności prewencyjnego stosowania aspiryny w zapobieganiu zawałowi serca. Aby ukazać niebezpieczeństwo analizy podgrup, autorzy tego badania wyróżnili podgrupy osób z uwagi na ich przynależność do różnych znaków astrologicznych i wykazali, że wyniki ich analizy wskazują na szczególnie korzystne oddziaływanie aspiryny u tych, którzy nie urodzili się pod znakiem bliźniąt lub wagi. Przyjrzyjmy się bliżej temu przykładowi.

	Przeżycie	Zgony spowodowane zawałem serca	Łącznie
Aspiryna T (+)	A = 7783	C = 804	8587
Placebo T (-)	B = 7584	D = 1016	8600

Tabela 2. Wyniki uzyskane z randomizowanego badania ISSIS-2

Na podstawie Tabeli 2 można oszacować ryzyko wystąpienia zdarzenia niekorzystnego w grupie T<sup>+</sup> (aspiryna) oraz w grupie T<sup>-</sup> (placebo). Następnie można obliczyć ryzyko względne dla T<sup>+</sup> oraz względną redukcję ryzyka uzyskaną dzięki podjętej interwencji.

Ryzyko wystąpienia zdarzenia niekorzystnego w grupie osób poddanych interwencji to:

$$R(T^+) = 804 / 8587 = 9,36\%$$

Odpowiednio, ryzyko wystąpienia zdarzenia niekorzystnego w grupie kontrolnej to:

<sup>12</sup> Podkreślić należy, że podane przykłady nie są odosobnione. W literaturze przedmiotu przytacza się wiele błędnych wniosków formułowanych na podstawie analizy podgrup (Yusuf i in. 1991).

$$R(T^-) = 1016 / 8600 = 11,8\%$$

Ryzyko względne wystąpienia zdarzenia niekorzystnego w  $T(+)$  to:

$$RR(T^+) = 9,36 / 11,8 = 79\%$$

Względna redukcja ryzyka w grupie  $(T^+)$  wynosi zatem:

$$RRR(T^+) = 1 - 0,79 = 21\%$$

Oznacza to, że prewencyjne zażywanie aspiryny zmniejsza o 21% ryzyko zawału serca w stosunku do grupy kontrolnej.

W dalszej kolejności autorzy badania ISIS-2 przeprowadzili analizę podgrup, wyróżniając dwie grupy osób z uwagi na ich przynależność do różnych znaków astrologicznych. Pierwsza grupa obejmowała osoby urodzone pod znakiem bliźniąt lub wagi (BW), druga — osoby spod pozostałych znaków. Dokonano zatem wskazania cechy (znak astrologiczny), która miałaby być rozważana jako potencjalnie przyczynowo istotna dla przebiegu terapii. Wyniki rozkładu pacjentów przedstawione są w Tabeli 3 i 4.

	Przeżycie w grupie bliźniąt lub wagi	Zgony spowodowane zawałem serca w grupie bliźniąt lub wagi	Łącznie
Aspiryna T (+)	$A_1 = 150$	$C_1 = 1207$	1357
Placebo T (-)	$B_1 = 147$	$D_1 = 1295$	1442

Tabela 3. Wyniki analizy podgrup badania ISIS-2 — pacjenci spod znaku bliźniąt lub wagi

	Przeżycie w grupie pozostałych znaków	Zgony spowodowane zawałem serca w grupie pozostałych znaków	Łącznie
Aspiryna T (+)	$A_2 = 654$	$C_2 = 6574$	7228
Placebo T (-)	$B_2 = 868$	$D_2 = 6319$	7187

Tabela 4. Wyniki analizy podgrup badania ISIS-2 — pacjenci spod innych znaków zodiaku

Na podstawie danych z Tabeli 3 i 4 można obliczyć ryzyka:

$$R_{BW}(T^+) = 150 / 1357 = 11\%$$

$$R_{inne znaki}(T^+) = 654 / 7228 = 9\%$$

$$R_{BW}(T^-) = 147 / 1442 = 10,2\%$$

$$R_{inne znaki}(T^-) = 868 / 7187 = 12,1\%$$

$$RR_{BW}(T^+) = 11,1 / 10,2 = 108\%$$

$$RR_{inne znaki}(T^+) = 9 / 12,1 = 75\%$$

Jak widać, ryzyko względne (RR) w grupie BW jest powyżej 1, a więc nie będziemy mieli do czynienia z redukcją ryzyka, lecz ze wzrostem ryzyka w porównaniu z grupą kontrolną. Zatem obliczając wzrost ryzyka dla grupy BW, należy brać pod uwagę wartość bezwzględną. Uzyskujemy tym samym następujące wartości:

$$RRR_{BW}(T^+) = |1 - 1,08| = 8\%$$

$$RRR_{inne\ znaki}(T^+) = 1 - 0,75 = 25\%$$

Podsumowując, badanie ISSIS-2 wykazało, że w grupie pacjentów z podjętą interwencją redukcja ryzyka względnego wyniosła 21% (obliczone wcześniej na podstawie Tabeli 2). Z kolei analiza podgrup (Tabele 3 i 4) pozwoliła stwierdzić, że wśród tych pacjentów, którzy są urodzeni pod znakiem bliźniąt lub wagi, terapia *zwiększyła* ryzyko względne wystąpienia zdarzeń niekorzystnych o 8% w stosunku do grupy kontrolnej. Dużo korzystniejsze okazało się natomiast przyjmowanie aspiryny wśród osób urodzonych pod pozostałymi znakami zodiaku, nastąpiła bowiem redukcja ryzyka względnego o 25%. Osoby spod znaków bliźniąt i wagi nie powinny zatem prewencyjnie zażywać aspiryny, ponieważ zwiększa to u nich (o 8% w stosunku do grupy kontrolnej) ryzyko wystąpienia zawału serca.

Wnioski wyprowadzone z tej analizy podgrup sugerujące wpływ znaków zodiaku na efekty terapii nie budzą jednak zaufania. Autorzy badania przedstawili je celowo, aby pokazać, że analiza podgrup nie jest metodą w pełni wiarygodną (Sun i in. 2014). Spełniony został warunek istotności statystycznej —  $P(r|t) > P(r)$  — jednak nie jesteśmy na tej podstawie skłonni akceptować wniosków ustalających zależności przyczynowe. Co jest powodem niskiej oceny wiarygodności tego badania? Dlaczego mielibyśmy uznać, że wnioskowanie o czynnikach ryzyka na podstawie statystycznie istotnych różnic w grupach osób urodzonych pod różnymi znakami zodiakalnymi jest czymś absurdalnym? Są to w istocie pytania o wiarygodność i poznawczą niezależność wnioskowań statystycznych przeprowadzanych na podstawie analizy podgrup.

## 5. PROBLEM WIARYGODNOŚCI WYNIKÓW ANALIZY PODGRUP

Wysokie koszty badania RCT sprawiają, że wiele wyników uzyskiwanych z tych badań jest dodatkowo opracowywanych z wykorzystaniem analizy podgrup (Parker, Naylor 2000, Sun i in. 2012: 3). Mimo stosunkowo dużej powszechności tej metody podaje się argumenty podważające wiarygodność formułowanych za jej pomocą wniosków. Można podzielić je na dwie grupy. Pierwsza obejmuje zastrzeżenia dotyczące błędów metodyki prowadzonych

analiz podgrup lub ich interpretacji. Ogólnie można powiedzieć, że błędy metodyki to błędy zaniechania lub braku dokładności badania. Nie są one tu dla nas interesujące.

Druga grupa obejmuje zarzuty, które można określić mianem epistemologicznych. Wskazują one, że badacze, dokonując analizy podgrup, przeceniają ich wartość poznawczą. Uznają, że metoda ta umożliwia ustalanie zależności przyczynowych, podczas gdy struktura tego badania na to nie pozwala. Argumenty o charakterze epistemologicznym, wskazując ograniczenia w stosowaniu analizy podgrup, dookreślają w istocie funkcję tej metody jako narzędzia poznawczego nauk biologiczno-medycznych. Przyjrzyjmy się zatem tym ograniczeniom.

#### 5.1. BRAK BIOLOGICZNEGO UZASADNIENIA DLA ROZWAŻANYCH W ANALIZIE PODGRUP CECH

Jednym z częściej formułowanych argumentów jest twierdzenie, że błędne wnioski z analizy podgrup są wynikiem braku biologicznego związku między wskazanymi w analizie cechami a efektami terapii. W analizie podgrup powinny być uwzględniane wyłącznie te własności pacjentów, których wpływ na przebieg terapii ma uzasadnienie w naukach biologicznych. Jak pisze Yusuf ze współpracownikami (1991: 95):

wierzmy, że hipotezy stanowiące podstawę analizy podgrup powinny mieć silne racjonalne uzasadnienie biologiczne (*strong biological rationale*).

Z kolei zespół Hernándeza (Hernández i in. 2006: 259) podkreśla, że „należy być sceptycznym, [...] jeżeli wyróżnienie podgrup nie jest biologicznie wiarygodne”. Spełnienie warunku biologicznego uzasadnienia hipotez wyeliminowałoby analizy określające np. wpływ znaków zodiaku na efekty terapii.

Niektórzy autorzy zauważają, że:

z uwagi na biologiczną różnorodność pacjentów zadziwiający byłoby, gdyby [w analizie podgrup] nie udało się znaleźć statystycznie istotnych zależności między efektami terapii a cechami charakteryzującymi pacjentów (Oxman, Guyatt 1992: 79).

Ta ostatnia uwaga wskazuje, że problemem nie jest w istocie zidentyfikowanie cechy biologicznej różnicującej pacjentów w sposób statystycznie istotny, lecz uzasadnienie, że jest ona biologicznie istotna dla efektów terapii. Pora roku, w jakiej przychodzimy na świat, a tym samym znak zodiakalny, może być uznana za cechę biologicznie różnicującą, wątpliwość może budzić jednak ustalenie jej przyczynowej doniosłości dla efektów terapii jedynie na podstawie uzyskania wyników istotnych statystycznie. Widać tu zasadniczą słabość omawianego zarzutu. Z jednej strony, akceptując postulaty epistemiczne



EBM, uznajemy, że badania biomedyczne dostarczają bardziej wiarygodnych danych niż badania podstawowe. Z drugiej, domagamy się, aby w analizie podgrup były brane pod uwagę wyłącznie te cechy, które mają biologiczne uzasadnienie, czyli uzasadnienie rozpoznawane na gruncie postulatów EBM jako najmniej wiarygodne.

Statystyczne zróżnicowanie cech biologicznych pacjentów nie zawsze może być podstawą do formułowania wniosków o ich wpływie na efekty terapii. Jeżeli jednak nie zawsze, to kiedy może? Dlaczego w innych badaniach biomedycznych wnioski formułowane na podstawie istotności statystycznej są rozpoznawane jako wiarygodne, a w analizie podgrup traktowane mają być z dużą dozą ostrożności?

#### 5.2. UPORCZYWE WYSZUKIWANIE ZWIĘKSZA PRAWDOPODOBIENSTWO ZNALEZIENIA STATYSTYCZNIE ISTOTNYCH RÓŻNIC

Kolejny argument podważający wiarygodność wyników uzyskiwanych z analizy podgrup związany jest z następującym faktem dotyczącym wnioskowań statystycznych:

(F) Im większa liczba czynników będzie brana pod uwagę w analizie podgrup, tym większa szansa na wyszukanie cech różnicujących efekty terapii w podgrupach w statystycznie istotny sposób (Oxman, Guyatt 1992: 81, Assmann i in. 2000: 1066, Wedeli in. 2001: 505, Yusuf i in. 1991: 94).

Jeżeli więc badacz bardzo chce znaleźć cechy różnicujące w sposób statystycznie istotny efekty terapii, to prawdopodobieństwo ich wskazania jest tym większe, im więcej cech będzie brał pod uwagę w analizie podgrup. Błędem byłoby jednak wnioskowanie, że cechy te warunkują przyczynowo efekty terapii. Można powiedzieć, że (F) podważa poznawczy sens działań motywowanych dewizą „szukajcie, a znajdziecie” i ogranicza w tym zakresie aktywność poznawczą badaczy określoną mianem *fishing expedition* (Oxman, Guyatt 1992: 78). Podkreślić należy, że (F) jest bardzo ważnym elementem argumentacji podważającej wiarygodność wyników analizy podgrup.

W badaniach statystycznych stosuje się jednak różne metody ograniczające niekorzystne konsekwencje (F). Pozwala to na uwzględnienie w końcowych wynikach badań liczby cech branych pod uwagę i proporcjonalne do niej zmniejszenie mocy testu, a więc liczby wykrywanych korelacji.

### 5.3. FORMUŁOWANIE HIPOTEZ *POST HOC* W ANALIZIE PODGRUP OBNIŻA WIARYGODNOŚĆ WNIOSKÓW

Wielu autorów utrzymuje, że wyniki dostarczane w analizie podgrup mają niski poziom wiarygodności, ponieważ sformułowanie hipotez wyróżniających czynniki różnicujące podgrupy często następuje dopiero po ukończeniu badania RCT. Innymi słowy, dopiero dysponując wynikami badania RCT, badacz decyduje się na sformułowanie hipotez o tych cechach pacjentów, które różnicują efekty terapii. Zdaniem krytyków wyniki analizy podgrup byłyby wiarygodniejsze, gdyby badacz wyróżnił cechy uwzględniane w analizie podgrup już na poziomie protokołu badania RCT. Dlatego oceniając wiarygodność analizy podgrup, dokonuje się podziału wyników na te uzyskane dzięki sformułowaniu hipotez *a priori* oraz te, w których hipotezy formułowane są *post hoc*. W pierwszym wypadku analiza podgrup jest zaplanowana już na etapie tworzenia protokołu RCT, w drugim – po uzyskaniu wyników badania RCT<sup>13</sup>. Zarzut ten podważa zatem status analizy podgrup jako niezależnej metody badawczej. Podkreślić należy, że argument wskazujący na błąd przyjmowania hipotez *post hoc* formułowany jest niezależnie od dwóch wcześniejszych argumentów, tj. argumentu z istotności biologicznej oraz faktu (F), jakkolwiek nie ulega wątpliwości, że występuje między nimi istotny związek.

Co jest jednak powodem krytykowania wiarygodności analizy podgrup opartych na hipotezach *post hoc*? Otóż analiza podgrup jest w tych przypadkach procedurą badawczą pozwalającą uzyskać nowe hipotezy. Za tego typu hipotezami nie stoi jednak żadne świadectwo niezależne od analizy podgrup, która stanowiła w istocie podstawę ich sformułowania. W takiej sytuacji nie można różnic statystycznych będących podstawą ich sformułowania traktować jako świadectw na ich rzecz.

Czy stosowanie analizy podgrup w celu uzyskania nowych hipotez badawczych jest uzasadnione? Można by przekonywać, że ze względu na fakt (F) takie postępowanie byłoby epistemologicznie nieprawomocne, ponieważ im więcej czynników będzie branych pod uwagę, tym większe prawdopodobieństwo znalezienia istotnych statystycznie różnic między efektami terapii w różnych podgrupach pacjentów. Z drugiej strony, można uznać, że analiza podgrup jest cenną metodą uzyskiwania hipotez, które oczywiście w dalszej kolejności powinny zostać poddane sprawdzeniu niezależnemu od metody ich pozyskania. Ignorowanie konsekwencji faktu (F) byłoby w tym wypadku realizacją

---

<sup>13</sup> Brak opisu, na którym etapie badania RCT formułowane są hipotezy, traktowany jest jako błąd metodyczny uniemożliwiający ocenę wiarygodności analizy podgrup (Assmann i in. 2000: 1066).

wytucznych pluralizmu teoriopoznawczego<sup>14</sup>. Postawione hipotezy mogą stanowić punkt wyjścia do dalszych badań w zakresie nauk podstawowych lub badań biomedycznych o większej wiarygodności, np. badań typu RCT.

#### 5.4. ANALIZA PODGRUP A BADANIE KLINICZNO-KONTROLNE

Przedstawiona w części 3 ogólna charakterystyka badań biomedycznych pozwala wskazać cechy, które wpływają na niską wiarygodność formułowanych w analizie podgrup wniosków:

- (i) jest to metoda o charakterze retrospektywnym: obejmuje dane uzyskane w przeszłości,
- (ii) prowadzone wnioskowanie ma charakter regresywny: wnioskujemy ze skutku o przyczynie,
- (iii) analiza podgrup jest wprawdzie oparta na wynikach badania RCT, ale czynnik interwencyjny nie jest wprowadzany intencjonalnie w przebiegu analizy podgrup, lecz w przebiegu poprzedzającego go badania RCT.

Z perspektywy metodologicznej jasne wydaje się zatem, dlaczego analiza podgrup traktowana jest jako metoda o niższej wiarygodności aniżeli badanie RCT. Dziwić może jednak inna kwestia. Otóż wszystkie wymienione cechy charakteryzują również badania kliniczno-kontrolne. Są to także badania nieeksperymentalne, retrospektywne, a prowadzone w nich wnioskowania mają charakter regresywny. Ponadto badaniom kliniczno-kontrolnym można postawić dokładnie te same zarzuty (1)-(3), które formułowane są wobec analizy podgrup.

Mimo podobieństw między analizą podgrup i badaniem CC, to jednak właśnie pierwsza z tych metod częściej uznawana jest jako źródło błędnych wniosków. W literaturze przeważa pogląd, że głównym powodem błędów analizy podgrup są uchybienia metodyczne. Wydaje się jednak, że można również wskazać na powody bardziej interesujące z perspektywy rozważań metodologicznych.

---

<sup>14</sup> W opisanym kontekście pluralizm teoriopoznawczy jest stanowiskiem postulującym wstępną akceptację wielu konkurencyjnych hipotez, które uzyskały potwierdzenie na podstawie analizy podgrup. O tym, że takie stanowisko jest w istocie przyjmowane w prowadzonych badaniach, świadczą najlepiej omówione wcześniej przykłady hipotez dotyczących wpływu takich czynników jak czas od wystąpienia zawału i wiek pacjentów na efekty terapii oraz płci w przypadku prewencyjnego stosowania aspiryny. Przyjęto wstępnie te hipotezy, a w dalszej kolejności w celu ich sprawdzenia przeprowadzono kolejne badania.

Po pierwsze, prowadząc wnioskowanie w analizie podgrup, uznajemy, że na obecny stan pacjentów (czyli S<sup>+</sup> lub S<sup>-</sup>) wpłynęły dwa czynniki występujące lub niewystępujące w ich historii przyczynowej: pierwszym jest interwencja podjęta w przebiegu badania RCT, drugim natomiast przypuszczalny czynnik ryzyka (płeć, wiek, znak zodiaku, czas od chwili zawału itd.). Nie wnioskujemy zatem z jedynej różnicy, jak to ma miejsce w badaniu CC.

Po drugie, w analizie podgrup wyznaczenie grupy kontrolnej nie jest zgodne z regresywnym charakterem prowadzonego wnioskowania. Wyjaśnijmy ten zarzut, odwołując się do charakterystyki badania CC. Otóż w badaniu tym wnioskujemy regresywnie od następstwa do racji i jednocześnie grupa kontrolna jest wyznaczona z uwagi na następstwo (część 2.1). Również w analizie podgrup wnioskujemy regresywnie od następstwa do racji, ale grupa kontrolna w tym badaniu nie jest określona ze względu na następstwo, lecz z uwagi na interwencję (por. Tabele 3 i 4). Mamy zatem do czynienia z rozbieżnością między dwoma cechami badania: kierunkiem wnioskowania i sposobem wyznaczania grupy kontrolnej. Wyznaczenie grupy kontrolnej w sposób uwzględniający kierunek wnioskowania ma na celu zwiększenie wiarygodności formułowanych wniosków (część 1, punkt 4), ponieważ warunek istotności statystycznej jest określony ze względu na grupy badaną i kontrolną. Zgodność kierunku wnioskowania ze sposobem wyznaczenia grupy kontrolnej jest zachowana zarówno w badaniach RCT, jak i w badaniach CC. Nie może być jednak zachowana w analizie podgrup.

## PODSUMOWANIE

Przeprowadzone rozważania pozwoliły wyodrębnić zasadnicze argumenty na rzecz ograniczeń poznawczych analizy podgrup. Dokonane ustalenia dopuszczają jej warunkowe stosowanie jako metody uzyskiwania hipotez badawczych, o ile przyjęta zostanie postawa pluralizmu teoriopoznawczego. Jednocześnie należy stwierdzić, że różnice między analizą podgrup a badaniem kliniczno-kontrolnym w sposobie wyznaczania grupy kontrolnej umniejszają wiarygodność formułowanych w analizie podgrup wniosków do poziomu, który wyklucza niezależność poznawczą tej metody badawczej. W gromadzeniu wiedzy w naukach biomedycznych może być ona traktowana wyłącznie jako jedna z metod wspomagających inne procesy badawcze. Uzyskiwane w niej wyniki wymagają bowiem dodatkowego uzasadnienia za pomocą bardziej wiarygodnych badań biomedycznych, np. badań eksperymentalnych.

## BIBLIOGRAFIA

- Assmann S., Pocock S., Enos L., Kasten L. (2000), *Subgroup Analysis and Other (Mis)uses of Baseline Data in Clinical Trials*, „The Lancet” 355(2909), 1064-1069.
- ATC. Antiplatelet Trialists' Colaboration (1988), *Secondary Prevention of Vascular Disease by Prolonged Antiplatelet Treatment*, „British Medical Journal” 296(6618), 320-331.
- Beaglehole R., Bonita R., Klejström T. (2002), *Podstawy epidemiologii*, Łódź: Instytut Medycyny Pracy im. Prof. J. Noefera.
- Brodniewicz T. (2015), *Słownik wybranych pojęć [w:] Badania kliniczne*, T. Brodniewicz (red.), Warszawa: CEDEWU, 27-32.
- Burke J., Sussman J., Kent D., Hayward R. (2015), *Three Simple Rules to Ensure Reasonable Credible Subgroup Analyses*, „British Medical Journal” 351, 1-5.
- Cartwright N. (1989), *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press.
- Cartwright N. (2010), *What Are Randomized Controlled Trials Good For?*, „Philosophical Studies” 147(1), 59-70.
- CCSG (1978), *A Randomized Trial of Aspirin and Sulfinpyrazol in Threatened Stroke. The Canadian Cooperative Study*, „New England Journal of Medicine” 299(2), 53-59.
- Concato J., Shah N., Horwitz R. (2000), *Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs*, „New England Journal of Medicine” 342(25), 1887-1892.
- Djubešovic B. (2000), *Evidentiary Challenges to Evidence-Based Medicine*, „Journal of Evaluation in Clinical Practice” 6(2), 99-109.
- GISSI. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (1986), *Effectiveness of Intravenous Thrombolytic Treatment in Acute Myocardial Infarction*, „The Lancet” 327(8478), 397-402.
- Hempel C. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York, NY: The Free Press.
- Hernández A., Boersma E., Murray G., Habbema J., Steyerberg E. (2006), *Subgroup Analyses in Therapeutic Cardiovascular Clinical Trials. Are Most of Them Misleading?*, „American Heart Journal” 151(2), 257-264.
- ISIS-2 (1988), *Randomised Trial of Intravenous Streptokinase, Oral Aspirin, Both, or Neither among 17 187 Cases of Suspected Acute Myocardial Infarction*, „The Lancet” 332(8607), 349-360.
- Jaeschke R., Cook D., Guyatt G. (1998), *Evidence based medicine (EBM), czyli praktyka medyczna oparta na wiarygodnych i aktualnych publikacjach (POWAP)*, „Medycyna Praktyczna”, 3(85), 147-1551.
- Jędrzychowski W. (2002), *Podstawy epidemiologii*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Kawalec P. (2006), *Przyczyna i wyjaśnianie*, Lublin: Wydawnictwo KUL.
- Lee T., Weisberg M., Brand D., Rouan G., Goldman L. (1989), *Candidates for Thrombolysis among Emergency Room Patients with Acute Chest Pain*, „Annals of Internal Medicine” 110(12), 957-962.
- Masębas W., Zastawny T. (2015), *Zarys procesu rozwoju leku [w:] Badania kliniczne*, T. Brodniewicz (red.), Warszawa: CEDEWU, 67-92.
- Oxman A., Guyatt G. (1992), *A Consumer's Guide to Subgroup Analyses*, „Annals of Internal Medicine” 116(1), 78-84.

- Papineau D. (1994), *The Virtues of Randomization*, „The British Journal of Philosophy of Science”, 45(2), 437-450.
- Parker A., Naylor C. D. (2000), *Subgroups, Treatment Effects, and Baseline Risks. Some Lessons from Major Cardiovascular Trials*, „American Heart Journal” 139(6), 952-961.
- Rzepiński T. (2012), *Ewidencja empiryczna w procesie podejmowania decyzji klinicznych*, „Filozofia Nauki” 20(3) [79], 13-33.
- Rzepiński T. (2014), *Randomized Controlled Trials versus Rough Set Analysis. Two Competing Approaches for Evaluating Clinical Data*, „Theoretical Medicine and Bioethics” 35(4), 271-88.
- Salmon W. (2006), *Four Decades of Scientific Explanation*, Pittsburgh, PA: University of Pittsburgh Press.
- Słowiński K. (1992), *Rough Classification of HSV Patients. [w:] Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, R. Słowiński (red.), Dordrecht: Kluwer, 77-93.
- Sun X., Briel M., Busse J., You J., Akl E., Mejza F., Bala M., Bassler D., Mertz D., Diaz-Granados N., Vandvik P., Malaga G., Srinathan S., Dahm P., Johnston B., Alonso-Coello P., Hassouneh B., Walter S., Heels-Ansdell D., Bhatnagar N., Altman D., Guyatt H. (2012), *Credibility of Claims of Subgroup Effects in Randomised Controlled Trials. Systematic Review*, „British Medical Journal” 12(344), e1553.
- Sun X., Joannidis J., Agoritsas T., Alba A., Guyatt G. (2014), *How to Use a Subgroup Analysis*, „Journal of American Medical Association” 311(4), 405-411.
- Urbach P., (1985), *Randomization and the Design of Experiment*, „Philosophy of Science” 52(2), 256-273.
- Wedel H., DeMets D., Deedwania P., Fageberg B., Goldstein S., Gottlieb S., Hjalmarsen A., Kjekshus J., Waagstein F., Wikstrand J. (2001), *Challenges of Subgroup Analyses in Multinational Clinical Trials. Experience from the MERIT-HF Trial*, „American Heart Journal” 142(3), 501-11.
- Worrall J. (2002), *What Evidence in Evidence-Based Medicine?*, „Philosophy of Science”, 69 (S3), 316-330.
- Worrall J. (2007), *Why There Is No Cause to Randomize?*, „British Journal for the Philosophy of Science” 58(3), 451-488.
- Yusuf S., Wittes J., Probstfield J., Tylor H. (1991), *Analysis and Interpretation of Treatment Effects in Subgroups of Patients in Randomized Clinical Trials*, „Journal of American Medical Association” 266(1), 93-98.