

Tomasz Rzepiński

Ekspertyza: Rodzaje badań epidemiologicznych

Spis treści

1. Cel ekspertyzy	4
2. Wprowadzenie.....	4
3. Badania obserwacyjne a badania eksperymentalne.....	5
Warunek korzyści praktycznych	7
Warunek uzasadnienia.....	8
Badania prospektywne vs. badania retrospektywne	8
Cele badania	8
Punkty końcowe badania.....	9
Niepożądane działanie leku.....	9
Kryteria włączenia i wykluczenia pacjentów z badania.....	10
Grupa kontrolna.....	11
Zaślepienie (<i>blinding</i>).....	12
Czynniki zakłócające (<i>confounders</i>)	12
Randomizacja i badania randomizowane	13
Ogólny podział badań eksperymentalnych.....	16
Klastrowe badania randomizowane (randomizacja klastrowa).....	17
Czynnikowe badania RCT (<i>factorial RCT</i>).....	17
Adaptacyjne badania eksperymentalne (<i>adaptive trials</i>)	18
Badania typu <i>non-inferiority</i> oraz badania równoważności.....	19
Badania typu <i>superiority</i> , równoważności oraz <i>non-inferiority</i>	21
Badania pragmatyczne.....	22
5. Rodzaje badań obserwacyjnych	23

Badania analityczne.....	24
Badania z doбором populacji według ekspozycji	25
Badania kohortowe prospektywne i retrospektywne.....	25
Progresywność i regresywność wnioskowań w badaniach epidemiologicznych.....	26
Badanie kohortowe prospektywne	27
Badanie kohortowe retrospektywne	27
Badania z doбором populacji według skutków ekspozycji.....	28
Badania przekrojowe	29
Badania kliniczno-kontrolne (<i>case-control studies</i>).....	29
Badanie kliniczno-kontrolne	31
Cechy badania kliniczno-kontrolnego istotne w ocenie komisji bioetycznych.....	33
Badanie kliniczno-kontrolne zagnieżdżone (<i>nested case-control study</i>)	33
Metaanaliza.....	34
Analiza podgrup w badaniach RCT i w badaniach kohortowych	35
Analiza podgrup	36
6. Strategie badań epidemiologicznych - podsumowanie	38
Strategie badań w epidemiologii	38
7. Ocena uzasadnienia dla projektu badań eksperymentalnych	40
8. Badania określające trafność testów diagnostycznych.....	43
9. Bibliografia.....	45

Indeks schematów

Schemat 1 – Model randomizacji prostej, równoległej	str. 15
Schemat 2 – Model randomizacji naprzemiennej, <i>cross-over</i> ,	str. 15
Schemat 3 – Ogólny podział badań eksperymentalnych	str. 16
Schemat 4 – Badania typu <i>superiority</i> , równoważności oraz typu <i>non-inferiority</i>	str. 21
Schemat 5 – Badania kohortowe prospektywne	str. 27
Schemat 6 – Badania kohortowe retrospektywne,	str. 27
Schemat 7 – Badania kliniczno-kontrolne	str. 31
Schemat 8 – Analiza podgrup	str. 36
Schemat 9 – Strategie badań w epidemiologii,	str. 38

1. Cel ekspertyzy

W niniejszym opracowaniu zaprezentowana zostanie charakterystyka różnych – pod względem metodologicznym – typów badań epidemiologicznych. Zasadniczym celem jest wskazanie tych cech badań epidemiologicznych, które powinny stanowić przedmiot szczególnej uwagi w procesie opiniowania tych badań przez komisje bioetyczne.

2. Wprowadzenie

Celem badań prowadzonych w epidemiologii jest uzyskanie wiedzy o wpływie różnych czynników i warunków środowiskowych na rozprzestrzenianie się chorób, jak również wiedzy o efektach podejmowanych działań terapeutycznych, diagnostycznych i prewencyjnych. Różne typy badań epidemiologicznych umożliwiają realizację odmiennych celów badawczych. Pewne badania lepiej nadają się dla oceny bezpieczeństwa stosowanej terapii, inne dla oceny jej skuteczności, a jeszcze inne dla oceny trafności testów diagnostycznych, czy stanu zdrowia populacji.

Zróżnicowanie procedur badawczych oraz realizowanych z ich wykorzystaniem celów poznawczych stawia zupełnie nowe wyzwania przed członkami komisji bioetycznych. Dla potrzeb oceny badania epidemiologicznego konieczne jest posiadanie coraz większych kompetencji w obszarze metodologii badań. W dobie bardzo szybkiego rozwoju nowych technologii oraz odkryć dokonywanych w naukach podstawowych: immunologii, genetyce, biochemii czy farmakologii, nie jest możliwe, aby wszyscy członkowie komisji bioetycznych posiadali wiedzę w obszarze wysoce wyspecjalizowanych dyscyplin badawczych. Wspólnego mianownika dla oceny tych merytorycznie zróżnicowanych badań winna dostarczać metodologia badań naukowych. Konieczne jest zatem pozyskanie podstawowych kompetencji w obszarze struktury badań epidemiologicznych, etapów rozwoju środka farmakologicznego oraz metod szacowania ryzyka. Niniejsze opracowanie jest poświęcone pierwszemu z tych obszarów. Celem jest omówienie w czytelny sposób ogólnej struktury badań epidemiologicznych, zarówno tych, którym członkowie komisji bioetycznych winni poświęcić więcej uwagi, jak i tych, których przeprowadzanie nie budzi zasadniczych wątpliwości natury etycznej. Liczę, że opracowanie dostarczy wskazówek umożliwiających lepszą orientację w strukturach badań epidemiologicznych, a co za tym idzie łatwiejszą identyfikację tych kwestii, które są ważne dla oceny badań z perspektywy komisji bioetycznych.

W niniejszym opracowaniu nie zostały bezpośrednio omówione zagadnienia związane z badaniami klinicznymi, fazami tych badań, czyli generalnie etapami rozwoju środka farmakologicznego od momentu identyfikacji cząsteczki biologicznie aktywnej do momentu rejestracji leku. Badania kliniczne stanowią podzbiór badań epidemiologicznych i dlatego można założyć, że pewna część ustaleń zaprezentowanych w niniejszym opracowaniu odnosi się również do badań klinicznych. Można nawet zaryzykować twierdzenie, że związek niektórych omawianych tu zagadnień z badaniami klinicznymi jest bardzo silny, np. podział badań na wyjaśniające i pragmatyczne, lub badania bez grupy kontrolnej, stosowane w zasadzie wyłącznie w I fazie badań klinicznych. Jednak są to zagadnienia, które musiały pojawić się w kontekście omawiania cech badań epidemiologicznych i dlatego też zostały ujęte w niniejszym opracowaniu.

3. Badania obserwacyjne a badania eksperymentalne

Zasadniczy podział badań epidemiologicznych to podział na badania eksperymentalne (interwencyjne) oraz badania obserwacyjne (nieeksperymentalne) (Jędrychowski, 2002; Beaglehole, Bonita, Kjellstrom, 2002). Różnica między nimi polega na kontrolowaniu podejmowanej w badaniu interwencji przez osobę wykonującą badanie.

Badania obserwacyjne to badania, w których różnice w podejmowanych interwencjach nie zostały wprowadzone w sposób intencjonalny przez badacza, lecz reprezentują pewien stan przezeń zastany. Są zwykle związane ze zwyczajowym stosowaniem przez różnych lekarzy różnych środków terapeutycznych, metod diagnostycznych lub też wcześniejszego wpływu czynników szkodliwych na pacjenta. Narażenie na działanie czynników szkodliwych może być związane z wykonywaną pracą (np. kontaktem ze środkami toksycznymi) lub trybem życia. Dlatego też w literaturze z zakresu epidemiologii stosowany jest zwykle termin: „ekspozycja” lub „narażenie” w miejsce węższego zakresowo terminu „interwencja terapeutyczna”.

Badania eksperymentalne natomiast to badania, w których interwencje są zastosowane wobec pacjentów w sposób **intencjonalny** przez badacza i mogą one różnić się od standardowych procedur, które są wykorzystywane w praktyce medycznej.

W celu wyraźnego zilustrowania różnic pomiędzy obydwoimi typami badań rozważmy następujące dwa przykłady. W pierwszym założmy, że badacz pragnie pozyskać informacje o różnicach w efektach stosowanych standardowo w leczeniu choroby D dwóch metod terapii: x

i y. Wie, że metoda x jest metodą z wyboru stosowaną w ośrodku zdrowia A, natomiast y jest metodą z wyboru w ośrodku B. Pozyskanie danych o charakterze poznawczym dotyczących efektów przeprowadzanych działań terapeutycznych w obu tych ośrodkach jest związane z uzyskaniem dostępu do rejestrów obu ośrodków. Osoba przeprowadzająca to badanie nie wdraża jednak terapii x i y w sposób intencjonalny w populacji pacjentów obu ośrodków, lecz analizuje pewien stan, który został przez nią zastany w obszarze praktyki medycznej.

Z odmienną sytuacją mamy natomiast do czynienia, gdy pragnąc porównać efekty stosowania dwóch standardowych metod terapii osoba przeprowadzająca badanie samodzielnie dokonuje przydziału pacjentów albo do grupy osób poddanych terapii x, albo do grupy osób poddanych terapii y. Interwencja terapeutyczna jest podjęta w tym przypadku intencjonalnie przez badacza w celu uzyskania określonych danych o charakterze poznawczym. Świadomość przeprowadzanej ingerencji decyduje o eksperymentalnym charakterze przeprowadzanego badania.

Różnica pomiędzy obydwoma typami badań pozostaje niezwykle istotna zarówno w wymiarze metodologicznym, legislacyjnym, jak i ze względu na prace komisji bioetycznych. Badania eksperymentalne dostarczają znacznie bardziej wiarygodnych informacji o skuteczności podejmowanych działań medycznych aniżeli badania obserwacyjne. Zgodnie z zaleceniami metodologii EBM (*Evidence-based medicine*, tzw. medycyna oparta na dowodach), badania eksperymentalne stanowią podstawę dla tworzenia wytycznych procedur decyzyjnych w praktyce medycznej.

Badaczem w eksperymentalnym badaniu epidemiologicznym może być wyłącznie lekarz lub lekarz dentyista posiadający odpowiednie dodatkowe kwalifikacje (Brodniewicz, 2015, s. 27 i 31). Z kolei w wielu badaniach obserwacyjnych, termin „badacz” jest używany w szerszym znaczeniu, osobą przeprowadzającą takie badania nie musi być lekarz.

Niezależnie od wymogu odpowiednio określonych kwalifikacji badacza, wykonywanie badań eksperymentalnych powinno pozostawać pod szczególną kontrolą komisji bioetycznych. Intencjonalny charakter podejmowanych przez badacza ingerencji wymaga od komisji bioetycznej ustalenia, czy w projekcie badania spełnione są dwa warunki reprezentujące system przekonań badacza. Nazwijmy je odpowiednio: warunkiem korzyści praktycznych i warunkiem uzasadnienia.

Warunek korzyści praktycznych

Badacz musi wykazać, że podejmowana przez niego w badaniu eksperymentalnym ingerencja pozwoli osiągnąć istotne cele poznawcze, które umożliwią realizację praktycznych działań o wysokiej skuteczności dla dobra przyszłych pacjentów. Oznacza to, że domniemane korzyści dla przyszłej praktyki medycznej wynikające z pozyskania wiedzy z badania eksperymentalnego muszą – w ocenie członków komisji bioetycznych – być wyższe aniżeli szacowane ryzyko narażenia zdrowia pacjentów biorących udział w takim badaniu.

Oczywiście nie jest możliwe bezpośrednie porównanie potencjalnych korzyści praktycznych z wartością ryzyka, z uwagi na ich niewspółmierność. Oznacza to, że nie jest możliwe zastosowanie narzędzia analizy, które na podstawie porównania obu tych wartości (korzyści praktycznych i ryzyka) pozwoliłoby generować normatywną regułę oceny projektu badania eksperymentalnego. Warto jednak zwrócić uwagę, że różne teorie podejmowania decyzji w sytuacji ryzyka dostarczają narzędzi, które umożliwiają, przy założeniu personalistycznej interpretacji prawdopodobieństwa, **zrekonstruowanie tych wartości ryzyka i wartości prawdopodobieństwa realizacji korzyści praktycznych**, które są przyjmowane przez osobę projektującą badanie eksperymentalne. Omówienie tych koncepcji nie byłoby przydatne dla celów realizowanych w niniejszym opracowaniu. Można jednak dokonać krótkiej rekapitulacji wniosków istotnych z perspektywy oceny projektu eksperymentalnego badania epidemiologicznego, w aspekcie oferowanych w tym badaniu potencjalnych korzyści praktycznych.

Ocena korzyści praktycznych, które potencjalnie mogą być uzyskane dzięki eksperymentalnemu badaniu epidemiologicznemu, musi obejmować między innymi wiedzę o:

- 1) częstości występowania choroby w populacji,
- 2) przewidywanych zmianach częstości występowania choroby, np. spowodowanych przewidywanymi zmianami stylu życia w społeczeństwie (choroby układu krążenia),
- 3) potencjalnych rezultatach rozwoju choroby dla indywidualnego pacjenta, jej przebiegu, rokowaniach, jakości życia po chorobie (podane wartości ryzyka wystąpienia istotnych zdarzeń klinicznych w przebiegu choroby),
- 4) konsekwencjach społecznych choroby, takich jak niezdolność do pracy czy do uczestniczenia w życiu społecznym i rodzinnym,

- 5) przewidywanych niepożądanych działaniach podejmowanej interwencji terapeutycznej,
- 6) dostępie do innych, stosowanych do tej pory metod terapii wraz z oceną ich skuteczności, ryzyka i rokowaniami dla pacjentów.

Warunek uzasadnienia

Po drugie, badacz musi dysponować uzasadnieniem, które uznaje za wiarygodną – w świetle obowiązującej metodologii badań naukowych w medycynie – przesłankę dla podejmowania badania eksperymentalnego. W szczególności oznacza to, że stopień uzasadnienia dla wykonania eksperymentalnego badania epidemiologicznego musi być znacznie wyższy aniżeli stopień uzasadnienia dla wykonania badania obserwacyjnego.

Kwestia uzasadnienia projektowanego badania eksperymentalnego zostanie przybliżona w dalszej części niniejszego opracowania (paragraf 7).

Badania prospektywne vs. badania retrospektywne

Badania prospektywne to badania zapoczątkowane przez zdarzenie, które ma miejsce obecnie i może stanowić potencjalną przyczynę przyszłego stanu zdrowia pacjenta lub grupy pacjentów. Badanie prospektywne jest prowadzone zatem od zdarzenia zaobserwowanego lub wywołanego w teraźniejszości (ekspozycji na czynnik szkodliwy, interwencji terapeutycznej) w przyszłość (skutek ekspozycji lub interwencji). Badaniami prospektywnymi są wszystkie rodzaje badań eksperymentalnych oraz niektóre z badań obserwacyjnych.

Badania retrospektywne to z kolei badania uwzględniające zdarzenia natury medycznej, które miały miejsce w przeszłości. Wykorzystuje się w nich dokumenty z minionych lat, historie chorób, karty szpitalne, roczniki statystyczne (Jabłonowski, 1996, s. 69).

Badaniami retrospektywnymi są wyłącznie niektóre typy badań obserwacyjnych.

4. Struktura badań eksperymentalnych

Cele badania

Najogólniej, badania eksperymentalne są przeprowadzane dla potrzeb realizacji dwóch celów: określenia skuteczności i bezpieczeństwa stosowanych środków farmakologicznych oraz skuteczności i bezpieczeństwa podejmowanych oddziaływań inwazyjnych w terapii i prewencji.

Punkty końcowe badania

Badania eksperymentalne mają na celu zidentyfikowanie zdefiniowanych wcześniej w protokole badania rezultatów podejmowanych ingerencji. W tym celu w protokole każdego badania eksperymentalnego określone zostają punkty końcowe (*endpoints, outcomes*). Są to zdarzenia natury medycznej, których wystąpienie pozwala oceniać efektywność ingerencji poddanej badaniu. Zdarzeniem natury medycznej może być określony stan zdrowia, jakość życia, wystąpienie określonej jednostki chorobowej, wznowienie choroby, zwiększenie lub zmniejszenie stężeń substancji biochemicznych w organizmie pacjentów. Należy podkreślić, że najczęściej w protokołach badań punktami końcowymi są niekorzystne dla zdrowia zdarzenia natury medycznej.

Wśród punktów końcowych badania wyróżnia się istotne klinicznie punkty końcowe oraz zastępcze punkty końcowe (*surrogate endpoints*) (Weintraub, Luscher, Pocock, 2015; Brody, 2016, s. 115-120; Evans, Ting, 2016, s. 75-80; FDA, 2018). Pierwsze z nich muszą być zdarzeniami natury medycznej istotnymi dla pacjenta, jego stanu zdrowia lub przebiegu choroby. Poza tym nie mogą stanowić subiektywnych odczuć pacjenta, takich jak stopień bólu, ilość zawrotów głowy itp. Przykładami istotnych klinicznie punktów końcowych są: zgon, zawał serca, udar, potrzeba transfuzji. Z kolei zastępcze punkty końcowe to takie, których wystąpienie pozwala wnioskować pośrednio o stanie zdrowia pacjenta. Przykładami takich punktów końcowych są wyniki badań biochemicznych, stężenie LDL, trójglicerydów, hemoglobiny, itp.

Przyjmowanie zastępczych punktów końcowych w badaniach epidemiologicznych budzi wiele kontrowersji. Ich omówienie jest jednak zbędne dla celów niniejszego opracowania. Dla potrzeb ocen dokonywanych przez komisje bioetyczne istotne są natomiast dwie kwestie.

Pierwszą jest metoda pomiaru punktu końcowego badania. Szczególną uwagę komisji opiniujących muszą bowiem zwracać te badania, w których punkty końcowe mierzalne są z wykorzystaniem diagnostyki inwazyjnej. Drugą ważną kwestią jest występowanie ubocznych, niekorzystnych efektów podejmowanych ingerencji.

Niepożądane działanie leku

Niepożądane działanie środka leczniczego (*adverse drug reaction – ADR*) to każde niekorzystne i niezamierzone działanie pojawiające się w efekcie stosowania dawek zalecanych

w celach profilaktycznych, diagnostycznych, lub leczniczych¹ (Maselbas, 2008; Wielahojewska, Orzechowska-Juzwenko, 2006). Z perspektywy komisji bioetycznej na podkreślenie zasługuje fakt, że występowanie wielu ADR-ów jest przewidywalne już na etapie projektu badania eksperymentalnego na podstawie wiedzy z zakresu nauk podstawowych, wiedzy o budowie cząsteczki leku, jego farmakodynamice i farmakokinetyce (Kaufman, 2016). Przewidywania dotyczące występowania lub braku ADR-ów są ważnym elementem oceny projektu badania eksperymentalnego. Komisja musi ocenić, czy oczekiwane korzyści wynikające z przeprowadzenia badania rekompensują ryzyko wystąpienia przewidywanych ADR-ów.

Całkowita ocena rezultatów badania eksperymentalnego tj. skuteczności i bezpieczeństwa ingerencji, obejmuje zarówno ocenę częstości występowania punktów końcowych, z wykorzystaniem parametrów redukcji ryzyka względnego i bezwzględnego, jak i ocenę częstości występowania ADR-ów z wykorzystaniem parametrów wzrostu ryzyka względnego i bezwzględnego. Omówienie tych wskaźników w niniejszym opracowaniu nie jest konieczne.

Kryteria włączenia i wykluczenia pacjentów z badania

Kryteria włączenia i wykluczenia pacjenta z badania to kryteria definiujące populację badaną w sposób pozytywny i negatywny. Celem badania jest określenie skuteczności i bezpieczeństwa podejmowanych oddziaływań terapeutycznych i prewencyjnych w określonej populacji: kobiet z nowotworem jajnika, osób z cukrzycą, mężczyzn z rakiem prostaty, itp. Kryteria włączenia pacjentów do badania są zatem kryteriami definiującymi populację badaną w taki sposób, aby była ona jak najbardziej reprezentatywna dla populacji docelowej, w której ma być stosowany dany lek lub zabieg. Kryteria wykluczenia to z kolei te, które wykluczają pacjenta z udziału w badaniu w przypadku wystąpienia pewnych dodatkowych cech, poza rozpoznaniem schorzeniem kwalifikującym do udziału w badaniach. Są to zatem często kryteria, które nie tylko określają jednostkę chorobową, ale mogą również uwzględniać stopień nasilenia choroby, czas jej trwania, wyniki określonych testów diagnostycznych, czy ocenę stanu pacjenta w badaniu fizykalnym z wykorzystaniem skal jakościowych (Brody, 2016, s. 83-130).

¹ Terminu „niepożądane działanie środka leczniczego” (ADR) nie należy mylić z terminem „zdarzenie niepożądane” (AE). Ten drugi pełni ważną funkcję w protokołach badań klinicznych, jednak dla celów obecnego opracowania nie ma potrzeby jego wprowadzania.

Zauważmy, że rozróżnienie pomiędzy obydwoimi typami kryteriów nie jest w pełni jasne z perspektywy metodologicznej. Kryteria włączania pacjentów do badania charakteryzują przecież cechy, których brak *de facto* wyklucza z udziału w badaniu. Przykładowo, rozważamy kryteria, które nakazują uwzględniać takie cechy pacjenta jak: stopień nasilenia danej choroby, czas jej trwania, wyniki określonych testów diagnostycznych, ocenę stanu pacjenta w badaniu fizykalnym, występowanie chorób współtowarzyszących, niepowodzenie lub sukces wcześniejszej terapii. Są to kryteria wyodrębniające w populacji pacjentów z daną chorobą pewną subpopulację. Kryteria te mogą być traktowane jako kryteria włączenia pewnego pacjenta do badania. Można jednak również uznać, że są zarazem kryteriami wykluczenia, w przypadku stwierdzenia u niego braku określonych cech.

Wydaje się, że najlepszym uzasadnieniem dla wyodrębniania w protokole kryteriów wykluczania pacjentów z badania nie są bezpośrednie wskazania merytoryczne, wynikające z realizowanych celów poznawczych, lecz kwestie etyczne. Zasadne jest podkreślenie przyjmowanych kryteriów wykluczenia pacjentów z badania wówczas, gdy wszyscy kandydaci spełniają kryteria scharakteryzowane dla populacji badanej z uwagi na cele poznawcze, lecz kwestie etyczne decydują o wykluczeniu z badania niektórych spośród tych kandydatów. Przykładem może być standardowe wykluczanie w protokołach badań eksperymentalnych dzieci poniżej 18 roku życia lub kobiet w ciąży².

Grupa kontrolna

Wśród badań eksperymentalnych wyróżniamy dwa typy: z grupą kontrolną i bez. W badaniach z grupą kontrolną następuje podział osób zakwalifikowanych do badania na grupę pacjentów otrzymujących nowy lek lub nowe leki oraz na grupę pacjentów przyjmujących terapię standardową lub otrzymujących placebo.

Wyróżnienie grupy kontrolnej ma umożliwić obliczenie wskaźników określających skuteczność nowych środków terapeutycznych tj. redukcji ryzyka względnego i redukcji ryzyka bezwzględnego (*relative risk reduction*, *absolute risk reduction*). Drugim ważnym celem

² Co skądinąd ma swoje bardzo negatywne konsekwencje w kwestii nie tylko realizowanych celów poznawczych, ale również samej praktyki medycznej. Przykładem mogą być doniesienia wskazujące na problem braku ocen terapii pacjentów pediatrycznych z uwagi na niewykonywanie badań z ich udziałem (Jędrzejowski, 2015). Podobny problem dotyczy braku danych z badań eksperymentalnych nad skutecznością i bezpieczeństwem środków farmakologicznych stosowanych przez kobiety w ciąży. Warto zwrócić uwagę w tym kontekście na wskazania CIOMS (2016).

poznawczym jest określenie ryzyka wystąpienia niepożądanego działania leków poprzez obliczenie, dla określonych zdarzeń klinicznych, wskaźników wzrostu ryzyka względnego i wzrostu ryzyka bezwzględnego. Badania eksperymentalne bez grupy kontrolnej nie pozwalają na określanie tego rodzaju wskaźników, ale też cel poznawczy realizowany w tych badaniach jest odmienny. Są to najczęściej badania kliniczne fazy 0 i I. Nie będą one jednak przedmiotem niniejszego opracowania.

Zaślepienie (*blinding*)

Badania eksperymentalne mogą podlegać zaślepieniu, które może mieć charakter zaślepienia pojedynczego, podwójnego lub potrójnego. Najogólniej, zaślepienie polega na utajnieniu tego, do jakiej grupy: badanej lub kontrolnej nastąpił przydział osoby uczestniczącej w badaniu. W badaniu zaślepionym pojedynczo o przydziale nie wiedzą albo uczestnicy badania, albo nie wie badacz. Szczególnym typem badania pojedynczo zaślepionego jest *assesor blind study*, czyli badanie zaślepienie dla kompetentnego oceniającego. Projektowanie takiego badania następuje w sytuacji, gdy nie jest możliwe zaślepienie uczestnika (np. badanie wpływu stosowanej diety na efekty zdrowotne) i oczekuje się, że obiektywna ocena stanu zdrowia uczestnika badania wymaga utajnienia przydziału wobec lekarza oceniającego efekty badania. W podwójnie ślepej próbie ani uczestnicy badania, ani badacz nie posiadają wiedzy o przydziale pacjentów. W potrójnie ślepej próbie zaślepieniu podlega cały personel łącznie z osobami dokonującymi analizy statystycznej wyników (Evans, Ting, 2016, s. 98-104).

Czynniki zakłócające (*confounders*)

W naukach zmatematyzowanego przyrodoznawstwa (fizyka, chemia), jak również w przypadku wykonywania doświadczeń na zwierzętach w naukach biologicznych możliwa jest pełna kontrola układu eksperymentalnego. Eksperymentator podejmuje pojedyncze ingerencje w układzie badanym i wie, że zastosowana przezeń ingerencja jest jedynym oddziaływaniem na przedmiot badany. W konsekwencji może z dużym stopniem pewności twierdzić, że zaobserwowany późniejszy stan układu eksperymentalnego jest skutkiem wprowadzonej przezeń intencjonalnie ingerencji. Pewności takiej nie można jednak posiadać w przypadku eksperymentalnych badań epidemiologicznych, w których przedmiotem badania jest populacja osób różniących się pod względem wielu cech istotnych dla wyników eksperymentu. Cechy te, mogące wpływać na przebieg podjętej ingerencji u poszczególnych pacjentów prowadząc do różnych efektów, określane są mianem „czynników zakłócających”.

Najogólniej, termin „czynniki zakłócające” (*confounders*) odnosi się do tych biologicznych i środowiskowych cech pacjentów, które mogą wpływać na przebieg podejmowanych w badaniach eksperymentalnych i obserwacyjnych ekspozycji (Jędrychowski, 2010, s. 95). Czynniki te mogą mieć charakter ujemny, działając antagonistycznie wobec czynnika badanego lub dodatni, działając z nim synergistycznie. Wiele z nich jest identyfikowanych w protokole badania poprzez odpowiednio sformułowane kryteria włączenia i wykluczenia pacjentów. Kryteria te pozwalają wyeliminować z badania osoby, których pewne cechy mogą w istotny sposób wpływać na przebieg badania. Definiują one docelową populację dla zastosowania danego leku.

Jednak nawet zawężenie, które uzyskujemy poprzez wprowadzenie kryteriów włączenia i wykluczenia pacjentów nie jest wystarczające, ponieważ pacjenci zakwalifikowani do badania cały czas różnią się dużą liczbą cech, a niektóre z nich mogą wpływać na przebieg terapii.

Dlatego też przyjmuje się, że wiarygodność eksperymentalnego badania epidemiologicznego zależy w zasadniczym stopniu od zdolności do równomiernego rozkładu czynników zakłócających w grupach badanej i kontrolnej, poprzez odpowiednio przeprowadzoną alokację pacjentów. Uznaje się powszechnie, że badaniem najpełniej realizującym ten cel jest badanie randomizowane.

Czynniki zakłócające można podzielić na znane i nieznanne. Znane to te, wobec których istnieje uzasadnione przekonanie, że mogą potencjalnie wpływać na przebieg podejmowanych w badaniu eksperymentalnym ingerencji. Uzasadnienie to jest uznawane za wiarygodne w świetle najlepszej dostępnej aktualnie wiedzy z zakresu nauk biologiczno-medycznych. Czynniki nieznanne to z kolei te, dla których takiego uzasadnienia nie posiadamy i w konsekwencji wszelkie próby identyfikowania ich jako czynników wpływających na przebieg terapii byłyby niewiarygodne. Nieracjonalnie byłoby jednak uznawanie, że na przebieg terapii wpływają tylko czynniki znane. Zakłada się zatem, że istnieją takie, które nie są brane pod uwagę jako potencjalnie modyfikujące efekty terapii.

Randomizacja i badania randomizowane

Wśród badań eksperymentalnych z grupą kontrolną wyróżnić można randomizowane kontrolowane badania kliniczne (*randomized controlled trials* - RCT) oraz badania nierandomizowane.

Podkreślić należy, że badań nierandomizowanych z grupą kontrolną praktycznie obecnie się nie przeprowadza. Koszty generowane przez badania z grupą kontrolną z udziałem dużych populacji pacjentów są tak bardzo wysokie, że nie warto obniżać wiarygodności uzyskiwanych wyników poprzez brak randomizacji.

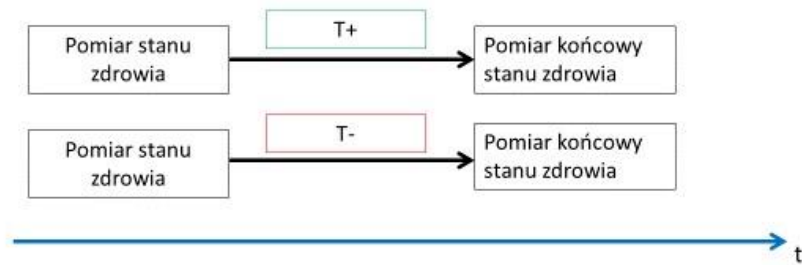
Zgodnie z metodologią EBM, badania RCT są traktowane jako dostarczające najbardziej wiarygodnych danych spośród wszystkich typów badań epidemiologicznych. Na rzecz tej tezy formułowane są dwa główne argumenty. Po pierwsze, randomizacja gwarantuje równomierny rozkład czynników zakłócających (znanych i nieznanych) lub przynajmniej zwiększa prawdopodobieństwo takiego rozkładu. Po drugie, randomizacja umożliwia zaślepienie przydziału i wyeliminowanie wpływu podmiotu na alokację pacjentów, dzięki czemu unika się błędów. Podkreślić należy, że efektem randomizacji może być podział populacji na dwie grupy (badaną i kontrolną) lub wyróżnienie grupy kontrolnej i kilku grup badanych (*multi-arm RCT*).

Wśród badań z randomizacją wyróżniamy badania z randomizacją prostą i randomizacją złożoną³. Randomizacja prosta może mieć charakter badania równoległego (ciągły) lub typu *cross-over* (naprzemiennego) (Jędrychowski, 1999, s. 95-96). W pierwszym przypadku populacja osób zakwalifikowanych do badania podzielona zostaje na grupę kontrolną i grupę/y badaną/e. W drugim, dany pacjent zakwalifikowany do badania otrzymuje naprzemiennie nowy lek i placebo / terapię standardową. W przypadku randomizacji naprzemiennnej ważne jest utrzymanie tzw. okresu wypłukania (*wash-time period*) pomiędzy naprzemiennymi dawkowaniami w celu uniknięcia błędów detekcji (np. nałożenia na siebie efektów stosowanych środków lub detekcji pomiaru punktu końcowego w grupie placebo). Poniżej zaprezentowane są poglądowe schematy obu typów randomizacji.

³ Osobnym typem są tzw. randomizowane badania klastrowe.

Schemat 1

Model randomizacji prostej, równoległej:



T+ to interwencja

T- to brak interwencji

- Stałe rozróżnienie grupy kontrolnej i poszczególnych grup badanych,
- Kohorty są poddane „badaniu” w tym samym czasie

Schemat 2

Model randomizacji prostej, układ naprzemienny – cross-over (szeregowy)



T+ to interwencja

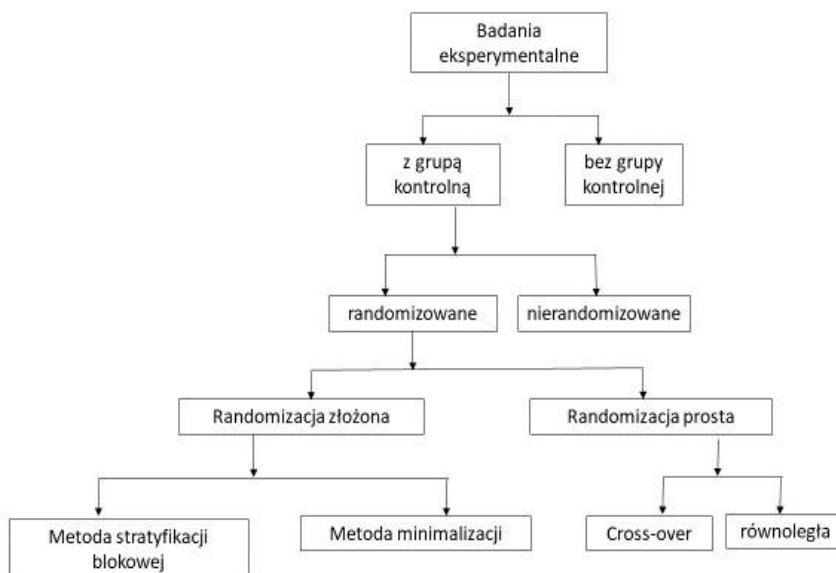
T- to brak interwencji

Naprzemienne podawanie leku i placebo lub standardowej terapii.

Randomizacja złożona to proces alokacji pacjentów do grup badanych, który umożliwia kontrolowanie wybranych (najczęściej nie więcej niż od 2 do 3) znanych czynników zakłócających. Taka randomizacja gwarantuje zatem bezdyskusyjnie (w przeciwieństwie do randomizacji prostej), że w grupie kontrolnej i w grupach badanych wystąpi taka sama liczba pacjentów ze zidentyfikowaną cechą zakłócającą. Stosowana jest w sytuacjach, gdy badacze posiadają **bardzo dobre uzasadnienie** dla przekonania, że określony czynnik/cecha wpływa na przebieg terapii. Wskazać można dwie podstawowe metody randomizacji złożonej: metoda minimalizacji i metoda stratyfikacji blokowej. Omówienie ich w niniejszym opracowaniu nie wniosłoby nic istotnego z perspektywy ocen badań dokonywanych przez członków komisji bioetycznych. W celu zilustrowania omówionych zagadnień w poniższym schemacie zaprezentowano ogólny podział badań eksperymentalnych.

Schemat 3

Ogólny podział badań eksperymentalnych



Klastrowe badania randomizowane (randomizacja klastrowa)

Badania klastrowe (*cluster randomized trial – CRT*) są to badania, w których randomizacja jest procedurą przeprowadzaną na grupach (klastrach), a nie na pojedynczych pacjentach. Innymi słowy, przydział losowy dotyczy wyszczególnionych klastrow, które są przydzielone do grupy badanej lub grupy kontrolnej. Uzasadnieniem poznawczym dla przeprowadzenia badań klastrowych jest wiedza o ewentualnym oddziaływaniu członków grupy na siebie wzajemnie, w efekcie czego może następować zaburzenie rezultatów badania (Evans, Ting, 2016, s. 97).

Przykładowo, trudno byłoby przebadać wpływ diety na występowanie określonych zdarzeń medycznych wówczas, gdy członkowie tego samego gospodarstwa domowego mieliby odmienne zalecenia dietetyczne. Łatwiej zachować rygory określonej diety w sytuacji jej narzucenia wszystkim członkom danego gospodarstwa domowego. W takiej sytuacji badanie, w którym zastosowano randomizację klastrową, dostarcza bardziej wiarygodnego rezultatu. Podkreślić należy, że w przypadku stosowania randomizacji klastrowej wyszczególniane są kryteria włączenia i wykluczenia zarówno dla indywidualnych pacjentów, jak i dla całych klastrow.

Podsumowując: przedmiotem procedury randomizacji mogą być albo indywidualni pacjenci, albo grupy pacjentów, w tym drugim przypadku mówimy o badaniach klastrowych.

Czynnikowe badania RCT (*factorial RCT*)

Badanie czynnikowe jest szczególnym przykładem badania, w którym wyróżniono kilka grup badanych (*multi-arm RCT*). Różnica polega jednak na tym, że w badaniu czynnikowym wyróżnia się grupy, które obejmują wszystkie możliwe wariacje wpływu rozważanych czynników. Jest to zatem badanie z pomocą którego można określić działanie dwóch lub większej liczby ingerencji, którym poddani są pacjenci zakwalifikowani do badania. Przykładem badania wieloczynnikowego jest badanie kliniczne HPS (*Herat Protection Study*), mające na celu ustalenie stosowania simwastatyny i antyoksydantów w prewencji chorób układu krążenia u osób z grupy ryzyka, ale z niskim stężeniem LDL. Wyróżnione zostały następujące grupy: simwastatyna + antyoksydanty, simwastatyna + placebo, placebo + antyoksydanty, placebo + placebo (Collins, Peto, Armitage, 2002).

Ogólny schemat badania czynnikowego 2 x 2 można zatem scharakteryzować następująco. Załóżmy, że badanie ma obejmować dwie ingerencje A i B. W klasycznym badaniu RCT typu *multi-arm* uzyskalibyśmy dwie grupy badane A i B oraz grupę kontrolną 0 (np. placebo). W badaniu czynnikowym uzyskamy cztery grupy: grupę osób A0, czyli takich, którym podano lek A oraz placebo leku B, grupę osób B0, czyli takich, którym podano lek B oraz placebo leku A, grupę AB, czyli osób, które uzyskały zarówno aktywny lek A jak i B, oraz grupę 00, czyli osób, którym nie podano żadnego aktywnego leku.

Bardzo istotne dla projektu badania czynnikowego RCT jest założenie, że pomiędzy podawanymi środkami farmakologicznymi nie zachodzą szkodliwe interakcje. Paradoksalnie natomiast jedną z korzyści takiego badania może być właśnie uzyskanie wiedzy o ewentualnych interakcjach podejmowanych terapii (McAlister et al., 2003).

Efekt terapeutyczny może się wzmacniać, czyli leki mogą wykazywać działania synergistyczne.

Badania czynnikowe odgrywają coraz ważniejszą rolę w sytuacji występowania chorób współistniejących. Ich zaletą jest to, że pozwalają ustalić w jednym badaniu zarówno efekty odrębnych oddziaływań, w grupach A0 i B0, jak i ich wspólny rezultat dla grupy AB. W badaniach tych wymagana jest jednak większa liczebność pacjentów dla potrzeb zachowania wiarygodności wniosków (Cipriani, Barbui, 2013; Montgomery, Astin, Peters, 2011; Evans, Ting, 2016, s.143-148).

Adaptacyjne badania eksperymentalne (*adaptive trials*)

Coraz częściej wśród badań naukowych stosowane są tzw. adaptacyjne badania kliniczne. Ponieważ są to badania eksperymentalne, których przebieg jest dosyć specyficzny, a jednocześnie istotny z perspektywy ocen dokonywanych przez komisje bioetyczne, idea tych badań zostanie bardzo krótko zaprezentowana w opracowaniu.

Najogólniej, badanie adaptacyjne to badanie, którego projekt pozwala na prospektywnie planowane **modyfikacje jednego lub większej liczby aspektów** projektu w oparciu o dane o pacjentach uzyskane w trakcie badania, w sposób nie umniejszający wiarygodności badania (FDA, 2019). Adaptacja oznacza w tym przypadku dopuszczalne w tym badaniu zmiany, jakie zostaną dokonane zarówno w zakresie **procedur badania** (jego przebiegu), jak i w zakresie zastosowanych **metod analizy statystycznej** (Chow, Chang, Pong, 2005). Poprzez adaptacje

procedur badania rozumie się zaplanowane zmiany w kryteriach włączania pacjentów, wielkości dawek środka leczniczego, czasu trwania leczenia, punktów końcowych, stosowanych testów i kryteriów oceny odpowiedzi klinicznej (Mahajan, Gupta, 2010).

Badania adaptacyjne określa się jako badania bardziej elastyczne niż klasyczne badania eksperymentalne przebiegające według tej samej procedury. Uważa się, że pozwalają one lepiej optymalizować korzyści, jakie pacjent może uzyskać z terapii z uwagi na analizę zmieniających się wielkości dawek, czasu leczenia i innych zmiennych aspektów procedury badania eksperymentalnego. Ponadto umożliwiają uzyskanie danych o podejmowanych interwencjach terapeutycznych w krótszym czasie i przy mniejszej kosztowności badania, niż badania tradycyjne. Nie może to być jednak zrealizowane za cenę obniżenia wiarygodności badania. Dlatego też duży nacisk kładzie się na prospektywność planowania zmian adaptacyjnych w badaniu. Cecha ta oznacza, że adaptacja zaprojektowana jest przed przeprowadzeniem jakichkolwiek analiz porównawczych danych gromadzonych z badania.

Adaptacja powinna być zaplanowana na etapie projektu badania i przewidziana w jego protokole lub przynajmniej musi być ustalona przed jego rozpoczęciem (FDA, 2019, s. 3). Te warunki mają na celu zachowanie wiarygodności badania.

Dla komisji bioetycznych najistotniejszą cechą jest zasadność planowanych zmian w badaniu. Bardzo ważne jest, aby właśnie na etapie jego projektowania wyraźnie określone zostało:

- (i) jakie zmiany w zakresie procedur badawczych są przewidywane,
- (ii) jakie czynniki mają inicjować wprowadzane modyfikacje badania (czy zmiana procedury ma nastąpić po określonym czasie trwania badania lub w efekcie uzyskania określonych rezultatów w danej grupie pacjentów),
- (iii) jakie jest uzasadnienie dla dokonywanych modyfikacji.

Badania typu *non-inferiority* oraz badania równoważności

Celem badań eksperymentalnych z grupą kontrolną jest najczęściej próba wykazania, że ingerencja eksperymentalna podjęta w grupie badanej (nowe leczenie) jest skuteczniejsza niż brak ingerencji. Następuje w nich porównanie grupy badanej z grupą kontrolną, w której

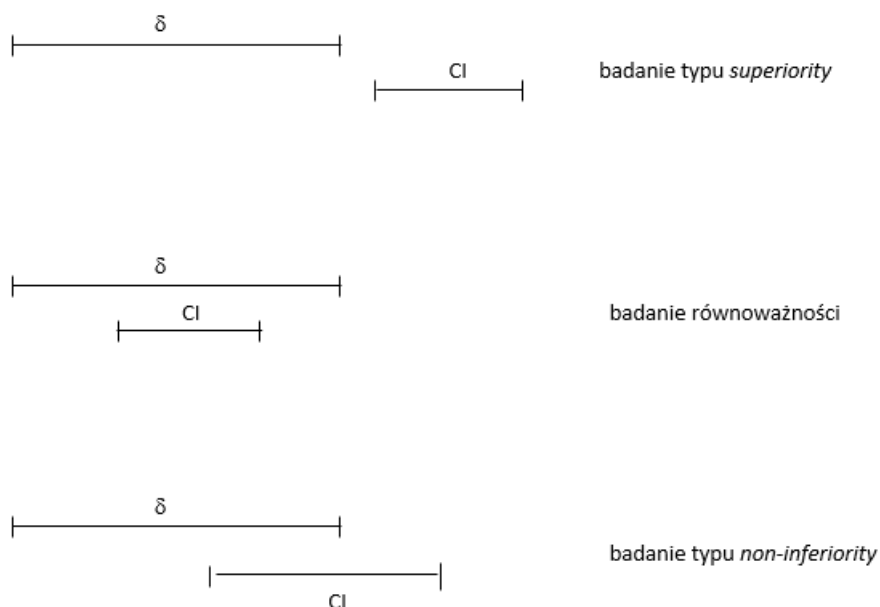
zastosowano placebo lub standardową metodę terapii. Większość badań eksperymentalnych z grupą kontrolną jest ukierunkowana na realizację takiego właśnie celu. Określane są one mianem *superiority trials* (Marko, Pawliczak, 2019).

Możemy mieć jednak również do czynienia z sytuacją, w której celem badania jest próba wykazania, że podejmowana ingerencja jest w przybliżeniu równoważna standardowej terapii, czyli działa niewiele gorzej lub niewiele lepiej niż ta użyta w grupie kontrolnej. Wówczas mamy do czynienia z tzw. *equivalence trials* (badaniami równoważności). Trzeci typ badań to badania, w których zakłada się, że testowana interwencja jest nie mniej skuteczna niż standardowa terapia zastosowana w grupie kontrolnej. Są to tzw. *non-inferiority trials* (badania nie mniejszej skuteczności) (Brody, 2016 s. 59; Leasffre, 2008; D'Agostino, Massaro, Sullivan, 2003). W niniejszym opracowaniu wskazane zostaną wyłącznie główne cechy badań typu *non-inferiority* i równoważności, które tworzą istotną perspektywę dla ocen dokonywanych przez komisje bioetyczne.

Różnica pomiędzy badaniami typu *non-inferiority* i badaniami typu *equivalence* dotyczy realizowanych w nich celów poznawczych. Jak wskazano wcześniej, badania równoważności mają na celu wykazanie, że obie terapie są równie skuteczne. Nie chodzi jednak o to, aby wykazać, że mają one dokładnie te same wartości skuteczności, lecz że ich skuteczności są podobne w pewnym określonym, akceptowalnym dla nas zakresie. Zakres ten określa się mianem marginesu równoważności (*equivalence margin*) i oznacza zwykle symbolem δ . Margines równoważności jest zatem maksymalnym zbiorem wartości, który identyfikuje akceptowalne klinicznie różnice skuteczności pomiędzy terapiami. Skuteczność terapii określa się z wykorzystaniem przedziału ufności (*confidence interval* – CI), czyli również pewnego przedziału wartości. Zatem, aby nowa terapia została oceniona w badaniu równoważności jako skuteczna, obie wartości graniczne przedziału ufności muszą mieścić się w przedziale δ . Jeżeli natomiast tylko jedna wartość przedziału ufności określonego dla potrzeb oceny skuteczności nowej terapii mieści się w przedziale δ , natomiast druga wykracza poza ten przedział, to mówimy o badaniach typu *non-inferiority* (Walker, Nowacki, 2010). Z kolei, gdyby obie wartości graniczne przedziału ufności wykraczały poza przedział δ , to mielibyśmy do czynienia z badaniami typu *superiority*, czyli badaniami, z którymi najczęściej kojarzymy realizowane cele badań eksperymentalnych. Różnice pomiędzy wymienionymi typami badań zostały zaprezentowane na poniższym schemacie.

Schemat 4

Badania typu *superiority*, równoważności oraz *non-inferiority*



CI – przedział ufności, δ - zakres marginesu równoważności

Z perspektywy oceny komisji bioetycznych istotne jest wskazanie na powody wykonywania badań równoważności i *non-inferiority*. Pierwszym, na który należy zwrócić uwagę jest fakt, że w przypadku wielu badań eksperymentalnych konstruowanie grupy kontrolnej z placebo byłoby nieetyczne z uwagi na dostępność zaakceptowanych już standardowych metod leczenia. W niektórych przypadkach wiedza, którą dysponujemy pozwala wnosić, że pozostawienie pacjentów bez leczenia w grupie z placebo prowadziło do ich śmierci lub znacząco obniżyłoby jakość lub długość ich życia. W tej sytuacji nowa terapia jest porównywana z pewnym akceptowalnym standardem leczenia, a celem badania nie jest wykazanie, iż nowa terapia jest lepsza, lecz że jest w jakimś zakresie równoważna lub nie gorsza od tego standardu. Takie sytuacje często mają miejsce w przypadku badania leków w obszarze onkologii. Inne powody przeprowadzania badań typu *non-inferiority* i równoważności wiążą się z przekonaniem, że nowa terapia wprawdzie nie jest lepsza od standardowej w realizacji jej głównych celów terapeutycznych, ale pozwala na szybsze wdrożenie leczenia; prowadzi do łagodniejszych powikłań lub mniejszej ich ilości; względnie jest tańsza.

Jednym z głównych problemów związanych z badaniami typu równoważności i badaniami typu *non-inferiority* jest ustalenie zakresu marginesu równoważności.⁴ Jeżeli zakres będzie zbyt wąski, to wówczas oczywiście niewiele terapii będzie spełniało warunki, które pozwoliłyby rozpoznać je jako równie skuteczne (równoważne) lub „niegorsze” od standardowej terapii (*non-inferior*). Jeżeli przyjęty zakres jest zbyt szeroki, to wówczas wiele terapii spełniałoby postawione warunki. Przyjmuje się, że badanie jest wiarygodne, jeżeli na etapie projektowania badania ustala się już zakres δ i podaje się dlań uzasadnienie, oparte najczęściej o praktykę kliniczną (Marko, Pawliczak, 2019).

Ocena i uzasadnienie wartości granicznych przedziału δ jest zatem kwestią kluczową z perspektywy poznawczej, ale posiada również istotne konsekwencje etyczne. Czym innym jest udział pacjentów w badaniu, którego celem jest wykazanie, że podejmowane terapie są równie skuteczne, a czym innym udział w klasycznym badaniu typu *superiority* lub badaniu równoważności. Podpisanie świadomej zgody wymaga uwzględnienia ich specyfiki. Tymczasem przeprowadzone analizy wykazują, że formularze świadomej zgody nie zawsze uwzględniają ten fakt (Marko, Pawliczak, 2019).

Badania pragmatyczne

Badania eksperymentalne można podzielić z uwagi na realizowane cele na badania wyjaśniające (*explanatory*) i badania pragmatyczne (*pragmatic*). Idea takiego rozróżnienia pojawiła się w latach 60-tych XX wieku (Schwartz, Lellouch, 1967). Podkreślić należy, że podział ten nie jest podziałem na badania realizujące cele poznawcze i praktyczne. Jest to raczej podział, który ma na celu wskazanie prób zwiększenia doniosłości klinicznej badań.

Otóż jednym z problemów wszystkich badań eksperymentalnych jest zaprojektowanie ich w taki sposób, aby populacja badana stanowiła jak najbardziej wierną reprezentację populacji docelowej, występującej w praktyce klinicznej (Ford, Norrie, 2016). Jednak kryteria włączenia i wykluczenia pacjentów z badania często definiują populację badaną w sposób, który może odbiegać w jakimś zakresie od populacji docelowej. Często populacja docelowa różni się od badanej czynnikami, których badacz nie mógł uwzględnić lub nawet takimi, których nie był świadomy. Przy dużej rozbieżności populacji badanej i populacji docelowej mamy do czynienia

⁴ Dla badań typu *non-inferiority* odpowiednio zakres ten określany jest mianem *non-inferiority margin*, czasami występuje też pojęcie „zakres porównywalności” (Marko, Pawliczak, 2019; Walker, Nowacki, 2010).

z problemem obniżonej wiarygodności zewnętrznej badania (*external validity*). Badania pragmatyczne mają na celu zniwelowanie powstałej w ten sposób luki poznawczej. Ich celem jest pozyskanie danych o efektach podejmowanej ingerencji (skuteczności i bezpieczeństwa) w populacji docelowej w praktyce klinicznej (Godwin, Ruhland, Casson, 2003; Rothwell, 2005).

Z uwagi na wskazane powyżej powody, badania pragmatyczne często obejmują bardziej złożone, wieloczynnikowe interwencje, podejmowane wobec osób z różnymi chorobami współistniejącymi, z udziałem personelu medycznego prezentującego różne umiejętności (pielęgniarki, chirurdzy, psychoterapeuci) i różne doświadczenie kliniczne. Z perspektywy komisji bioetycznych zasadniczym problemem, który pojawia się w kontekście badań pragmatycznych jest fakt, że ich protokół z definicji odbiega od protokołu badań wyjaśniających przeprowadzonych z udziałem bardziej jednorodnej populacji (Sugarman, Califf, 2014). W konsekwencji trudno ustalić, jakie korzyści zdrowotne i czynniki ryzyka, które zostały zidentyfikowane na podstawie badań wyjaśniających, będą zachowane w badaniach pragmatycznych. Na badania pragmatyczne należy zawsze zatem patrzeć jako na badania, które urealniamy. Im bardziej niejednorodna będzie populacja zakwalifikowana do badania pragmatycznego, tym większe ryzyko uzyskania wyników odstępujących od rezultatów badania wyjaśniającego.

5. Rodzaje badań obserwacyjnych

Wśród badań obserwacyjnych wyróżniamy dwa główne rodzaje: badania opisowe oraz badania analityczne.

Badania opisowe

Badania opisowe tworzą najbardziej podstawowy etap badań epidemiologicznych. Obejmują one opisy stanu zdrowia w danej populacji z uwzględnieniem zjawisk demograficznych i czynników środowiskowych (wiek, płeć), w tym również badania głównych wskaźników epidemiologicznych monitorujących stan zdrowia populacji: zapadalności, chorobowości, umieralności i śmiertelności. W badaniach tych można również identyfikować korelacje o charakterze statystycznym, poprzez porównywanie stanu zdrowia populacji w różnych regionach w tym samym czasie lub poprzez porównywanie stanu zdrowia populacji danego regionu w różnym czasie (Beaglehole, Bonita, Kjellstrom, 2002, s. 48).

Podkreślić należy, że badania opisowe nie określają, czy pomiędzy analizowanymi czynnikami środowiskowymi i demograficznymi oraz stanami zdrowia populacji identyfikowanymi za pomocą wskaźników epidemiologicznych zachodzą zależności przyczynowe. Identyfikacja takich zależności następuje dopiero w dalszych badaniach (Jędrychowski, 2010 s. 60). Przykładowo w 1981 roku opis kilku przypadków nietypowego zapalenia płuc stworzył podstawę dla rozwoju dalszych badań, które doprowadziły do identyfikacji nowej jednostki chorobowej (AIDS) (Gottlieb, 1981). Podobnie, doniesienia o zwiększającej się ilości przypadków fokomelii (niedorozwój kończyn wywołany zmianami genetycznymi lub działaniem teratogennym w życiu płodowym) u niemowląt (czyli zwiększająca się wartość współczynnika zapadalności) doprowadziły do sformułowania w 1961 roku hipotezy Lentza stanowiącej podstawę dla badań kliniczno-kontrolnych, w których zidentyfikowano talidomid zażywany przez kobiety w okresie ciąży jako czynnik przyczynowy fokomelii (Rzepiński, 2019).

Bardziej wyczerpująca charakterystyka wymienionych wskaźników epidemiologicznych stosowanych w badaniach opisowych nie jest istotna z perspektywy celów realizowanych w niniejszym opracowaniu. Badania opisowe nie budzą – poza jednym zastrzeżeniem, które zostanie wskazane w dalszej części opracowania (gdzie stanowią uzasadnienie dla badań eksperymentalnych) – kontrowersji natury etycznej.

Badania analityczne

Badania analityczne to najogólniej badania, których celem jest identyfikacja zależności przyczynowo-skutkowych pomiędzy zdarzeniami natury medycznej. Są to zatem badania, których cele poznawcze wykraczają znacząco poza cele realizowane w badaniach opisowych. Z tego powodu do badań analitycznych niektórzy autorzy zaliczają badania eksperymentalne, które w największym stopniu pozwalają na identyfikację zależności przyczynowo-skutkowych (Jędrychowski, 2002, s. 74). Biorąc jednak pod uwagę szczególny status poznawczy badań eksperymentalnych, jak również uwzględniając uwarunkowania etyczne tych badań, zostały one w niniejszym opracowaniu wyróżnione jako osobna kategoria opozycyjna wobec badań obserwacyjnych.⁵ Przyjmując takie rozstrzygnięcie można dokonać podziału obserwacyjnych

⁵ Podobne rozstrzygnięcie przyjmuje (Beaglehole, Bonita, Kjellstrom, 2002).

badan analitycznych na badania, w których dobór pacjentów następuje według (i) ekspozycji, lub (ii) skutków ekspozycji (Jędrychowski, 1999, s. 49).

Badania z doborem populacji według ekspozycji

W badaniach z doborem populacji według ekspozycji wyróżniamy dwa główne typy: badania kohortowe (*follow-up*) o charakterze prospektywnym oraz badania kohortowe retrospektywne, określane również jako kohortowe historyczne. Oba rodzaje badań są badaniami z grupą kontrolną.

Badania kohortowe prospektywne i retrospektywne

W prospektywnych badaniach kohortowych monitoruje się skutki ekspozycji, które mają miejsce obecnie i porównuje się je z efektami braku ekspozycji w grupie kontrolnej. Można zatem powiedzieć, że są to badania, które prowadzą od zdarzeń (ekspozycja lub jej brak) mających miejsce w teraźniejszości i prowadzą ku skutkom tych zdarzeń w przyszłości. Ponieważ są to badania obserwacyjne, zatem wybór populacji poddanej ekspozycji oraz wybór populacji niepoddanej ekspozycji nie leży w gestii badacza, lecz jest wynikiem zwyczajowo stosowanych procedur medycznych w różnych ośrodkach zdrowia lub przez różnych lekarzy.

Badania kohortowe są bardzo przydatne w określaniu stopnia natężenia ekspozycji na chorobotwórcze czynniki środowiskowe, w szczególności czynniki związane z wykonywanym zawodem, np. badania kohortowe nad występowaniem raka płuc u osób pracujących z azbestem lub górników (Jędrychowski, 2010, s. 100)

Często uważa się, że generalnie badania kohortowe mają charakter prospektywny, ale jest to – jak zauważają niektórzy autorzy – termin mylący. Termin „prospektywne” dotyczy czasu, w jakim miały miejsca dane zdarzenia medyczne, a nie związku między narażeniem i skutkiem. Zatem należy uznać, że badania kohortowe mogą być zarówno prospektywne, jak i retrospektywne (Beaglehole, Bonita, Kjellstrom, 2002, s. 55). Badania kohortowe retrospektywne (badania kohortowe historyczne) byłyby zatem badaniami, w których narażenie / ekspozycja miały miejsce w przeszłości, a dane o tych zdarzeniach zostały pozyskane z rejestrów szpitali i przychodni.

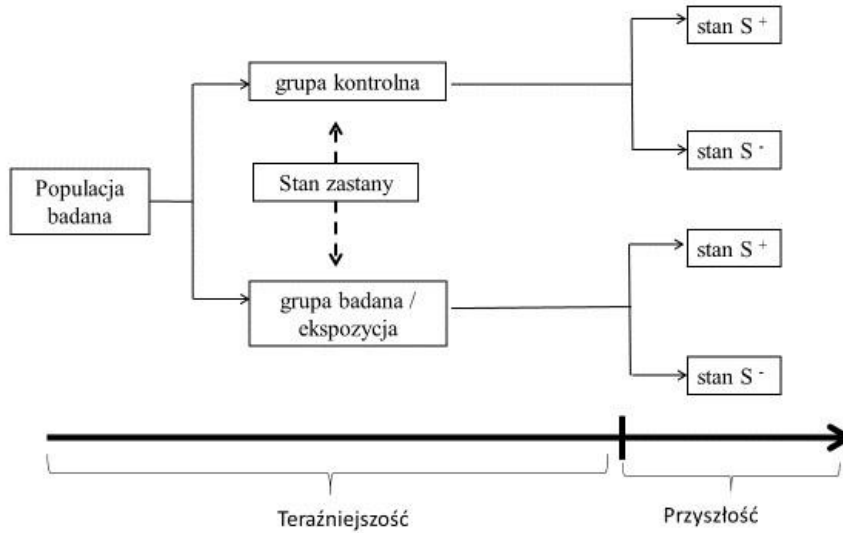
Progresywność i regresywność wnioskowań w badaniach epidemiologicznych

Należy zwrócić uwagę na fakt, że w obu typach badań kohortowych, zarówno prospektywnych jak i retrospektywnych, wnioskuje się od przyczyny (ekspozycji, narażenia) do skutku, pomimo że w badaniach prospektywnych przyczyna występuje w teraźniejszości, a w retrospektywnych – w przeszłości. W celu uniknięcia nieporozumień wskazywanych w literaturze epidemiologicznej (Beaglehole, Bonita, Kjellstrom, 2002) warto w charakterystyce badań epidemiologicznych, odróżnić – zgodnie z propozycją W. Galewicza – kierunek prowadzonego badania (prospektywny lub retrospektywny, w zależności od czasu, w jakim występuje ekspozycja) od kierunku prowadzonego wnioskowania. Wnioskowanie w badaniu może mieć charakter progresywny, czyli od przyczyny (ekspozycji) do skutku, lub regresywny: od skutku (np. choroby) do potencjalnej przyczyny.⁶ Zgodnie z dokonanym ustaleniem wnioskowania przeprowadzone zarówno w badaniach kohortowych retrospektywnych, jak i prospektywnych można zidentyfikować jako progresywne. Pomylenie kierunku prowadzonych wnioskowań z kierunkiem badania może skutkować nieporozumieniami w kwestii wiarygodności tych badań i nieścisłościami terminologicznymi. Poniżej zaprezentowane zostały schematy badań kohortowych z uwzględnieniem omawianych różnic.

⁶ Ma to szczególne znaczenie w charakterystyce porównawczej badań kohortowych i badań kliniczno-kontrolnych, na co zostanie zwrócona uwaga w dalszej części niniejszego opracowania.

Schemat 5

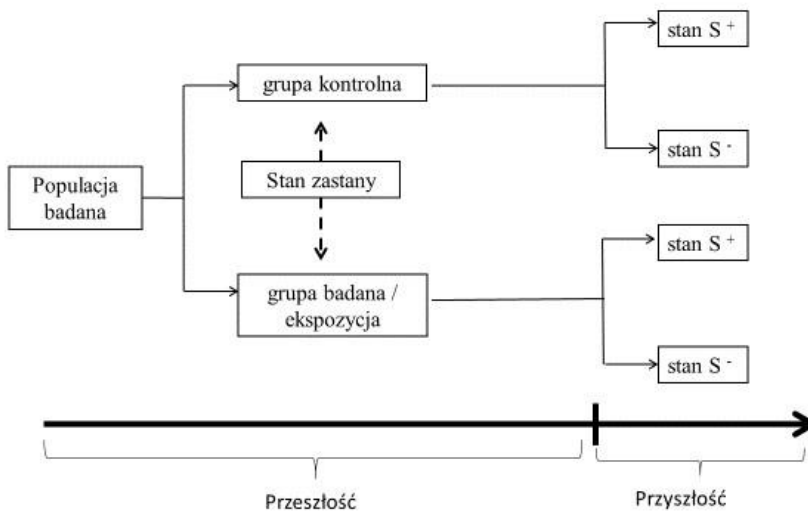
Badanie kohortowe prospektywne



Badanie prospektywne – z teraźniejszości w przyszłość
Kierunek wnioskowania progresywny – od ekspozycji do skutku: stan S⁺ lub S⁻

Schemat 6

Badanie kohortowe retrospektywne



Badanie retrospektywne – z przeszłości w przyszłość
Kierunek wnioskowania progresywny – od ekspozycji do skutku: stan S⁺ lub S⁻

Należy zwrócić uwagę na jeszcze jedną ważną kwestię: czynnik stanowiący ekspozycję w obu typach badania kohortowego może być albo określonym działaniem medycznym, albo czynnikiem o charakterze środowiskowym. W przypadku działania medycznego czynnikiem takim może być podjęcie terapii farmakologicznej, zabiegowej, działania prewencyjne (badania wpływu diety, ćwiczeń fizycznych itp.), poddanie określonym procedurom diagnostycznym. Wówczas będziemy mówili, że kohorta badana jest ustalana z uwagi na **ekspozycję medyczną**. W przypadku czynnika o charakterze środowiskowym analizowany może być stopień uprzemysłowienia miejsca zamieszkania, zanieczyszczenia tam występujące, wykonywany przez badanych zawód, klimat. Wówczas będziemy mówili, że kohorta badana jest ustalana z uwagi na **narażenia środowiskowe**.

Zaproponowane odróżnienie ekspozycji medycznej i narażenia środowiskowego nie jest istotne z perspektywy metodologicznej i z perspektywy projektowania badań epidemiologicznych. Wydaje się jednak, że może być istotne z perspektywy oceny badań dokonywanych przez komisję bioetyczną. Pozwala bowiem zwrócić uwagę na potencjalne różnice w sposobach pomiaru skutków ekspozycji medycznych i narażeń środowiskowych. W przypadku badań kohortowych, w których występują narażenia środowiskowe, wykorzystuje się zwykle standardowe procedury medyczne stosowane zwyczajowo w przypadku podejrzenia występowania danej choroby. W przypadku monitorowania skutków ekspozycji medycznej może zostać wykorzystana bardziej złożona diagnostyka medyczna. Przeprowadzanie takiej diagnostyki, wykraczającej poza standardowe postępowanie, jest w pełni zrozumiałe ze względów poznawczych, jednak należy mieć oczywiście świadomość, że zastosowane metody diagnostyki mogą tworzyć dodatkowe ryzyko dla pacjenta. Zagadnienie to zostanie bardziej obszernie scharakteryzowane w dalszej części opracowania w paragrafie 8.

Badania z doborem populacji według skutków ekspozycji

Wśród badań z doborem populacji według skutków ekspozycji zwyczajowo wyróżnia się trzy rodzaje: badania przekrojowe, badania kliniczno-kontrolne oraz badania kliniczno-kontrolne zagnieżdżone. Z uwagi na zadania realizowane w niniejszym opracowaniu, do wymienionych trzech rodzajów dołączone zostaną dwie metody analizy danych, mianowicie metaanaliza oraz analiza podgrup, której przedmiotem mogą być badania RCT, metaanalizy i badania kohortowe.

Badania przekrojowe

Badania przekrojowe to badania, w których rozpoznanie określonej jednostki chorobowej u pacjenta stanowi jednocześnie o pomiarze pewnych cech pacjenta, które potencjalnie mogły posiadać wpływ na rozwój danej choroby. Przykładowo, przyjęta zostaje hipoteza, zgodnie z którą niski poziom aktywności fizycznej wpływa na rozwój chorób układu krążenia. Dla potrzeb przeprowadzonego badania przyjęta zostaje typologia stopni aktywności fizycznej z uwagi na liczbę godzin aktywności tygodniowo. Rozpoznanie u danego pacjenta choroby układu krążenia stanowi jednocześnie o pomiarze, wykonanym za pomocą ankiety, typu aktywności fizycznej. Widzimy zatem, że w badaniu przekrojowym identyfikacja jednostki chorobowej, czyli pomiar skutku zapoczątkowuje pomiar domniemanej ekspozycji. Podkreślić należy, że w badaniu przekrojowym przedmiotem pomiaru ekspozycji mogą być zarówno cechy stałe, np. grupa krwi czy rasa, jak i cechy względnie stałe, czyli takie, które długotrwale charakteryzują pacjenta (Beaglehole, Bonita, Kjellstrom, 2002, s. 52). Warto podkreślić, że podstawą dla badań przekrojowych jest hipoteza sformułowana na podstawie wniosku regresyjnego: od zaobserwowanej jednostki chorobowej do domniemanej jej przyczyny. Natomiast nie są to badania, które w przyjętym wyżej sensie można by określić mianem retrospektywnych lub prospektywnych. Określenie „badania przekrojowe” najlepiej oddaje ich charakter.

Badania przekrojowe posiadają bardzo duże znaczenie w początkowym etapie badań epidemiologicznych jednostki chorobowej. Niemniej do wniosków, w których identyfikuje się zależności przyczynowe na podstawie tych badań, należy podchodzić z dużą dozą ostrożności, w szczególności w odniesieniu do pomiaru cech, które są względnie stałe. Problemem może bowiem być w tych przypadkach ustalenie, czy domniemana ekspozycja rzeczywiście poprzedzała skutek, czy też była wyłącznie korelacją współwystępującą ze zidentyfikowaną jednostką chorobową. Najczęściej zatem badania te są punktem wyjścia dla dalszych badań umożliwiających bardziej wiarygodną identyfikację związków przyczynowych.

Badania kliniczno-kontrolne (*case-control studies*)

Badania kliniczno-kontrolne to badania, w których punktem początkowym badania jest rozpoznanie jednostki chorobowej i następnie na podstawie sformułowanej hipotezy poszukuje się potencjalnego czynnika przyczynowego choroby w historii pacjenta. Przykładowo

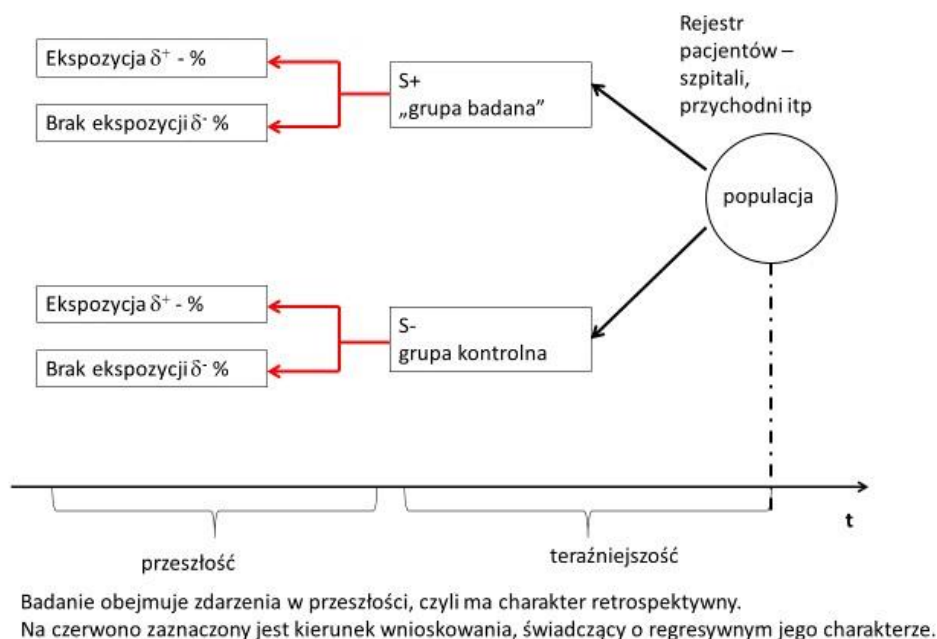
identyfikuje się populację osób z rozpoznaniem raka płuc i następnie sprawdza się, jak wiele z tych osób paliło papierosy lub jak wiele osób żyło w rejonach wysoce uprzemysłowionych o dużym stopniu zanieczyszczeń powietrza.

Badania kliniczno-kontrolne różnią się od badań przekrojowych tym, że identyfikacja potencjalnej przyczyny zdiagnozowanej choroby dotyczy zdarzeń, które miały miejsce w przeszłości lub zapoczątkowane zostały w przeszłości i trwają do dzisiaj. Badania takie są zatem badaniami retrospektywnymi. Przypomnijmy, że badaniami retrospektywnymi, ujmującymi zdarzenia z przeszłości, były również historyczne badania kohortowe. Należy jednak przypomnieć, że wnioskowania w tamtych badaniach nie mają charakteru regresywnego lecz progresywny, są prowadzone od ekspozycji do obserwowanych obecnie skutków. Z kolei w badaniach kliniczno-kontrolnych wnioskowania w nich przeprowadzane są wnioskowaniami regresywnymi: od obserwowanych skutków do domniemyanych ekspozycji.

Badania kliniczno-kontrolne są badaniami, w których występuje grupa badana (osoby ze zidentyfikowaną jednostką chorobową, S^+) oraz grupa kontrolna S^- , którą powinny tworzyć osoby, których charakterystyka (wiek, płeć, miejsce zamieszkania, praca, choroby współwystępujące) jest jak najbardziej zbliżona do charakterystyki osób z grupy badanej, poza tym, że są one zdrowe. Ten warunek przysparza zwykle najwięcej kłopotów, a jego niespełnienie może być źródłem błędnych wniosków. Poniżej przedstawiony jest schemat badania kliniczno-kontrolnego (Beaglehole, Bonita, Kjellstrom, 2002, s. 53).

Schemat 7

Badanie kliniczno-kontrolne



Zauważmy, że schemat badania kliniczno-kontrolnego jest pomyślany w taki sposób, aby w jak największym stopniu umożliwić identyfikację pewnych zdarzeń z przeszłości jako warunków wystarczających i koniecznych obserwowanego w teraźniejszości stanu chorobowego. Jeżeli dla każdego pacjenta z grupy badanej, można wskazać na wystąpienie w przeszłości ekspozycji δ^+ , to można wnosić, że czynnik ten jest warunkiem wystarczającym wystąpienia obserwowanej jednostki chorobowej, oczywiście abstrahując od zawodności wnioskowania indukcyjnego. Jeżeli natomiast dodatkowo udałoby się ustalić, że dla poprawnie dobranej grupy kontrolnej u żadnego z pacjentów z tej grupy nie występuje dany objaw chorobowy, to można wnosić, że δ^+ jest również warunkiem koniecznym wystąpienia danej choroby.

Badania kliniczno-kontrolne doskonale nadają się do identyfikowania czynników szkodliwych. Dobrym przykładem jest badanie kliniczno-kontrolne identyfikujące talidomid jako przyczynę fokomelii. Pierwsze doniesienia o zwiększonej liczbie przypadków fokomelii to końcówka 1959 roku. Nikt jednak nie łączył tych danych z zażywaniem talidomidu przez matki dzieci. Przełom nastąpił dopiero 18 listopada 1961 roku w trakcie konferencji pediatrycznej.

Kilku lekarzy opisało wówczas zaobserwowane przez nich przypadki fokomelii. Był to zatem pierwszy etap badań epidemiologicznych – badania opisowe przypadków i serii przypadków. Jeden z uczestników konferencji, Widukind Lentz zwrócił uwagę, że miał do czynienia również z przypadkami fokomelii, a ich cechą wspólną było zażywanie talidomidu przez matki dzieci. Hipoteza Lentza została przyjęta jako robocza hipoteza dla drugiego etapu badań epidemiologicznych – badań kliniczno-kontrolnych. Ich przeprowadzenie pozwoliło stwierdzić, że zasadnicza część kobiet, których dzieci urodziły się z fokomelią, zażywała talidomid (Rzepiński, 2019, s. 164).

Problem polega jednak na tym, że badanie kliniczno-kontrolne obciążone jest nie tylko błędami wynikającymi z zawodności wnioskowania indukcyjnego, ale również błędami o charakterze metodologicznym. Wskazuje się głównie na cztery rodzaje błędów metodologicznych, które mogą zostać popełnione w badaniu kliniczno-kontrolnym.

- 1) Wybór czynnika z przeszłości stanowiącego domniemaną ekspozycję. W wielu przypadkach można wskazać cechy, które charakteryzują pacjentów z grupy badanej, ale ich występowanie ma charakter korelacji statystycznej niezwiązanej przyczynowo z występowaniem choroby. Wprowadzenie grupy kontrolnej w badaniu zmniejsza wprawdzie szansę na wybór cechy przyczynowo nieistotnej, ale nie wyklucza takiej sytuacji. Podjęcie badania kliniczno-kontrolnego związane jest ze sformułowaniem hipotezy, że dana cecha z przeszłości jest przyczynowo istotna. Można w konsekwencji formułować kolejne hipotezy tak długo, aż uzyskamy różnice istotne statystycznie pomiędzy grupą badaną a grupą kontrolną. Projekt badania kliniczno-kontrolnego musi posiadać zatem **uzasadnienie**, nie tyle nawet w obszarze badań opisowych lub przekrojowych, lecz przede wszystkim pochodzące z ustaleń nauk podstawowych.
- 2) Drugi problem związany jest z uformowaniem grupy kontrolnej, na co wskazano już wcześniej. Dla każdego pacjenta z grupy badanej należy wskazać pacjenta, który pod względem charakterystyki będzie najbardziej doń zbliżony, ale nie będzie u niego występowała dana choroba. Jest to żmudne przeszukiwanie rejestrów i często nie kończy się satysfakcjonującymi wynikami. Różnice pomiędzy charakterystyką pacjentów z grup badanej i kontrolnej zmniejszają wiarygodność formułowanych w badaniu wniosków. Pacjenci z obu grup mogą się bowiem tak bardzo różnić historiami przyczynowymi, że wystąpienie danej choroby nie musi być spowodowane domniemaną ekspozycją, lecz

synergistycznym działaniem pewnych czynników, przy jednoczesnym antagonistycznym działaniu innych czynników u różnych pacjentów (Jędrychowski, 1999, s. 71; Jędrychowski, 2010, s. 83-85).

- 3) Trzeci problem związany jest z tzw. wypaczeniem Neymana w badaniu kliniczno-kontrolnym. Ma ono miejsce wówczas, gdy z uwagi na ustalony punkt końcowy grupa badana nie jest reprezentatywna dla całej populacji chorych. Przykładowo, założmy, że choroba D może w pewnych sytuacjach kończyć się zgonem pacjenta, a w innych pewnym zdarzeniem S+. W takiej sytuacji, jeżeli stanem inicjującym badanie jest zgon pacjenta w przebiegu choroby D, to okazuje się, że ci pacjenci, którzy wyzdrowieli, nie zostaną uwzględnieni i dołączeni do grupy badanej.
- 4) Problem pozyskiwania danych o ekspozycji / narażeniu. Bardzo często w badaniu kliniczno-kontrolnym nie korzysta się z rejestrów, lecz w procesie wywiadu należy uzyskać informację o tym, czy pacjent był poddany ekspozycji. Problem polega na tym, że pacjenci mogą nie pamiętać o takim fakcie. Przykładem może być ponownie badanie kliniczno-kontrolne identyfikujące talidomid jako czynnik odpowiedzialny za powodowanie fokomelii. Talidomid w wielu krajach był dostępny bez recepty. Pacjentki często nie pamiętały o jego zażyciu, szczególnie gdy fakt ten miał jednorazowy charakter. Często też, domyślając się, że jest to środek odpowiedzialny za wystąpienie fokomelii, starały się wyprzeć ten fakt z pamięci.

Cechy badania kliniczno-kontrolnego istotne w ocenie komisji bioetycznych

Badanie kliniczno-kontrolne, z uwagi na swój retrospektywny charakter i regresywne wnioskowanie nie tworzy szczególnych problemów natury etycznej w płaszczyźnie podejmowanych działań medycznych. Należy jednak zwrócić szczególną uwagę na jedną kwestię. Dotyczy ona sytuacji, w której diagnostyka stanu chorobowego S+ stanowiącego punkt wyjścia dla tego badania, wykraczałaby poza stosowane standardy praktyki medycznej i wymagałaby diagnostyki inwazyjnej lub zmniejszającej komfort funkcjonowania pacjenta.

Badanie kliniczno-kontrolne zagnieżdżone (*nested case-control study*)

Badanie kliniczno-kontrolne zagnieżdżone ma na celu identyfikację pewnych stałych cech pacjentów jako potencjalnych ekspozycji dla choroby zdiagnozowanej u tych pacjentów w przyszłości. Cechami stałymi identyfikowanymi w tym badaniu są najczęściej pewne

czynniki biochemiczne, jak np. stężenia określonych witamin we krwi, antyoksydantów, cholesterolu, itp. Próbkę krwi zostają zamrożone, a następnie po dłuższym czasie, np. kilkunastu latach można ocenić, w jaki sposób występowanie określonego czynnika stałego skorelowane jest z wystąpieniem określonej choroby. Badania te mają zatem stanowić mniej kosztowną alternatywę wobec badań kohortowych, w których konieczne byłoby monitorowanie pacjentów przez długi czas (Jędrychowski, 2010, s. 89).

Z perspektywy ocen dokonywanych przez komisje bioetyczne, badania kliniczno-kontrolne zagnieżdżone nie budzą zasadniczych kontrowersji.

Metaanaliza

Metaanaliza jest badaniem wtórnym, co oznacza, że stanowi jedną z procedur stosowanych dla potrzeb opracowywania wyników. W przypadku metaanalizy są to wyniki pochodzące z różnych badań, najczęściej eksperymentalnych RCT i kohortowych. Metaanalizy są również często wykorzystywane w opracowywaniu wyników uzyskanych w II fazie badań klinicznych, jednak to zagadnienie nie jest przedmiotem niniejszego opracowania. W metaanalizie następuje zestawienie wyników uzyskanych z wielu badań pierwotnych danego środka leczniczego lub interwencji, dzięki czemu zwiększa się wiarygodność uzyskanych wyników z uwagi na większą populację pacjentów.

Problem, na który należy zwrócić uwagę polega na tym, że badania pierwotne dostarczające materiału dla metaanalizy często przebiegają według różnych protokołów. Oznacza to, że różnice pomiędzy badaniami pierwotnymi mogą być bardzo duże. Badania te mogą uwzględniać odmienne kryteria włączenia pacjentów, ich wykluczenia, różne dawki środka farmakologicznego, różną częstotliwość jego podawania, różny stopień zaawansowania choroby stanowiącej wskazanie do podjęcia leczenia oraz nawet różne punkty końcowe badania. W takiej sytuacji ceną poznawczą, jaką płacimy przeprowadzając metaanalizę jest zmniejszenie precyzji formułowanych wniosków z uwagi na zróżnicowanie populacji pacjentów i z uwagi na różnice w operacjonalizowaniu badań pierwotnych.

Oczywiście poziom heterogeniczności metaanaliz może być różny. Niektóre z nich obejmują bardziej ujednolicone badania pierwotne, w innych uwzględnia się badania o bardzo dużej rozbieżności. Dąży się oczywiście do uwzględniania tych badań pierwotnych, które cechują się wspólnie jak najwyższą jednorodnością. Nie zawsze jest to jednak możliwe. Stopień

akceptacji poziomu niejednorodności metaanalizy zależy od realizowanych celów poznawczych. W badaniach chorób sierocych, rzadko występujących w populacji, metaanaliza jest jednym z podstawowych narzędzi opracowywania danych, nawet przy dużej niejednorodności badań pierwotnych.

Pomimo wskazanych problemów metaanalizy pełnią bardzo istotne funkcje poznawcze. Przede wszystkim są one narzędziami umożliwiającymi zsyntetyzowanie dotychczasowych wyników badań i ujęcie ich w spójną całość. Nie bez znaczenia dla wzrastającej popularności metaanaliz jest fakt, iż stanowią one również doskonałe narzędzie diagnozowania „wątpliwych praktyk badawczych” [*questionable research practices* - QRP]. Pozwalają bowiem zidentyfikować niespójności pojedynczych badań pierwotnych, które spowodowane są celowym oszustwem. Metaanaliza jest zatem swoistym „audytem badawczym nauki” (Karwowski, 2018).

Nie umniejszając wskazanych powyżej walorów, należy wskazać na cechę szczególnie istotną z perspektywy ocen badań dokonywanych przez komisje bioetyczne. Otóż wyniki metaanaliz mogą stanowić uzasadnienie dla formułowania nowych hipotez badawczych i w konsekwencji projektowania nowych badań naukowych, w tym również badań eksperymentalnych. Zasadne jest w takiej sytuacji pytanie o to, czy wyniki metaanalizy rzeczywiście można traktować jako wystarczająco dobre uzasadnienie dla projektu nowego badania eksperymentalnego. O ile zatem sama decyzja o przeprowadzeniu metaanalizy nie stanowi przedmiotu kontrowersji natury etycznej, o tyle już wyniki metaanalizy, jak również założenia dotyczące stopnia heterogeniczności przyjęte dla jej przeprowadzenia są dla komisji bioetycznych interesujące. To właśnie założenia metaanalizy są bowiem kluczowe dla sformułowania potencjalnych hipotez testowanych w późniejszych badaniach eksperymentalnych. Komisje bioetyczne muszą rozważyć, czy uzasadnienie dla badania eksperymentalnego pochodzące z metaanalizy o dużym stopniu niejednorodności jest wystarczające w świetle przewidywanego ryzyka wynikającego z eksperymentalnego charakteru projektowanego badania.

Analiza podgrup w badaniach RCT i w badaniach kohortowych

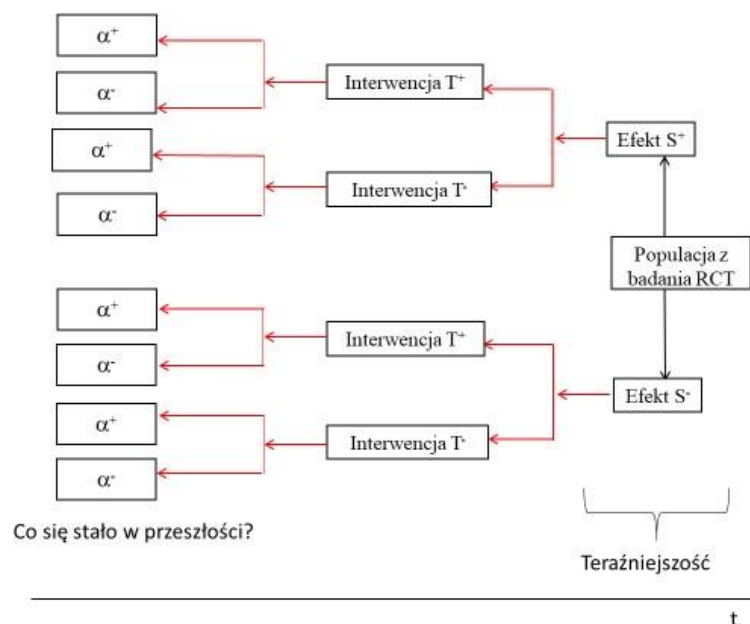
W literaturze z zakresu epidemiologii analiza podgrup (*subgroup analysis*) nie jest traktowana jako osobny rodzaj badania epidemiologicznego, lecz jako metoda opracowywania danych pozyskanych z innych badań, najczęściej RCT i kohortowych, jak również z metaanaliz

tych badań. Wydaje się jednak, że przybliżenie struktury tych badań może być w pewnych sytuacjach istotne z perspektywy ocen dokonywanych przez komisje bioetyczne. Zasadnicza idea tej metody nie różni się w wymienionych typach badań, dlatego też omówiona zostanie skrótowo w odniesieniu do RCT.

Idea analizy podgrup badania RCT jest stosunkowo prosta i opiera się na założeniu, że na przebieg interwencji medycznej wpływ mogą mieć pewne czynniki zakłócające. Ponieważ, jak wspomniano we wcześniejszej części opracowania, nie wszystkie te czynniki można kontrolować z wykorzystaniem metod randomizacji złożonej, zatem warto dokonać analizy po zakończeniu badania dla potrzeb ustalenia, czy określona cecha (α) pacjenta wpływa na efekty podjętej interwencji. W tym celu w grupach pacjentów z określonymi wynikami S+ oraz S- wyróżnia się tych pacjentów, którzy zostali poddani interwencji i posiadają daną cechę (α^+) lub jej nie posiadają (α^-). Schemat analizy podgrup RCT zaprezentowany jest poniżej.

Schemat 8

Analiza podgrup



Czerwone strzałki oznaczają kierunek wnioskowania.

Analiza podgrup jest zatem metodą, w której podejmowane jest wnioskowanie regresywne mające na celu identyfikację potencjalnego czynnika α wpływającego na rezultat terapii. Zauważmy, że hipotezę identyfikującą domniemany czynnik zakłócający przebieg

terapii można sformułować na etapie projektu badania RCT lub po jego zakończeniu, gdy znane są już rezultaty tego badania. W pierwszym przypadku mówimy o sformułowaniu **hipotezy poprzedzającej** badanie RCT, w drugim o hipotezie *post hoc* tego badania. Pomimo że analiza podgrup nie jest badaniem epidemiologicznym sensu stricto, i wykazuje wiele zbieżności z obserwacyjnym badaniem kliniczno-kontrolnym, to jednak jest ona niezwykle istotna z perspektywy ocen i monitorowania badań przez komisje bioetyczne.

W pierwszej kolejności należy zwrócić uwagę na fakt, że analiza podgrup dostarcza wyników nieporównywalnie mniej wiarygodnych aniżeli badanie RCT, metaanaliza lub nawet badanie kohortowe, które mogą stanowić dla niej podstawę. W konsekwencji istotne jest bardzo wyraźne rozgraniczenie wniosków z badania RCT, metaanalizy lub badania kohortowego od wniosków sformułowanych na podstawie analizy podgrup tych badań. Wszelkie próby zacierania różnicy pomiędzy wynikami analizy podgrup a pozostałymi badaniami traktowane są jako świadome oszustwo w badaniu naukowym.

Spektakularnym przykładem oszustwa opartego na analizie podgrup jest raport Scotta Harkonne'a dotyczący wyników fazy III leku Actimmune. Celem raportu było zwiększenie sprzedaży poprzez wskazanie na subpopulację osób, dla których leczenie z wykorzystaniem Actimmune byłoby wyjątkowo korzystne. Ściślej, Harkonnen miał nadzieję, że FDA zatwierdzi lek w niektórych przypadkach śmiertelnej choroby płuc. Ze względu na brak statystycznie istotnych wyników RCT, Harkonnen wykorzystał analizę podgrup. W rezultacie odkryto podgrupę ze statystycznie istotnymi korzyściami w zakresie przeżycia (Mayo, 2020). Sprawa toczyła się dalej w sądzie, gdzie Harkonnen został uznany za winnego oszustwa za wydanie wprowadzającej w błąd informacji prasowej lekarzom, nie podając, że uzyskane wnioski zostały otrzymane w analizie podgrup. Innymi słowy, zaprezentowane one zostały jako wnioski pochodzące bezpośrednio z badania RCT.

Kolejną ważną kwestią jest to, że wyniki analizy podgrup stanowią istotny argument uzasadniający projekt nowego badania. Bardzo często badania eksperymentalnego, a więc ze zwiększonym ryzykiem dla pacjentów biorących w nim udział. Problem dobrze ilustruje przykład badania klinicznego poświęconego terapii digoksyną w przypadkach niewydolności serca. W analizie podgrup w badaniu postawiono hipotezę, że obecność tzw. „trzeciego tonu serca” jest czynnikiem wpływającym na korzyści z leczenia digoksyną w przypadku pacjentów z niewydolnością serca. Uzyskano statystycznie istotne różnice. Wynik otrzymany z analizy

podgrup stanowił podstawę dla zaprojektowania odrębnego badania RCT. Na marginesie należy zauważyć, że wyniki tego badania eksperymentalnego nie potwierdziły wniosków uzyskanych z analizy podgrup (Oxman, Guyatt, 1992).

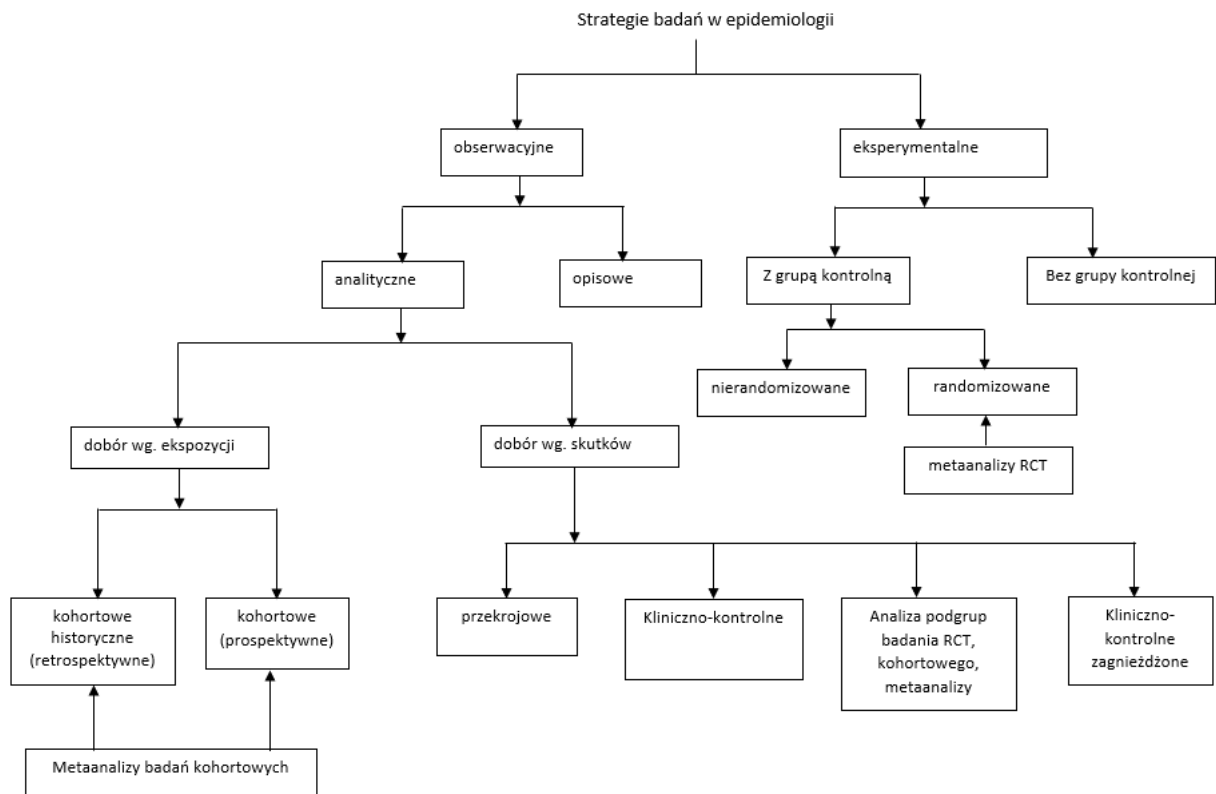
Problem polega na tym, że wiarygodność wniosków z analizy podgrup, które często są podstawą dla projektowania nowych badań eksperymentalnych, w zasadniczy sposób zależy od sposobu uzasadnienia tej analizy. Analiza podgrup dokonywana na podstawie hipotez *post hoc* dostarcza najmniej wiarygodnych danych i jest powszechnie uznawana za tzw. *phaunting* lub *data dredging process* czyli uporczywe poszukiwanie hipotezy, która spełni metodologiczny warunek istotności statystycznej.

6. Strategie badań epidemiologicznych - podsumowanie

Dokonane w poprzednich paragrafach ustalenia można zaprezentować z wykorzystaniem następującego schematu.

Schemat 9

Strategie badań w epidemiologii



Dodatkowo należy pamiętać oczywiście o następujących podziałach, które w powyższym ogólnym schemacie nie zostały ujęte:

- badania wyjaśniające i pragmatyczne,
- badania typu *superiority*, równoważności i *non-superiority*,
- badania klastrowe,
- badania adaptacyjne.

Uwzględniając zaprezentowaną strukturę badań epidemiologicznych można stwierdzić, że z perspektywy oceny badań naukowych dokonywanej przez komisje bioetyczne, najistotniejsze są dwie następujące kwestie:

(1) Eksperymentalny charakter badania.

Szczególnej uwadze należy poddać badania, w których ingerencje mają zostać wprowadzone intencjonalnie przez badacza, zarówno badania typu RCT, jak i badania bez grupy kontrolnej. Nie ulega wątpliwości, że eksperymentalny charakter badania jest jedną z kluczowych cech – najczęściej identyfikowanych w różnego typu wytycznych – stanowiących przedmiot oceny komisji, która musi rozstrzygnąć, czy domniemane korzyści z podjętej ingerencji eksperymentalnej rekompensują podejmowane ryzyko.

(2) Niestandardowe metody monitorowania choroby.

Uwadze należy poddać wszystkie badania, **w tym również badania obserwacyjne**, w których planuje się użycie metod monitorowania przebiegu i efektów choroby nie stosowanych standardowo w diagnostyce medycznej danego schorzenia. W szczególności takie metody diagnostyczne, które tworzą ryzyko zagrożenia życia, zdrowia lub zmniejszają komfort życia pacjenta.

Poza powyższymi, zasadne jest również zwrócenie uwagi na dwie dalsze kwestie, które w niektórych przypadkach mogą stać się przedmiotem kontrowersji natury etycznej i w efekcie podlegać ocenie komisji:

(3) Uzasadnienie projektu badań eksperymentalnych.

(4) Badania, których przedmiotem nie jest ocena skuteczności i bezpieczeństwa terapii, lecz ocena trafności nowych metod diagnostycznych.

Oba wymienione zagadnienia zostaną przybliżone w dalszej części opracowania.

7. Ocena uzasadnienia dla projektu badań eksperymentalnych

Nie ulega wątpliwości, że wszystkie rodzaje badań naukowych powinny posiadać odpowiednie uzasadnienie. Nie chodzi przy tym wyłącznie o identyfikację potencjalnych oczekiwanych korzyści praktycznych wynikających z wdrożenia nowej metody terapii do praktyki medycznej, lecz również o uzasadnienie merytoryczne. Uzasadnienie merytoryczne badania naukowego powinno identyfikować przesłanki świadczące o tym, że badanie oddziaływania terapeutycznego (farmakologicznego, zabiegowego, prewencyjnego) dostarczy w tym badaniu preferowanego rezultatu. Ściślej, że na podstawie posiadanej do tej pory wiedzy można zasadnie wnosić, że prawdopodobieństwo uzyskania w projektowanym badaniu eksperymentalnym preferowanego rezultatu jest większe niż prawdopodobieństwo uzyskania rezultatu przeciwnego. Merytorycznej oceny uzasadnienia dla wykonania danego badania dokonują eksperci w zakresie nauk podstawowych. Dodatkowo w ocenie takiej mogą być wykorzystywane narzędzia analizy statystycznej, których zaprezentowanie w niniejszym opracowaniu byłoby zbędne. Celem jest natomiast wskazanie na specyfikę kompetencji tych członków komisji, którzy mają dokonywać oceny projektów badań nie dysponując wiedzą z obszaru nauk podstawowych i biostatystyki.

Uzasadnienie badań eksperymentalnych można rozpatrywać w dwóch różnych kontekstach. Pierwszy dotyczy procesu rozwoju środka farmakologicznego w kolejnych fazach badań klinicznych prowadzących do rejestracji tego produktu. Ten aspekt oceny uzasadnienia dla podejmowanych ingerencji stanowi osobną kwestię, która nie jest przedmiotem rozważań w niniejszym opracowaniu. Drugi dotyczy projektów badań eksperymentalnych produktów leczniczych, zabiegów terapeutycznych lub prewencyjnych, które są już wykorzystywane w praktyce medycznej. Celem takich badań jest uzyskanie dodatkowej wiedzy o skuteczności podejmowanych działań w wybranych populacjach lub porównania ich skuteczności z innymi metodami terapii. Przedmiotem dalszych ustaleń niniejszego opracowania jest właśnie drugi z tych kontekstów, pozbawiony specyfiki rozstrzygnięć w obszarze badań klinicznych.

Przyjmijmy, że projektowane jest badanie eksperymentalne, w którym możemy uzyskać rezultaty S^+ lub S^- , z których drugi jest rezultatem preferowanym, czyli świadczy o korzyściach zdrowotnych uzyskanych przez pacjenta. Mając na uwadze ryzyko, które wiąże się z wykonaniem każdego badania eksperymentalnego oczekujemy, że można wskazać na wiarygodne dane uzasadniające, że uzyskanie preferowanego celu badania jest bardziej

prawdopodobne aniżeli uzyskanie rezultatu niepożądanego. Problem polega jednak na tym, że dane te mogą być uzyskane z badań o różnym stopniu wiarygodności. W konsekwencji możemy mieć do czynienia z dwoma sytuacjami, szczególnie kontrowersyjnymi z perspektywy ocen dokonywanych przez komisje bioetyczne.

W pierwszej, uzasadnienie dla przeprowadzonego badania eksperymentalnego jest uzyskane z badań nieeksperymentalnych o bardzo niskim poziomie wiarygodności. W konsekwencji, kontrowersyjne może być podejmowanie ryzyka badania eksperymentalnego. Możemy bowiem żywić uzasadnioną wątpliwość, że z uwagi na niską wiarygodność badania prawdopodobieństwo uzyskania rezultatu preferowanego nie jest większe niż prawdopodobieństwo rezultatu niepożądanego. W drugiej sytuacji, wiarygodność uzasadnienia dla przeprowadzenia badania eksperymentalnego jest tak wysoka, że pojawić może się wątpliwość co do zasadności wykonania badania eksperymentalnego – szczególnie z grupą kontrolną – z uwagi na wysoki poziom pewności, że wynik badania eksperymentalnego nic nie zmieni w posiadanej wiedzy. W drugiej z tych sytuacji uzasadnione byłoby projektowanie raczej badań typu *non-inferiority* lub równoważności. Mając na uwadze obie powyższe sytuacje można zastanowić się, które sposoby uzasadniania badań eksperymentalnych winny wzbudzić szczególną uwagę komisji bioetycznych.

Rezultaty badań kohortowych, zarówno prospektywnych jak i retrospektywnych, nie budzą wątpliwości jako podstawa dla projektowania badań eksperymentalnych. Z jednej strony badania kohortowe obu typów mają niższą wiarygodność niż eksperymentalne, zatem zasadne jest potwierdzenie wyników, które są w nich uzyskane, z wykorzystaniem badań eksperymentalnych. Z drugiej strony, wiarygodność badań kohortowych jest na tyle wysoka, że nie ma obawy, iż stanowiłyby punkt wyjścia dla nieuzasadnionego projektu badania eksperymentalnego, o ile oczywiście spełniają warunki poprawności metodologicznej.

Dobrym uzasadnieniem dla projektowania badań eksperymentalnych są również metaanalizy badań kohortowych (obu typów) oraz badań eksperymentalnych. Druga z tych sytuacji może być pewnym zaskoczeniem, ponieważ przedmiotem metaanaliz są w tym przypadku badania eksperymentalne. Dlaczego wyniki metaanaliz badań eksperymentalnych miałyby stanowić uzasadnienie dla kolejnych badań eksperymentalnych? W szczególności, czy nie mielibyśmy w tym przypadku do czynienia z wątpliwością dotyczącą zasadności badania

eksperymentalnego z uwagi na domniemaną oczywistość uzyskanych wyników? Wbrew pozorom sprawa nie jest taka oczywista.

Przedmiotem metaanalizy często są badania eksperymentalne przebiegające według odmiennych protokołów, znacząco różniące się kryteriami włączenia i wykluczenia pacjentów. W badaniach zakwalifikowanych do metaanalizy mogą również występować rozbieżności w przyjętych punktach końcowych oraz identyfikowanych ADR-ach (niepożądanych działaniach środka leczniczego). Zaprojektowanie badania eksperymentalnego ma wówczas sens w kilku sytuacjach. Przykładowo, gdy subpopulacja pacjentów tego badania będzie różniła się od populacji pacjentów z wcześniejszych badań eksperymentalnych pod względem kryteriów włączenia i wykluczenia lub po prostu będzie znacznie liczniejsza niż w którymkolwiek z wcześniejszych badań.

Powody podobne jak w przypadku metaanaliz mogą sprawiać, że projekty nowych badań eksperymentalnych są uzasadnione na podstawie wcześniejszych badań eksperymentalnych z mniej liczną populacją lub odmiennie scharakteryzowanymi celami badania.

Odmienne sytuacja przedstawia się w przypadku uzasadnienia projektu badania eksperymentalnego, które byłoby uzyskane na podstawie badania kliniczno-kontrolnego lub analizy podgrup. W przypadku obu tych badań celem jest ustalenie korelacji pomiędzy zidentyfikowaną w przeszłości ekspozycją / narażeniem, a obecnym stanem zdrowia lub wynikiem podjętej ingerencji (np. w badaniu RCT). Obie procedury badawcze są narażone na specyficzne błędy, na które zwrócono uwagę wcześniej w niniejszym opracowaniu. W konsekwencji, dane uzyskiwane z tych badań stanowią mało wiarygodne uzasadnienie dla podjęcia badania eksperymentalnego. Podkreślić dodatkowo należy fakt, że pomimo swojego podobieństwa, wiarygodność obu tych typów procedur może być różna.

Analiza podgrup, jak opisano wcześniej, może być oparta na *hipotezie poprzedzającej* sformułowanej na etapie projektu badania RCT (lub metaanalizy) lub może mieć charakter *post hoc*. W drugim przypadku hipoteza stanowiąca podstawę przeprowadzenia analizy podgrup sformułowana została po ukończeniu badania RCT. Uzasadnienie badania eksperymentalnego na podstawie rezultatów uzyskanych z badania kliniczno-kontrolnego lub na podstawie analizy podgrup opartej na *hipotezie poprzedzającej* posiadają podobną wiarygodność, z pewnością niższą, aniżeli uzasadnienie uzyskane z badań kohortowych czy metaanaliz.

Szczególną uwagę Komisji bioetycznych winny jednak wzbudzić projekty tych badań eksperymentalnych, dla których uzasadnienie jest uzyskane z analizy podgrup przeprowadzonej na podstawie hipotez *post hoc* lub badań opisowych. Są to badania o najniższym poziomie wiarygodności. Konieczna jest zatem bardzo wnikliwa ocena projektu, przede wszystkim stopnia ryzyka, jakie związane jest z projektowanym badaniem. Zasadne jest w szczególności pytanie o to, czy poziom ryzyka podejmowania ingerencji eksperymentalnej jest akceptowalny w świetle niskiej wiarygodności przesłanki uzasadniającej przeprowadzenie badania eksperymentalnego.

8. Badania określające trafność testów diagnostycznych

Poza wymienionymi wcześniej badaniami, których celem jest ocena skutków i bezpieczeństwa podejmowanych oddziaływań terapeutycznych, prewencyjnych, zabiegowych oraz narażeń środowiskowych, uwaga komisji bioetycznych musi być również ukierunkowana na badania, których celem jest ocena nowych metod diagnostycznych. Badania te stanowią, pod względem ilościowym, znaczącą część wszystkich badań naukowych podejmowanych w medycynie.

Testy diagnostyczne wykonywane w praktyce medycznej mają za zadanie potwierdzenie choroby (wyniki dodatnie) lub jej wykluczenie (wyniki ujemne). Problem polega na tym, że żadne testy nie mają charakteru ostatecznie rozstrzygającego. Bardzo często dostarczają one wyników fałszywych (fałszywie dodatnich lub fałszywie ujemnych). Każda metoda diagnostyczna może być oceniona z uwagi na jej zdolność do potwierdzenia choroby w populacji osób chorych (tzw. czułość diagnostyczna) oraz zdolność do wykluczenia choroby w populacji osób zdrowych (tzw. swoistość diagnostyczna). Szczegółowa charakterystyka tych parametrów w niniejszym opracowaniu jest zbędna, wystarczy stwierdzić, że wysoka wartość tych parametrów wiąże się z niższą częstotliwością wyników fałszywych (Rzepiński, 2018). Ocena obu tych parametrów (tzw. trafności testu) dokonuje się z wykorzystaniem testu referencyjnego, czyli najlepszej dostępnej metody stosowanej w diagnostyce danej choroby. Dzięki temu można określić wartości liczbowe czułości i swoistości diagnostycznej nowej metody diagnostycznej.

Mamy zatem rozróżnienie na test referencyjny oraz testy podstawowe oceniane w odniesieniu do wyników testu referencyjnego. Przykładem może być wysiłkowe EKG, które jest testem podstawowym w diagnostyce niektórych chorób układu krążenia oraz

koronarografia pełniąca rolę testu referencyjnego. Dzięki porównaniu dużej liczby pacjentów poddanych obu tym testom można określić trafność, czyli czułość i swoistość diagnostyczną testu wysiłkowego EKG, a więc ustalić, w jakim stopniu jest to test przydatny w rozpoznawaniu lub wykluczeniu danego schorzenia.⁷ Można określić prawdopodobieństwo wyników prawdziwych uzyskanych w tym teście.

Wiele badań naukowych prowadzonych w medycynie ma na celu ustalenie przydatności nowych metod diagnostycznych w rozpoznawaniu choroby. Ocena taka musi być dokonana poprzez porównanie wyników nowej metody diagnostycznej z testem referencyjnym.

Przyjrzyjmy się tej kwestii bliżej.

Lekarz zleca wykonanie badań dodatkowych wówczas, gdy na podstawie wywiadu i badania fizykalnego dostrzega uzasadnienie dla wykonania danego testu. Jego celem jest bądź potwierdzenie, bądź wykluczenie pewnej hipotezy diagnostycznej identyfikującej chorobę. Można zatem powiedzieć, że decyzje lekarzy zlecających wykonanie danego testu wyodrębniają w społeczeństwie pewną populację pacjentów, dla których istnieją wskazania do przeprowadzenia testu. Niemniej nie ulega wątpliwości, że bardzo niewielka część z tych pacjentów kwalifikowana byłaby jako grupa ze wskazaniem do wykonania testu referencyjnego. Testy referencyjne są bowiem zwykle bardziej kosztowne, często bardziej pracochłonne, wymagają zaangażowania specjalistycznego personelu medycznego oraz często zwiększają ryzyko pogorszenia stanu zdrowia pacjenta (np. ryzyko zgonu w przypadku koronarografii). Co w opisanej sytuacji jest szczególnie istotne z perspektywy działań komisji bioetycznych?

Założmy, że badacz pragnie w zaprojektowanym badaniu określić, jaka jest trafność testu T w identyfikacji lub wykluczeniu określonej choroby, poprzez jego porównanie z wynikami testu referencyjnego T_r . Oznacza to, że cała populacja pacjentów, wobec których przeprowadzono diagnostykę za pomocą testu T, musi następnie być skierowana do testu referencyjnego, bez względu na wskazania lekarskie do wykonania takiego testu. Każdy pacjent, który uzyskał wynik z testu T, a nie został poddany testowi referencyjnemu, zwiększa

⁷ Podkreślić w tym miejscu należy, że zdolność danego testu do wykluczenia choroby nie dopełnia się ze zdolnością testu do jej potwierdzenia. Innymi słowy, z wiedzy o zdolności testu do potwierdzenia nie możemy wnosić o zdolności testu do wykluczenia choroby. W praktyce oznacza to, że w sytuacji gdy uzyskano ujemny wynik testu, o którym wiadomo, że jego zdolność do potwierdzenia choroby (czułość diagnostyczna) wynosi 75%, nie można wnosić, że choroba jest wykluczona w 25%.

ryzyko błędu w ocenie trafności testu T. Błąd taki jest określany mianem błędu weryfikacji. Nieprzeprowadzenie testu referencyjnego wobec wszystkich pacjentów z testem T powoduje wypaczenie w ocenie testu T. Widoczne staje się w tym miejscu, jak bardzo istotna jest ocena komisji bioetycznej, która musi ustalić, czy ryzyko związane ze szkodliwymi efektami testu referencyjnego jest rekompensowane przez realizowane cele poznawcze, czyli uzyskanie wiedzy o trafności nowej metody diagnostycznej.

9. Bibliografia

Beaglehole R., Bonita R., Kjellstrom T. (2002), *Podstawy epidemiologii*, tł. W. Hanke pod red. N. Szeszeni-Dąbrowskiej, Oficyna Wydawnicza Instytutu Medycyny Pracy im. prof. J. Nofera, Łódź.

Brodniewicz T. (2015), *Słownik wybranych pojęć*, [w:] T. Brodniewicz (red.), *Badania kliniczne*, Wydawnictwo CeDeWu, Warszawa: 27–32.

Brody T. (2016), *Clinical trials. Study Design, Endpoints and Biomarkers, Drug Safety, and FDA and ICH Guidelines*, Elsevier/Academic Press, Amsterdam.

Chow S.C., Chang M., Pong A. (2005), *Statistical consideration of adaptive methods in clinical development*, „Journal of Biopharmaceutical Statistics” 15 (4): 575–591.

CIOMS (Council for International Organizations of Medical Sciences) (2016), *International Ethical Guidelines for Health-related Research Involving Humans*, Fourth Edition, Geneva.

Cipriani A., Barbui C. (2013), *What is a factorial trial?*, „Epidemiology and Psychiatric Sciences” 22 (3): 213–215.

Collins R., Peto R., Armitage J. (2002), *The MRC/BHF Heart Protection Study: Preliminary results*, „International Journal of Clinical Practice” 56 (1): 53–56.

D’Agostino R.B., Massaro J.M., Sullivan L. M. (2003), *Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics*, „Statistics in Medicine” 22 (2): 169–186.

Evans S., Ting N. (2016), *Fundamental Concepts for New Clinical Trialists*, Taylor & Francis Group, Chapman & Hall/ CRC Biostatistics Series, London, New York.

FDA (Food and Drug Administration) (2018), *Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Guidance for Industry*, URL = <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics> [dostęp 15.05.2022].

FDA (Food and Drug Administration) (2019), *Adaptive design clinical trials for drugs and biologics Guidance for Industry*, URL = <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry> [dostęp 16.05.2022].

Ford I., Norrie J. (2016), *Pragmatic trials*, „New England Journal of Medicine” (375): 454–463.

Godwin M., Ruhland L., Casson I., MacDonald S., Delva D., Birtwhistle R., Lam M., Seguin R. (2003), *Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity*, „BMC Medical Research Methodology” (3): 28.

Gottlieb M., Schroff R., Schanker H.M., Weisman J.D., Fan P.T., Wolf R.A., Saxon A. (1981), *Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency*, „The New England Journal of Medicine” 305 (24): 1425–1431.

Jabłonowski L. (1996), *Klasyfikacja badań epidemiologicznych*, [w:] L. Jabłoński (red.), *Epidemiologia. Podręcznik dla lekarzy i studentów*, Wydawnictwo Folium, Lublin: 67–74.

Jędrzychowski W. (1999), *Epidemiologia. Wprowadzenie i metody badań*, Wydawnictwo Lekarskie PZWL, Warszawa.

Jędrzychowski W. (2002), *Podstawy epidemiologii*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

Jędrzychowski W. (2010), *Epidemiologia w medycynie klinicznej i zdrowiu publicznym*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

Jędrzejowski A. (2015), *Badania kliniczne w populacji pediatrycznej*, [w:] T. Brodniewicz (red.), *Badania kliniczne*, Wydawnictwo CeDeWu, Warszawa: 199–226.

- Karwowski M. (2018), *Metaanaliza: o ilościowym syntetyzowaniu ustaleń empirycznych*, „Polskie Forum Psychologiczne” 23 (2): 252–276.
- Kaufman G. (2016), *Adverse drug reactions: classification, susceptibility and reporting*, „Nursing Standard” 30 (50): 53–63.
- Lesaffre E. (2008), *Superiority, equivalence, and non-inferiority trials*, „Bulletin of the N.Y.U. Hospital for Joint Diseases” 66 (2): 150–154.
- Mahajan R., Gupta K. (2010), *Adaptative design clinical trials: Methodology, challenges and prospect*, „Indian Journal of Pharmacology” 42 (4): 201–207.
- Marko M., Pawliczak R. (2019), *Badania typu noninferiority w praktyce klinicznej*, „Alergologia Polska” 6 (4): 152–159.
- Masełbas W. (2008), *Działania niepożądane leków: definicje, przyczyny, podział*, [w:] I. Łagocka, A. Maciejczyk (red.), *Nadzór nad bezpieczeństwem farmakoterapii. Pharmacovigilance*, OINPHARMA, Warszawa: 67–86.
- Mayo D. (2020), *P-Values on Trial: Selective Reporting of (Best Practice Guides Against) Selective Reporting*, „Harvard Data Science Review” 2 (1), URL = <https://doi.org/10.1162/99608f92.e2473f6a> [dostęp 10.12.2023].
- McAlister F.A., Straus S.E., Sackett D.L., Altman D.G. (2003), *Analysis and reporting of factorial trials: a systematic review*, „Journal of the American Medical Association” 289 (19): 2545–2553.
- Montgomery A.M., Peters T.J., Little P. (2003), *Design, analysis and presentation of factorial randomised controlled trials*, „BMC Medical Research Methodology” 3: 26, URL = <https://doi.org/10.1186/1471-2288-3-26> [dostęp 18.05.2022].
- Montgomery A.A., Astin M.P., Peters T.J. (2011), *Reporting of factorial trials of complex interventions in community settings: a systematic review*, „Trials” 12: 179.
- Oxman A., Guyatt G., (1992), *A Consumer Guide to Subgroup Analysis*, „Annals of Internal Medicine” 116 (1): 78–84.
- Rothwell P.M. (2005), *External validity of randomised controlled trials: to whom do the results of this trial apply?*, „Lancet” 365 (9453): 82–93.

- Rzepiński T. (2018), *Twierdzenie Bayesa w projektowaniu strategii diagnostycznych w medycynie*, „Diametros” 57: 39–60.
- Rzepiński T. (2019), *O przedmiocie badań chemii i problemach określania biologicznej aktywności środków farmakologicznych – studium historii talidomidu*, [w:] J. Boruszewski, R. Kazibut, S. Leciejewski, T. Rzepiński, Z. Tworak (red.), *Między filozofią a chemią. Księga Jubileuszowa dla Profesora Zeidlera*, Wydawnictwo UAM, Poznań: 153–170.
- Schwartz D., Lellouch J. (1967), *Explanatory and pragmatic attitudes in therapeutical trials*, „Journal of Chronic Diseases” 20 (8): 637–648.
- Sugarman J., Califf R.M. (2014), *Ethics and Regulatory Complexities for Pragmatic Clinical Trials*, „Journal of American Medical Association” 311 (23): 2381–2382.
- Walker E., Nowacki A.S. (2010), *Understanding Equivalence and Noninferiority Testing*, „Journal of General Internal Medicine” 26 (2): 192–196.
- Weintraub W.S., Lüscher T.F., Pocock S. (2015), *The perils of surrogate endpoints*, „European Heart Journal” 36 (33): 2212–2218.
- Wiela-Hojeńska A., Orzechowska-Juzwenko K. (2006), *Niepożądane działania leków*, [w:] K. Orzechowska-Juzwenko, (red.), *Farmakologia kliniczna. Znaczenie w praktyce medycznej*, Górnicki Wydawnictwo Medyczne, Wrocław: 209–246.