

Warszawski Uniwersytet Medyczny
Zakład Metodologii Badań Naukowych

Wstęp do medycyny opartej na dowodach naukowych (*evidence-based medicine*)

Skrypt dla studentów kierunku lekarskiego

Autor

Albert Stachura,

MSc in Evidence-Based Healthcare



WARSZAWSKI
UNIwersYTET
MEDYCZNY

Warszawski Uniwersytet Medyczny
Zakład Metodologii Badań Naukowych

Wstęp do medycyny opartej na dowodach naukowych (*evidence-based medicine*)

Skrypt dla studentów kierunku lekarskiego

Autor

Albert Stachura,

MSc in Evidence-Based Healthcare



WARSZAWSKI
UNIwersYTET
MEDYCZNY

Warszawa 2026

Autor dr n. med. i n. o zdr. Albert Stachura, MSc in Evidence-Based Healthcare

Recenzenci Prof. dr hab. n. med. Hanna Szajewska
 Prof. dr hab. n. med. i n. o zdr. Mariusz Panczyk

Wydano za zgodą Uczelnianej Rady ds. Kształcenia WUM

ISBN 978-83-7637-671-4

Wydano w Sekcji Druków Uczelnianych
Warszawskiego Uniwersytetu Medycznego
tel. (22) 5720 327,
e-mail: wydawnictwo@wum.edu.pl
www.drukiuczelniane.wum.edu.pl

Spis treści:

<i>Mamy problem</i>	5
<i>...ale i rozwiązanie, czyli czym jest EBM?</i>	8
Pułapka mechanistycznego rozumowania – historia flekainidu.....	12
<i>5 kroków EBM – przypadek kliniczny</i>	16
<i>Precyzowanie pytania klinicznego – PICO/PECO</i>	20
Leczenie	20
Prognoza	21
<i>Rodzaje odpowiedzi</i>	24
Opisy przypadków	24
Badania obserwacyjne	27
Zakłócanie (<i>confounding</i>) jako ograniczenie badań obserwacyjnych.....	29
Randomizowane badania kliniczne	31
Przeglądy systematyczne i meta-analizy	34
Hierarchia medycznych dowodów naukowych	36
<i>Jak czytać publikacje naukowe?</i>	41
Jak zacząć czytać artykuły, czyli sztuka wyrzucania nauki do kosza	41
Przeglądy systematyczne i meta-analizy: filtry w świecie informacyjnego szumu ...	43
Badania jakościowe: Jak zrozumieć to, czego nie da się policzyć?	45
Badania randomizowane: Jak oddzielić naukę od marketingu?	48
Dlaczego nie możemy ufać wszystkiemu, co opublikowano?.....	51
<i>Gdzie znajdę dobrą odpowiedź na moje pytania?</i>	55
Korzystanie z UpToDate	56
Skuteczność leczenia naszej pacjentki – czego się dowiadujemy?.....	58
Działania niepożądane terapii – jak znaleźć dowody?	62
Szukanie w UpToDate.....	63
Szukanie w PubMedzie.....	64
Jak znaleźć informacje nt. prognozy?.....	70
Dlaczego AI zawodzi, gdy szukamy wysokiej jakości danych?.....	74

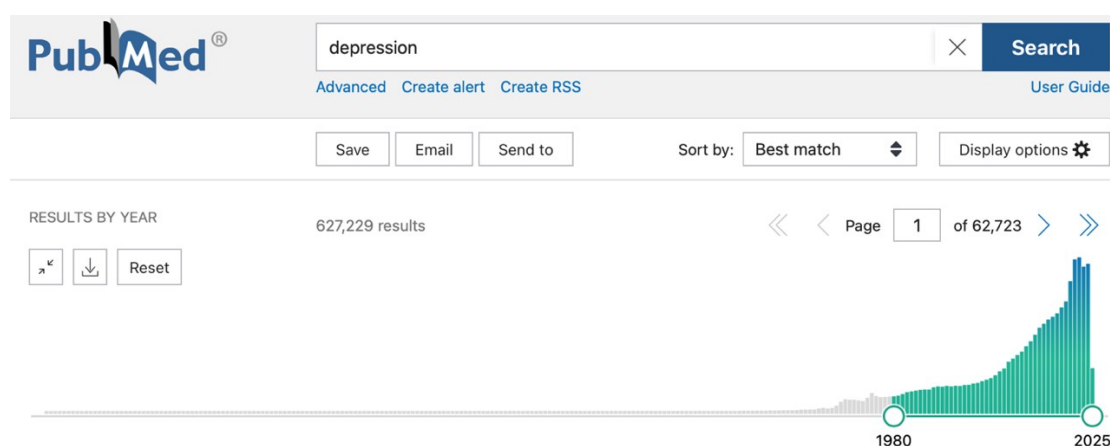
<i>Jak rozgryźć dane statystyczne?</i>	77
W jaki sposób można przedstawić dane z badania klinicznego?	77
Liczenie ryzyka, szans, relatywnego ryzyka (RR) i relatywnych szans (OR).....	78
Interpretacja RR, OR i HR	81
Czy uzyskane wyniki są przypadkowe, czyli dwa słowa o 95% przedziale ufności...83	
Odpowiedzi na pytania dot. naszej pacjentki	86
Statystyka dla klinicystów: Jak nie dać się zwieść liczbom?	87
Istotność statystyczna kontra kliniczna: Czy każda „gwiazdka” w badaniu ma znaczenie dla Twojego pacjenta?.....	92
<i>Prawidłowe i błędne użycie dowodów naukowych</i>	95
Kiedy dowody zmieniły świat na lepsze	95
Historia tytoniowej epidemii: Jak Doll i Hill udowodnili, że palenie zabija	95
Sterydy w COVID-19 – od teorii do twardych dowodów	98
Kiedy EBM jest źle używany	100
Polityka kontra nauka: Jak Donald Trump ogłosił, że paracetamol powoduje autyzmem.....	100
Skandal Vioxx: Lekcja o tym, czego nie widać na pierwszy rzut oka.....	103
<i>Więcej o EBM</i>	106
Ocena jakości i raportowanie badań.....	106
Nowe kierunki rozwoju.....	107
<i>Podziękowania</i>	110
<i>Piśmiennictwo</i>	111
<i>Spis rycin i tabel</i>	118

Mamy problem...

„85% opublikowanych artykułów medycznych nie zmieni Waszej praktyki klinicznej”. Byłem w szoku, kiedy pierwszy raz to usłyszałem podczas zajęć w Oxfordzkim Centrum Medycyny Opartej na Dowodach Naukowych. Będąc uprzednio studentem kierunku lekarskiego na naszej uczelni sądziłem, że badania naukowe być może nie są idealne, ale duża ich część może pomóc lekarzom i pacjentom. Odnosimy się do osiągnięć nauki, argumentując nasze decyzje kliniczne. Przekonujemy pacjentów o skuteczności leczenia w oparciu o badania naukowe. Różnimy się od szarlatanów oraz praktykantów medycyny alternatywnej odwołując się do rzetelnych dowodów.

Jednak już w 2009 roku Iain Chalmers i Paul Glasziou oszacowali, że ponad 85% nakładów na badania biomedyczne może być traconych (wasted) na różnych etapach procesu badawczego - od wyboru problemu badawczego po publikację wyników (zapewne dotyczy to też badań przedklinicznych). Wskazywali na to, że konkluzje takich badań mogą wprowadzać w błąd lub być publicznie niedostępne.¹ Głównymi przyczynami złego inwestowania zasobów są: (1) nieadresowanie pytań ważnych dla pacjentów i klinicystów, (2) używanie złych metod badawczych, (3) publikowanie wyników badania jedynie w postaci streszczenia, (4) błędne lub niepełne raportowanie wyników badań. Autorzy sami przyznali, że są zszokowani wynikami swoich analiz! W podobnym duchu jeszcze w latach 90. Doug Altman – główny doradca

statystyczny the BMJ (British Medical Journal) – pisał o niskiej jakości badań medycznych.² Uznawał za skandal, że „ogromne sumy pieniędzy są wydawane corocznie na badania, które są poważnie wadliwe poprzez zastosowanie niewłaściwych rodzajów badań, małych, niereprezentatywnych grup pacjentów, nieprawidłowych metod analizy i błędnej interpretacji”. Altman przekonywał, że potrzebujemy „mniej badań, lepszych badań i badań przeprowadzanych z właściwych powodów”. Pomimo tych apelów liczba publikowanych corocznie doniesień naukowych stale rośnie – rozważmy poniższy przykład (Rycina 1).



Rycina 1. Liczba opublikowanych badań (627 229) na temat depresji w bazie danych PubMed w latach 1980-2025 (stan na 26.03.2025 r.)

W 2024 roku opublikowano wyniki ponad 41 250 badań na temat depresji. Zakładając, że chcielibyśmy być na bieżąco ze „wskazaniami aktualnej wiedzy medycznej”, do czego zobowiązuje nas ustawa o zawodach lekarza i lekarza dentysty (art. 4),³ musielibyśmy czytać około 113 badań dziennie! Nawet zakładając, że jedynie 15% badań jest „dość dobrych”, by kierować się nimi przy podejmowaniu decyzji klinicznych, byłoby to nadal 17 badań dziennie – a mówię tu o jednej chorobie w wąskiej dziedzinie psychiatrii! Poza przytłaczającą liczbą doniesień publikowanych codziennie, wyzwaniem pozostaje wybranie tych 15%.

Powstaje zatem zasadniczy problem. Jak odnaleźć się w szumie informacyjnym i skutecznie identyfikować dobre i istotne klinicznie dane? Jak odróżnić słabe badania, wprowadzające w błąd od tych, które powinny nas informować w praktyce klinicznej? W dzisiejszej dobie poza lekarzami wiele grup walczy o uwagę pacjentów – influencerzy, „specjaliści” medycyny alternatywnej, oszuści, żerujący na cierpieniu chorych ludzi. Sami pacjenci decydują się nieraz na pozyskiwanie informacji z Internetu czy korzystają ze wsparcia sztucznej inteligencji. Nigdy nie mieliśmy dostępu do tak szerokiego morza informacji, jednocześnie przeżywając tak dużą niepewność, dotyczącą tego, co jest użyteczne i prawdziwe, a co stanowi szum i powinno zostać odrzucone.

W obliczu ciężaru tego wyzwania można poczuć zniechęcenie. Trudno się dziwić – zdaje się, że identyfikacja rzetelnych dowodów przypomina poszukiwanie igły w stogu siana. Jednak w tym skrypcie proponuję rozwiązanie – ustrukturyzowane podejście do praktykowania medycyny w oparciu o aktualne i adekwatne dowody naukowe. Przedstawię Ci, drogi Czytelniku, jak w praktyce szybko i skutecznie znajdować odpowiedzi na pytania, które mogą pojawić się w Twojej przyszłej (lub obecnej) pracy.

...ale i rozwiązanie, czyli czym jest EBM?

Medycyna oparta na dowodach naukowych, czyli *evidence-based medicine* (EBM) polega na integracji (1) najlepszych adekwatnych dowodów naukowych z (2) doświadczeniem klinicznym oraz (3) wartościami i preferencjami pacjenta (Rycina 2).



Rycina 2. Elementy EBM

Celem tego podejścia jest poprawa wyników leczenia, tj. zmniejszenie śmiertelności pacjentów, wyleczenie z choroby, złagodzenie dolegliwości, redukcja szkód leczenia czy poprawa jakości życia. Stosując to podejście, możemy też lepiej informować pacjenta o prognozie oraz odpowiednio interpretować wyniki badań laboratoryjnych. Z własnego doświadczenia oraz relacji wielu moich znajomych lekarzy dodam, że jest jeszcze jedna korzyść praktykowania w ten sposób – satysfakcja lekarza. Uważam, że nie należy umniejszać wartości, jaką jest przyjemność z dobrze wykonanej pracy, a EBM jest narzędziem, które pomaga osiągnąć możliwie jak najlepszy efekt leczenia.

EBM ma swoje korzenie jeszcze w XVIII wieku, kiedy to James Lind przeprowadził (być może) pierwsze na świecie badanie kliniczne z grupą kontrolną, w którym wykazał, że cytrusy są pomocne w leczeniu szkorbutu. Jak sam pisał w swoim raporcie „Zanim temat ten [leczenie szkorbutu] mógł zostać przedstawiony w jasnym i właściwym świetle, konieczne było usunięcie dużej ilości śmieci”.⁴ Innymi słowy, zanim Lind wpadł na to, żeby dać chorym marynarzom cytrusy, musiał przejrzeć w systematyczny sposób dostępną wówczas literaturę i odsiać ziarna od plew, by zidentyfikować potencjalnie skuteczną interwencję leczniczą. Można powiedzieć, że mimo upływu czasu wyzwanie odróżnienia dobrych dowodów od złych pozostaje aktualne. Formalnie określenie *evidence-based medicine* zostało przedstawione w latach 90. XX wieku przez kanadyjski zespół epidemiologów i lekarzy z Gordonem Guyattem na czele.⁵ EBM miał być wówczas nowym sposobem nauczania medycyny i kształcenia kadry lekarskiej. Autorzy tego terminu pisali, że podejście to „odrzuca intuicję, niesystematyczne doświadczenie kliniczne i uzasadnienie patofizjologiczne jako wystarczające podstawy

do podejmowania decyzji klinicznych i kładzie nacisk na analizę dowodów z badań klinicznych”. Co to jednak oznacza w praktyce?

Rozważmy przykład przedstawiony w pierwszym manifeście o EBM:⁵ wyobraź sobie, że pracujesz jako młody rezydent, który przyjmuje 43-letniego pacjenta po napadzie drgawek uogólnionych. Pacjent nigdy wcześniej nie miał epizodu drgawek. Historia choroby, badania z krwi ani badania obrazowe nie wskazują na ewidentną przyczynę zaistniałego zdarzenia. Dajesz pacjentowi dożylnie fenytoinę (lek przeciwpadaczkowy) i utrzymujesz terapię w formie tabletek. Pacjent jest bardzo przejęty i chciałby wiedzieć jakie jest ryzyko kolejnego takiego epizodu. Co robisz?

Wedle historycznego modelu praktykowania medycyny spytałbyś starszego kolegę, który (wsparty w swej opinii przez specjalistę) powiedziałby Ci, że ryzyko nawrotu jest duże (bez konkretnej liczby) i że należy to przekazać pacjentowi. Stosujesz się do zaleceń, podkreślasz w rozmowie z pacjentem, że nie może prowadzić, powinien brać leki, udawać się na wizyty kontrolne do lekarza rodzinnego. Pacjent opuszcza szpital z mglistym pojęciem na temat swojej prognozy.

Istnieje jednak alternatywa – możesz zadać sobie pytanie czy wiesz coś na temat prognozy w kontekście przypadku pacjenta, a jeśli nie – poszukać informacji w dostępnych bazach danych. Mając wiedzę na temat rodzajów badań oraz interpretacji podstawowych danych statystycznych znajdujesz artykuł, który odpowiada na Twoje pytanie. Dowiadujesz się z niego, że ryzyko nawrotu w ciągu 1. roku wynosi między 43% a 51%, a w ciągu 3 lat między 51% a 60%. Po 18-miesięcznym okresie wolnym od napadów ryzyko nawrotu spada poniżej 20%. Przedstawiasz te informacje pacjentowi razem z zaleceniem przyjmowania leków, udawania się na wizyty kontrolne do lekarza rodzinnego i rekomendujesz ocenę zasadności leczenia w przypadku

braku nawrotów przez 18 miesięcy. Pacjent wychodzi ze szpitala z jasnym przekonaniem na temat możliwej prognozy.

Stare podejście opiera się fundamentalnie na doświadczeniu klinicznym, przestrzeganiu dotychczas wypracowanej praktyki i autorytecie starszych lekarzy. Dostępnych tu mamy kilka opcji rozwoju: refleksja nad własnym doświadczeniem, refleksja nad biologią danego zagadnienia, zajrzenie do podręcznika lub spytanie lokalnego eksperta. Podejście to ma następujące założenia:

1. Niesystematyczne obserwacje z doświadczenia klinicznego są ważnym sposobem budowania wiedzy na temat rokowania pacjenta, wartości testów diagnostycznych, skuteczności leczenia.
2. Badanie i zrozumienie podstawowych mechanizmów choroby i zasad patofizjologicznych są wystarczającymi przewodnikami dla praktyki klinicznej.
3. Połączenie gruntownego tradycyjnego szkolenia medycznego i zdrowego rozsądku jest wystarczające, aby umożliwić ocenę nowych testów i metod leczenia.
4. Wiedza merytoryczna i doświadczenie kliniczne stanowią wystarczającą podstawę do generowania ważnych wytycznych dla praktyki klinicznej.

Założenia **nowego podejścia** są następujące:

1. Doświadczenie kliniczne i rozwój intuicji klinicznej (szczególnie w odniesieniu do diagnozy) są kluczową i niezbędną częścią stania się kompetentnym lekarzem. Wiele aspektów praktyki klinicznej nie może lub nigdy nie zostanie odpowiednio przetestowanych. Doświadczenie kliniczne i płynące z niego wnioski są szczególnie ważne w takich sytuacjach. Jednocześnie systematyczne obserwacje przeprowadzone w sposób powtarzalny i bezstronny

znacznie zwiększają zaufanie do wiedzy na temat rokowania pacjenta, wartości testów diagnostycznych i skuteczności leczenia. W przypadku braku systematycznej obserwacji należy zachować ostrożność w interpretacji informacji pochodzących z doświadczenia klinicznego i intuicji, ponieważ czasami mogą one wprowadzać w błąd.

2. Zgłębienie podstawowych mechanizmów chorób stanowi kluczowy, choć sam w sobie niewystarczający fundament dla praktyki klinicznej. Wnioski diagnostyczne i terapeutyczne oparte wyłącznie na zasadach patofizjologii mogą okazać się błędne.
3. Zrozumienie pewnych zasad dotyczących określenia siły dowodów jest niezbędne do prawidłowej interpretacji literatury na temat przyczyn choroby, prognozy, testów diagnostycznych i strategii leczenia.

EBM nie polega na odrzuceniu doświadczenia klinicznego. Nie jest też ślepym podążaniem za badaniami. Jest integrowaniem dostępnych możliwie jak najlepszych dowodów z intuicją i doświadczeniem klinicznym przy świadomości ograniczeń obu tych elementów.

Pułapka mechanistycznego rozumowania – historia flekainidu

Model patofizjologiczny wielu chorób można wytłumaczyć następująco: przyczyna A prowadzi do mechanizmu B, co skutkuje chorobą C. Intuicja podpowiada, że jeśli znajdziemy lek, który zablokuje mechanizm B, to wyleczymy pacjenta. To klasyczne rozumowanie oparte o mechanizmy (*mechanism-based reasoning*). Brzmi to logicznie, prawda? Niestety, w historii medycyny to właśnie ta logika wielokrotnie zawiodła, a idealnym przykładem jest historia leku o nazwie flekainid.⁶

W latach 70. i 80. XX wieku kardiolodzy mierzyli się z poważnym problemem: pacjenci po zawałach serca często mieli tzw. dodatkowe

skurcze komorowe (komorowe zaburzenia rytmu – PVCs). Zauważono korelację – osoby z dużą liczbą takich skurczów umierały nagle znacznie częściej. Mechanistyczny wniosek wydawał się oczywisty: PVCs są przyczyną nagłej śmierci, więc jeśli podamy lek, który je usunie, uratujemy życie pacjentów. W 1981 roku na łamach *The New England Journal of Medicine* opublikowano wyniki badania, które zdawało się potwierdzać ten sukces.⁷ Badacze przetestowali flekainid na małej grupie pacjentów i uzyskali następujące wyniki:

- U 9 z 10 pacjentów udało się całkowicie wyeliminować dodatkowe skurcze.
- Średnia redukcja skurczów komorowych wyniosła aż 98,3%.
- Lek był „dobrze tolerowany” i wykazywał „wysoką skuteczność”.

Z punktu widzenia patofizjologii lekarze odnieśli sukces. Zbadali lek, który niemal całkowicie eliminował widoczny w EKG problem. Flekainid stał się standardem leczenia.

W czym tkwi haczyk? W badaniu z 1981 roku mierzono jedynie tzw. punkt zastępczy (*surrogate end point*) – czyli parametr laboratoryjny (liczbę skurczów w EKG), o którym myśleliśmy, że przewiduje przeżycie. To założenie wcale jednak nie musi przekładać się na realne zmniejszenie śmiertelności. Jak dowiesz się w kolejnych rozdziałach, punkty zastępcze bywają zwodnicze. To, że EKG wygląda lepiej, nie zawsze oznacza, że pacjent czuje się lepiej lub żyje dłużej.

Pod koniec lat 80. naukowcy postanowili wreszcie sprawdzić to, co najważniejsze: czy pacjenci leczeni flekainidem faktycznie żyją dłużej. Tak powstało badanie CAST (*Cardiac Arrhythmia Suppression Trial*).⁸ CAST było rygorystycznym, wielośrodkowym randomizowanym badaniem klinicznym kontrolowanym placebo. W kolejnych rozdziałach dowiesz się więcej o tym, na czym polega takie badanie. Zamiast zakładać, że lek działa, badacze losowo przydzielili pacjentów do dwóch grup: jednej

podawano lek (flekainid lub enkainid), a drugiej placebo (tabletkę bez substancji czynnej).

Wyniki były niepokojące:

- Całkowita śmiertelność w grupie otrzymującej leki wyniosła 7,7%, podczas gdy w grupie placebo tylko 3,0%. Oznacza to, że osoby przyjmujące lek, który miał je ratować, umierały ponad dwa i pół raza częściej niż te, które brały placebo.
- Badanie trzeba było przerwać przedwcześnie ze względów etycznych – dalsze podawanie leku byłoby narażaniem pacjentów na śmierć.

Okazało się, że choć flekainid faktycznie poprawiał EKG, to jednocześnie miał ukryte działanie arytmogenne – u niektórych pacjentów po zawale prowokował nowe, śmiertelne zaburzenia rytmu, których nie widać było w krótkich testach.

Przypadek flekainidu to podręcznikowy dowód na to, dlaczego rozumowanie oparte o mechanizmy może zawodzić. Mechanizmy są niezbędne do wymyślania nowych terapii, ale są niewystarczające, by uznać leczenie za skuteczne. Dlaczego? Bo ludzki organizm jest systemem niezwykle złożonym. Nasza wiedza o patofizjologii jest przeważnie niepełna. Lek może naprawiać jedną rzecz (skurcze PVCs), jednocześnie psując inną, znacznie ważniejszą (np. stabilność elektryczną serca), o której istnieniu w danym kontekście nie mieliśmy pojęcia.

Ta historia nauczyła nas, że samo ufanie autorytetom czy przekonującym teoriom nie jest wystarczające, żeby potwierdzić skuteczność lub szkody danego leczenia. Potrzebna jest integracja najlepszych możliwych dowodów z doświadczeniem klinicznym oraz wartościami i preferencjami pacjenta. W tym wypadku kluczowe było uzupełnienie rozumienia działania leku o wyniki wysokiej jakości badania

klinicznego. W jaki sposób zatem mamy stosować EBM w codziennej praktyce klinicznej, nie polegając jedynie na mechanizmach patofizjologicznych oraz opinii starszych kolegów?

5 kroków EBM – przypadek kliniczny

Żeby zrozumieć jak aplikować EBM w praktyce, przedstawię przypadek pacjentki, który omówię w kolejnych rozdziałach.

Wyobraź sobie, że 30-letnia kobieta trafia do oddziału psychiatrycznego z powodu psychozy poporodowej. Niekoniecznie musisz się wyobrazić w roli lekarza tej pacjentki – to potencjalnie mogłaby być Twoja przyjaciółka, znajoma znajomej lub ktoś z rodziny. Zanim przejdziemy do omówienia pytań, jakie moglibyśmy sobie zadać, wytłumaczę na czym polega choroba pacjentki.

Psychoza poporodowa obejmuje szybkie wystąpienie objawów psychotycznych, w tym możliwych halucynacji i urojeń, dziwacznych zachowań, dezorientacji i dezorganizacji.⁹ Psychoza poporodowa stanowi stan nagły i zazwyczaj wymaga szybkiej interwencji i hospitalizacji, a także kompleksowej oceny medycznej i leczenia psychiatrycznego. Mówiąc prostszym językiem, kobieta z psychozą poporodową może słyszeć nieistniejące głosy, które każą jej zrobić ludziom wokół krzywdę. Może też mieć niezwykle silne przekonanie, że jest zagrożona (co wiąże się z silnym, choć nieraz skrywanym lękiem) i musi natychmiast uciekać ze szpitala z dzieckiem. Potencjalnych manifestacji tej choroby jest wiele. Wiemy z opowieści kobiet dotkniętych psychozą poporodową i ich rodzin, że doświadczenie to odciska piętno na relacji kobiety z samą sobą oraz z bliskimi, w tym nowo narodzonym dzieckiem.¹⁰

Rodzi się wiele pytań. Jak leczyć pacjentkę? Jak skuteczne jest leczenie? Jak długo powinno trwać? Jakie są szkody tego leczenia? Jakie jest rokowanie pacjentki? Czy może karmić piersią w związku z przyjmowanymi lekami? Jaki to będzie miało wpływ na rozwój dziecka? Jak rozmawiać z taką pacjentką? Co ona przeżywa? Co przeżywa rodzina? Pytań jest wiele i dotyczą różnych zagadnień. Można je jednak podzielić zgrubnie na kilka kategorii tematycznych:

1. **Leczenie** – skuteczność i szkody
2. **Prognoza** – co się stanie w przyszłości z tą pacjentką?
3. Powszechność – jak częsta jest choroba?
4. Etiologia – jakie są czynniki ryzyka lub czynniki chroniące przed wystąpieniem choroby?
5. Diagnoza – czy mamy test diagnostyczny (np. test z krwi lub objaw kliniczny), który pozwala przewidzieć wystąpienie lub potwierdzić diagnozę psychozy poporodowej?
6. Ludzkie doświadczenie – w jaki sposób pacjentka przeżywa chorobę? W jaki sposób przeżywa to jej rodzina? W jaki sposób lekarz może zapewnić adekwatne wsparcie?

W ramach tego skryptu omówię szerzej oparte o EBM podejście do dwóch pierwszych kategorii pytań, tj. leczenia i prognozy. Są to uniwersalne problemy, których solidne zrozumienie powinien mieć lekarz-praktyk. Żeby zaadresować pytania z kategorii 3-6 wymagana jest bardziej specjalistyczna wiedza oraz znajomość rodzajów badań, które wykraczają poza wstęp do EBM-u. Niektóre z nich omówię jedynie pobieżnie w rozdziale poświęconym czytaniu poszczególnych rodzajów badań. Dla lekarza lub rodziny pacjentki pytania dotyczące leczenia oraz prognozy zdają się też najbardziej palące. Kobieta w psychozie niekoniecznie będzie zainteresowana powszechnością choroby, czynnikami ryzyka czy metodą diagnozy. Jej lekarz prawdopodobnie

też nie – przynajmniej nie w momencie przejęcia opieki nad chorą. I z pewnością nie, jeśli będzie zestresowanym rezydentem psychiatrii 1. roku.

Skupmy się na dwóch zagadnieniach:

1. Leczenie

- a. **Czy** istnieje skuteczne leczenie?
- b. **Jak skuteczne** jest leczenie?
- c. **Jakie** są szkody leczenia?
- d. **Jak duże** są szkody leczenia?

2. Prognoza

- e. **Czy** epizod psychozy poporodowej może przerodzić się w inną, przewlekłą chorobę?
- f. **Jak duże** jest prawdopodobieństwo rozwinięcia przewlekłej choroby po epizodzie ostrej psychozy poporodowej?

Zwróć proszę uwagę, że zadaję dwa rodzaje pytań – „czy?” i „jak bardzo?”. Podręczniki, z których się uczymy, często mówią o skuteczności lub braku skuteczności danej metody. Nie zawsze jednak mówią jak dużego efektu należy się spodziewać. Pamiętam, gdy jako młody lekarz włączyłem lek przeciwkrzepliwy u pacjenta, który przeszedł zatorowość płucną (zatkanie tętnicy płucnej skrzepem pochodzącym zazwyczaj z żył kończyn dolnych). Oczywiście chciałem zapobiec kolejnemu takiemu zdarzeniu, byłem bardzo przejęty i miałem nadzieję, że moja interwencja będzie skuteczna (jak to młody rezydent). Pamiętam, jak konsultantka chorób wewnętrznych powiedziała mi wtedy „Panie doktorze, proszę pamiętać, że ten lek zmniejszy prawdopodobieństwo, ale nie da panu gwarancji, że zatorowość nie nawróci.” Uświadomiłem sobie wtedy, że wiem, jak należy leczyć, ale w zasadzie nie wiem jak dużego efektu mogę się spodziewać po zastosowaniu interwencji – zmniejszenia ryzyka o 5%, 10%, 50%, a może 95%? Taka niewiedza może być źródłem

niepewności, która prowadzi do lęku i stresu – a tego w pracy mamy dość. Umiejętność interpretacji wyników badań ogranicza tę niepewność i pozwala zrozumieć jak dużego efektu spodziewać się po zastosowanym leczeniu.

Do tej pory udało nam się zadać pytania dotyczące dwóch ważnych obszarów. **Zadanie precyzyjnego pytania** jest pierwszym z tzw. 5 kroków EBM (Rycina 3). W następnej kolejności należy **znaleźć odpowiedź** na pytanie. Po zidentyfikowaniu badania, które odpowiada na nasze pytanie, należy choć w minimalnym zakresie **zweryfikować**, czy badanie jest dość dobrej jakości. Później zdobyte informacje należy **zastosować w praktyce** i ewentualnie **zastanowić się, co można usprawnić**.



Rycina 3. 5 kroków medycyny opartej na dowodach naukowych

W kolejnej części wytłumaczę Ci, w jaki sposób można sformułować pytanie kliniczne oraz czemu to robimy.

Precyzowanie pytania klinicznego – PICO/PECO

Leczenie

Wyobraź sobie, że nic nie wiesz o leczeniu ani prognozie tej pacjentki (jeżeli jesteś studentem przed zajęciami z psychiatrii, to prawdopodobnie zresztą tak jest). Chcesz sprawdzić czy jakiś lek pomoże zmniejszyć dolegliwości pacjentki, a najlepiej wyleczyć stan psychotyczny. Przy okazji dobrze jest też zweryfikować czy lek nie będzie powodował u pacjentki działań niepożądanych, czyli szkód leczenia. Te pytania można zawrzeć w ramach tzw. PICO, czyli akronimu opisującego składowe pytania klinicznego: (P)opulation, (I)ntervention, (C)omparison, (O)utcome, czyli populacja, interwencja, porównanie i wynik (leczenia). Na czym to polega?

Populacja, którą jesteśmy zainteresowani to pacjenci podobni do naszej pacjentki. Jak możemy ją opisać? Jest kobietą, ma 30 lat, ma psychozę poporodową.

Nie określiliśmy interwencji, jaką chcielibyśmy zbadać. Nie mamy na to na razie konkretnego pomysłu, ale jest to zapewne jakiegoś rodzaju lek.

Interesuje nas czy lek będzie lepszy niż... zrobienie niczego. W końcu zawsze wyjściową alternatywą dla leczenia jest po prostu nieleczenie. W przypadku leków podanych w tabletkach można oczywiście zastanowić się czy aktywna substancja będzie lepsza niż tzw. placebo, czyli tabletką bez aktywnej substancji.

W końcu interesuje nas wynik czy efekt leczenia. W tym wypadku pacjentka ma psychozę. A my chcemy, żeby jej nie miała. To jest wynik, który chcemy osiągnąć – ustąpienie psychozy. Moglibyśmy sformułować go inaczej – np. jako zmniejszenie nasilenia dolegliwości psychotycznych. Nawet jeśli w pełni nie bylibyśmy w stanie wyeliminować procesu psychotycznego, czy możemy chociaż uśmierzyć dolegliwości? Pamiętaj też proszę, że wyniki leczenia mogą być ujemne, tj. lek może wyrządzić szkody. Dobrze by było wiedzieć jakich szkód oczekiwać u pacjentki przy włączeniu leczenia.

Podsumujmy nasze pytanie kliniczne w formacie PICO:

(P)opulacja – kobiety w wieku rozrodczym z psychozą poporodową

(I)nterwencja – lek (nie wiemy jeszcze jaki)

(C)omparison, tj. porównanie – brak leku lub placebo (tj. interwencja porównawcza)

(O)utcome, tj. wyniki leczenia – ustąpienie psychozy lub zmniejszenie nasilenia dolegliwości psychotycznych; szkody leczenia

Prognoza

Teraz rozważmy kwestię prognozy. Zastanawia nas czy i w jakim stopniu wystąpienie psychozy poporodowej wpłynie na zdrowie psychiczne naszej pacjentki w przyszłości. Jesteśmy zainteresowani tym, czy wystąpienie tej choroby w porównaniu z brakiem ekspozycji na tę chorobę będzie miało wpływ na losy chorej. Tego rodzaju pytanie można opisać przy pomocy akronimu PECO – trzy z czterech liter już omówiliśmy, zaś (E)xposition oznacza ekspozycję – w tym wypadku zachorowanie na psychozę poporodową. Jak będą wyglądały składowe pytania klinicznego zadanego w formacie PECO?

(P)opulacja – kobiety w wieku rozrodczym

(E)xposition, tj. ekspozycja – zachorowanie na psychozę poporodową

(C)omparison, tj. porównanie – brak zachorowania na psychozę poporodową

(O)utcome, tj. wynik kliniczny – zachorowanie w przyszłości na przewlekłą chorobę psychiczną (np. schizofrenię lub chorobę afektywną dwubiegunową)

Dlaczego w przypadku leczenia mówimy o (I)ntervention, a w przypadku prognozy o (E)xposition? W pierwszym wypadku zadaliśmy pytanie dotyczące skuteczności czy szkód leczenia. Leczenie z definicji jest interwencją – czymś dodanym. Interwencja nie zawsze musi mieć charakter farmakologiczny — może obejmować także modyfikacje stylu życia, takie jak zmiany żywieniowe czy 45-minutowa praktyka jogi. Prawdopodobnie w tym wypadku byłby to kiepski pomysł i raczej nikt nie dostałby pieniędzy na zbadanie takiej formy leczenia. Niemniej byłaby to ingerencja w normalny tryb życia pacjentki. Gdy zadajemy pytanie o skuteczność czy szkodę leczenia zawsze używamy PICO, bo pytamy o interwencję.

W drugim przypadku my jako lekarze nie stosujemy żadnej interwencji. Obserwujemy zjawisko, na które pacjentka została wyeksponowana – w tym wypadku na epizod psychozy. Pytamy czy ta ekspozycja (epizod psychozy) ma wpływ na dalsze życie pacjentki (np. rozwinięcie przewlekłej choroby psychicznej). Nie stosujemy jednak psychozy jako interwencji. Po pierwsze, nie umiemy. Po drugie, nawet gdybyśmy umieli, to nie powinniśmy. Obserwujemy zatem to, co dzieje się z pacjentką. Innymi przykładami ekspozycji mogłoby być palenie tytoniu, obecność choroby psychicznej u jednego z rodziców czy obecność danego wariantu genetycznego. W każdym z tych przypadków moglibyśmy spytać czy ekspozycja na te bodźce lub dane z wywiadu pacjentki w porównaniu z brakiem takiej ekspozycji, są powiązane z wystąpieniem choroby w przyszłości.

Dlaczego precyzujemy pytania kliniczne? Po pierwsze, żeby uporządkować nasze myślenie. Czego dokładnie chcemy się dowiedzieć z literatury? Jeżeli pytanie jest ogólne lub brak mu struktury, zmniejszamy swoje szanse na znalezienie konkretnej odpowiedzi. Po drugie, jeżeli precyzyjnie określimy nasze pytanie, będziemy wiedzieć, jakiego rodzaju dowodu szukać – o tym więcej w kolejnym rozdziale. Po trzecie, określenie pytania pomaga nam skutecznie wyszukiwać informacje w medycznych bazach danych – o tym więcej w rozdziale nt. szukania informacji.

Na razie zastanówmy się, jakiego rodzaju odpowiedzi na zadane przez nas pytania w ogóle istnieją.

Rodzaje odpowiedzi

„Słuchaj, następnym razem jak będziesz miał katar, jedz codziennie pół awokado. Ono działa przeciwzapalnie, zmniejsza stres oksydacyjny, jest bogate w dobre kwasy tłuszczowe. To oczywiste, że silnie działa przeciwinfekcyjnie. Ja brałem codziennie jak byłem chory i katar mi przeszedł po 5 dniach, a mówią, że normalnie trwa 7!”

Przekonało Cię to? Mnie nie. W tym rozdziale pokażę Ci, że nauka może dawać bardzo różne rodzaje odpowiedzi na to samo pytanie — i że nie wszystkie są równie wiarygodne. Poniżej przedstawię rodzaje badań oraz ich miejsce w tzw. hierarchii dowodów naukowych.

Opisy przypadków

Powyższy opis można potraktować jako tzw. **opis przypadku** – inaczej **case report** albo dowód anegdotyczny. Takie historie bywają powtarzane, ale pytanie brzmi, na ile są wiarygodne? Jeden skuteczny przypadek leczenia nie znaczy, że interwencja jest skuteczna dla innych chorych. Pamiętaj proszę, że w literaturze medycznej wyniki pozytywne (tj. wskazujące na skuteczność interwencji) są zdecydowanie częściej i szybciej publikowane niż wyniki negatywne (tj. brak efektu).¹¹ Oznacza to, że na każdego opublikowanego udanego „case’a” gdzieś na świecie w szufladach naukowców i lekarzy może być 5 innych nieudanych. Po drugie, do każdej interwencji medycznej można stworzyć hipotezę,

uzasadniająca, dlaczego leczenie powinno zadziałać. W wielu przypadkach jednak myślenie takie jest niewystarczające, jak zobaczyliśmy wcześniej na podstawie historii flekainidu. Kolejnym argumentem podważającym wiarygodność opisów przypadków jest brak grupy kontrolnej. Oznacza to, że opisany przypadek nie został porównany z podobnym, który otrzymałby inne leczenie lub brak leczenia. W naszym przypadku (infekcji górnych dróg oddechowych) naturalny przebieg choroby jest samoograniczający się. Innymi słowy, nawet bez dodatkowej interwencji dojdzie do ustąpienia objawów.

Warto wspomnieć tutaj o tzw. **efekcie placebo**. Polega on na tym, że pacjent zgłasza poprawę stanu zdrowia pomimo tego, że stosowana substancja nie ma rzeczywistego działania leczniczego. W przypadku leczenia zaburzeń depresyjnych czy lękowych pacjenci przyjmujący placebo zgłaszają zmniejszenie nasilenia dolegliwości w czasie.¹² Wiemy jednak, że leki przeciwdepresyjne czy przeciwłękowe mają wyraźnie większy wpływ na redukcję dolegliwości niż placebo.¹³ Efekt placebo jest zróżnicowany w zależności od leczzonej choroby i wyniku, który mierzymy. Chociażby w zaburzeniach lękowych efekt placebo jest większy niż w przypadku zaburzeń psychiatrycznych takich jak schizofrenia. Efekt ten jest też bardziej wyrażony w przypadku subiektywnej oceny objawów przez pacjenta (np. bólu), zaś bardziej ograniczony lub nieobecny w przypadku leczenia tzw. twardych punktów końcowych jak zgon czy nawrót choroby nowotworowej.¹⁴ Warto jednak zaznaczyć, że poprawa stanu zdrowia osób przyjmujących placebo nie zawsze wynika z samego efektu placebo. Często jest to też konsekwencja naturalnego przebiegu choroby, której objawy mogą ograniczać się w czasie. Może to być też konsekwencja **regresji do średniej**.¹⁵ Zjawisko to polega na tym, że pojedyncze, ekstremalne pomiary mają naturalną skłonność do normalizacji w czasie. Np. gdy wartość ciśnienia skurczowego pacjenta

wynosi 190 mmHg, jest wysoce prawdopodobne, że jeśli zmierzmy ciśnienie ponownie, wartość ta zmaleje. Często nasilenie objawów wśród pacjentów włączonych do badania jest najwyższe na początku obserwacji. W naturalny sposób nasilenie to może maleć nawet bez zastosowania interwencji.

Artykuły omawiające **serię przypadków** inaczej nazywają się **case series**, natomiast ich ograniczenia są podobne. Podsumowując:

- Przypadki udanego leczenia są częściej publikowane niż nieudane próby.
- Brak możliwości uogólnienia wyników ze względu na to, że wnioski odnoszą się jedynie do danego pacjenta lub grupy pacjentów.
- Brak grupy kontrolnej nie pozwala na określenie skuteczności w porównaniu z brakiem leczenia lub alternatywnym leczeniem.
- Brak standaryzacji daje autorom dużą wolność w zakresie rygoru opisywania procesu leczenia.
- Brak możliwości odróżnienia przypadkowego następstwa czasowego od związku przyczynowego.

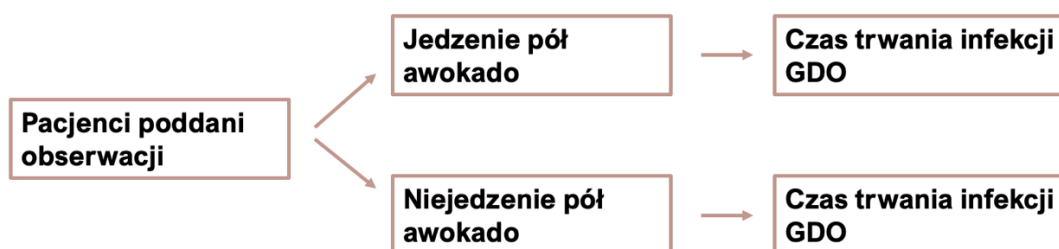
Tego typu publikacje mają jednak swoje miejsce w literaturze medycznej, szczególnie jeśli mamy do czynienia z:

- Rzadką populacją (np. rzadka choroba genetyczna), w przypadku której przeprowadzenie większego badania byłoby niemożliwe lub bardzo trudne.
- Rzadką manifestacją powszechnie występującej choroby lub nietypowym przebiegiem choroby bądź procesu leczniczego.
- Nowym zjawiskiem, które nie zostało do tej pory opisane – dobrym przykładem jest wczesny opis występowania rzadkiego nowotworu (mięśaka Kaposiego) u mężczyzn zakażonych HIV na początku lat 80.¹⁶

- Niestandardowym zastosowaniem leków (*off-label*) bądź użyciem innowacyjnej procedury (np. chirurgicznej).

Badania obserwacyjne

Jaką mamy alternatywę? Wróćmy do przykładu z pół awokado celem leczenia infekcji górnych dróg oddechowych (GDO). Powiedzmy, że chcemy przeprowadzić bardziej rygorystyczne badanie na większej grupie osób, a do tego chcielibyśmy uwzględnić grupę kontrolną. W tym celu moglibyśmy podjąć się **obserwacji** osób z infekcją GDO i zbierać dane na temat tego, kto z nich korzystał z metody awokado, a kto nie. W tym wypadku nie ingerujemy w decyzje pacjentów, czyli nie interweniujemy, lecz jedynie **obserwujemy** naturalny rozwój wydarzeń. Tak zaplanowane badanie moglibyśmy przedstawić graficznie przy pomocy poniższego schematu (Rycina 4).



Rycina 4. Schemat badania obserwacyjnego (kohortowego)

Powyżej widzimy klasyczny schemat badania kohortowego (*cohort study*). Badanie kohortowe polega na obserwacji kohorty (grupa ludzi) oraz zbieraniu danych na temat ekspozycji, tj. czynników, na które uczestnicy badania są wystawieni. Nawet w kontekście leczenia mówimy tutaj o ekspozycji ze względu na to, że uczestnicy sami decydują, czy przyjmować awokado – nie jest to interwencja zlecona przez lekarza czy badacza. Ekspozycjami mogłyby też być inne czynniki – palenie

papierosów, picie alkoholu, przyjmowanie witaminy D, miejsce zamieszkania itd. Są to zatem czynniki obserwowane w życiu danego uczestnika badania. W tym kontekście pytanie kliniczne mogłoby zostać sformułowane w formacie PECO:

P(opulation) – osoby z infekcją GDO

E(xposition) – jedzenie pół awokado dziennie

C(omparison) – niejedzenie awokado

O(utcome) – czas trwania infekcji GDO

Tego typu badanie kohortowe może mieć charakter **prospektywny**, tj. zaczynamy zbierać dane w momencie wystąpienia infekcji u uczestnika badania. Może też mieć charakter **retrospektywny**, tj. po przebytej infekcji zbieramy dane na temat tego, czy dana osoba jadła awokado w jej trakcie. Dane takie można by pozyskać od samego uczestnika badania (co wiąże się z ryzykiem, że nie przypomni sobie, jak było w rzeczywistości lub poda nieprawdziwe informacje) lub z innych źródeł – od bliskich osób czy z dokumentacji medycznej (choć odnotowanie faktu jedzenia awokado z oczywistych względów mogłoby zostać pominięte). Właśnie z tego względu, że jakość zbieranych informacji jest bardziej wątpliwa w przypadku badania retrospektywnego, silniejszej jakości dowody uzyskujemy, prowadząc badania prospektywne. Podsumowując, badanie kohortowe definiuje grupy według ekspozycji (np. narażeni i nienarażeni), a następnie obserwuje wystąpienie wyniku w czasie. Istnieją także inne rodzaje badań obserwacyjnych, jednak ich szczegółowe omówienie wykracza poza zakres tematyczny niniejszego skryptu.

Na czym jednak polega wyzwanie w interpretacji wyników tego typu badania? Otóż na tym, że osoby decydujące się korzystać z awokado mogą istotnie różnić się od osób, które się na to nie zdecydowały. Osoby z pierwszej grupy mogą być młodsze, mieć wyższy status socjoekonomiczny, mieć mniej nasilone objawy infekcji. Głównym

ograniczeniem wpisanym w jakikolwiek rodzaj badania obserwacyjnego jest to, że może istnieć cecha lub kilka cech, które różnią porównywane grupy. Wtedy powstaje pytanie – czy obserwowana między grupami różnica w czasie trwania infekcji GDO wynika z jedzenia awokado czy też z innych względów? To zjawisko nosi nazwę zakłócania (*confounding*) i jest jednym z głównych ograniczeń badań obserwacyjnych.

Zakłócanie (*confounding*) jako ograniczenie badań obserwacyjnych

Żeby lepiej zrozumieć ryzyko, które niesie ze sobą zakłócanie rozważmy bardziej jaskrawy przykład.¹⁷ Wyobraź sobie, że prowadzisz badanie obserwacyjne i zauważasz, że osoby, które noszą w kieszeni zapalniczkę, znacznie częściej chorują na raka płuca. Dane statystyczne wskazują na silną i nieprzypadkową korelację. Czy to oznacza, że noszenie zapalniczki powoduje raka płuca? Oczywiście nie. Zapewne domyślasz się, że za tym związkiem stoi coś więcej – mianowicie palenie papierosów. To właśnie jest **zjawisko zakłócania** (*confounding*) – częsty problem w medycynie. Sprawia, że w badaniu możemy dostrzec przypadkowe związki przyczynowe, które w rzeczywistości nie istnieją, lub – co gorsza – nie dostrzegamy tych, które faktycznie tam są.

Zmienna zakłócająca to czynnik wpływający na chorobę, który jest nierówno rozłożony między grupami, które porównujemy (np. osoby palące vs niepalące). Aby dany czynnik (nazwijmy go Z) mógł zostać uznany za zakłócający, w relacji między ekspozycją (E) a chorobą (C), musi spełnić trzy żelazne warunki:

1. Musi być samodzielnie powiązany z badaną chorobą. Jak pokazałem wyżej, istnieje korelacja pomiędzy posiadaniem zapalniczki a diagnozą raka płuca.

2. Musi być powiązany z ekspozycją w grupie badanej. Osoby posiadające zapalniczkę to zazwyczaj te same osoby, które palą papierosy.
3. Nie może być skutkiem ekspozycji. Zakładamy, że noszenie zapalniczki nie „powoduje”, że ktoś zaczyna palić (to palenie powoduje posiadanie zapalniczki).

W przedstawionym wcześniej badaniu na temat czasu trwania infekcji GDO spożywanie awokado może wiązać się ze zdrowszym stylem życia (w tym mniejszym ryzykiem palenia papierosów), lepszą higieną snu czy unikaniem innych używek. Te z kolei mogą mieć korzystny wpływ na czas trwania infekcji górnych dróg oddechowych. Równocześnie nie możemy stwierdzić, że jedzenie awokado powoduje wszystkie inne zachowania prozdrowotne. Raczej z nimi współwystępuje.

W badaniach obserwacyjnych grupy tworzą się same w zależności od wystąpienia lub braku wystąpienia ekspozycji. Prawie zawsze grupy różnią się między sobą czymś więcej niż tylko badaną cechą. Wpływ zakłócania na wnioski może być trojaki:

- Przeszacowanie efektu: Widzisz silny związek (jak zapalniczka i rak), który jest w dużej mierze sztuczny.
- Niedoszacowanie efektu: Czasem zakłócając działa w drugą stronę i „maskuje” prawdziwą zależność (tzw. *negative bias*).
- Odwrócenie kierunku: W skrajnych przypadkach zakłócanie może sprawić, że czynnik szkodliwy wyda się pomocny lub odwrotnie.

To prowadzi do tego, że badane grupy nie są ze sobą w 100% porównywalne. Idealne badanie to takie, w którym grupa eksponowana jest identyczna z nieeksponowaną pod każdym względem (wiek, dieta, geny, styl życia) poza badaną ekspozycją. Jeśli tak nie jest, wyniki należy interpretować bardzo ostrożnie.

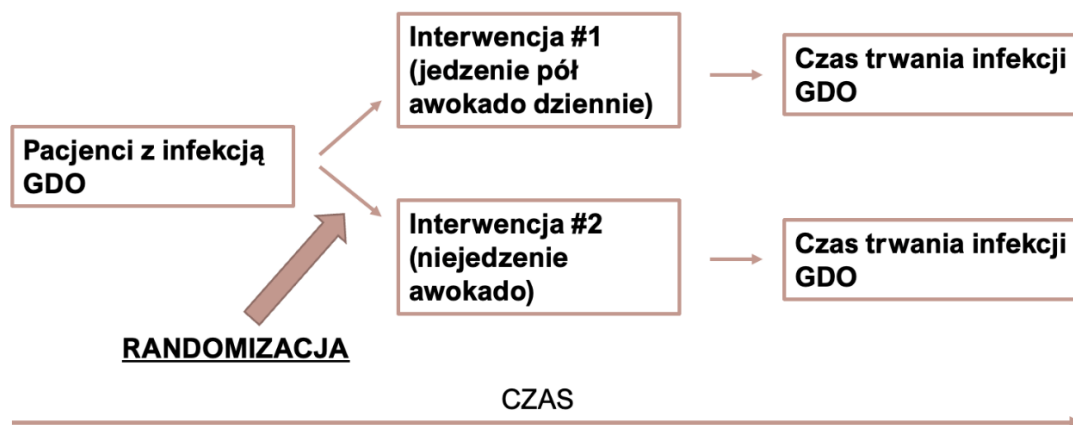
Jest sposób, żeby ograniczyć wpływ czynników zakłócających na wyniki. Czytając artykuł możesz natrafić na informację na temat poprawek statystycznych (*statistical adjustment*).¹⁸ Autorzy rzetelnych prac starają się matematycznie „wyczyścić” wyniki z wpływu znanych zakłócaaczy (np. wieku czy palenia). Pamiętaj jednak o dwóch pułapkach:

1. Niezmierzone zakłócacze: Możemy poprawić wynik o to, co zmierzaliśmy. Ale co z czynnikami, o których badacze zapomnieli lub których nie potrafili zmierzyć (np. predyspozycje genetyczne)? One nadal tam są i mają wpływ na wyniki.
2. Nadmierna poprawka: Nie wolno „poprawiać” wyniku o czynniki, które są częścią mechanizmu działania (np. nie poprawiamy wpływu otyłości o nadciśnienie, jeśli to otyłość powoduje nadciśnienie). To doprowadziłoby do błędnych wniosków.

Badania obserwacyjne mówią nam zatem przeważnie o korelacjach czy powiązaniach. Jedynie rygorystycznie przeprowadzone badania obserwacyjne z dużą liczbą uczestników mogą sugerować związek przyczynowo-skutkowy. Jak w takim razie możemy przeprowadzić badanie w taki sposób, żeby możliwie jak najbardziej trafnie odpowiedziało na nasze pytanie dot. skuteczności danej interwencji?

Randomizowane badania kliniczne

Złotym standardem badania skuteczności oraz szkód leczenia jest randomizowane badanie kliniczne (*randomised controlled trial – RCT*). Schemat takiego badania przedstawiono poniżej (Rycina 5).



Rycina 5. Schemat randomizowanego badania klinicznego

W tego typu badaniu pacjenci są przydzielani przez badaczy do jednej z (najczęściej) dwóch grup: grupy z badaną interwencją (w tym wypadku awokado) oraz grupy porównawczej. Przydział pacjentów nie odbywa się jednak na podstawie preferencji co do stosowanej interwencji, ale w sposób losowy. Proces ten nazywa się randomizacją.

Randomizacja odpowiednio dużej grupy pacjentów zapewnia równe rozłożenie potencjalnych czynników zakłócających (znanych i mierzalnych, jak i nieznanymi czy niemierzalnych) między dwoma grupami. Pomyśl o tym jak o rzucie monetą. W tym przykładzie dana moneta symbolizowałaby daną zmienną zakłócającą – np. płeć. Jeżeli losowo rzucimy monetą 100 razy, szansa, że wypadnie idealnie 50 razy reszka i 50 razy orzeł jest dość niewielka (ok. 8%). Jednak prawdopodobieństwo tego, że orzeł wypadnie między 40 a 60 razy (zaś reszka odpowiednio między 60 a 40) wynosi już około 95%. Innymi słowy, losowe przydzielenie pacjentów do grup daje nam bardzo dużą szansę tego, że większość potencjalnych zmiennych zakłócających rozłoży się podobnie między dwoma grupami. Może to jednak nie być prawdziwe dla małych grup badawczych. W takich przypadkach ryzyko nierównomiernego rozkładu zmiennych między grupami rośnie. Istnieją narzędzia metodologiczne

(stratyfikacja, minimalizacja)¹⁹, które ograniczają ten problem, jednak ich omówienie wykracza poza zakres tego skryptu.

W badaniach randomizowanych często używa się zaślepienia (*blinding / masking*). Polega to na tym, że pacjent nie wie, czy dostaje lek, czy placebo (pierwsze zaślepienie) i osoba dająca interwencję również nie wie, czy wydaje lek, czy placebo (drugie zaślepienie). Trzecie zaślepienie jest możliwe na etapie analizy danych, kiedy najczęściej statystyk nie wie czy grupa A lub B otrzymała aktywny lek czy placebo. Zaślepienie zapobiega przedwczesnej rezygnacji z udziału w badaniu, preferencyjnemu traktowaniu własnych pacjentów oraz manipulacji danych na poziomie analizy. Zapobiega wpływowi oczekiwań i przekonań na zgłaszane objawy, zachowanie uczestników, decyzje terapeutyczne personelu oraz interpretację wyników, co jest szczególnie istotne przy ocenach subiektywnych (np. jakość życia, nastrój, ból). Maskowanie (zaślepienie) zwiększa wiarygodność i interpretowalność wyników, minimalizując ryzyko zawyżenia efektu leczenia lub zaniżenia skuteczności placebo.²⁰

W badaniach randomizowanych, jak i w badaniach obserwacyjnych wynik leczenia może nosić nazwę **punktu końcowego**. Odnosi się to do zdarzenia (pożądanego lub nie), które odnotowuje się u danego pacjenta w trakcie badania. Przykładem może być wyzdrowienie, zgon, nawrót choroby, remisja. Są to zatem wyniki zero-jedynkowe w przeciwieństwie do innych wyników, które mogą być mierzone na skalach, np. nasilenie bólu (od 0 do 10), nasilenie objawów depresyjnych czy czas trwania infekcji.

Ze względu na powyższe cechy randomizowane badanie kliniczne uznaje się za złoty standard testowania skuteczności leczenia.

Do najbardziej istotnych zalet należą:

1. Randomizacja zapewnia losowe rozłożenie **znanych i nieznanymi czynników** (np. wiek, płeć, stopień nasilenia choroby, ale również predyspozycje genetyczne), mogących wpływać na badany wynik
2. Zapobiega selektywnemu przydzielaniu pacjentów do grup
3. Badanie może być **zaślepienie** na trzech poziomach – pacjenta, osoby stosującej interwencję (np. podającej lek) i osoby analizującej dane (podwójne/potrójne zaślepienie)
4. Rygorystyczna metodologia pozwala na rzetelną ocenę działania nowego leku względem interwencji kontrolnej (placebo/standardowej opieki/starego leku) – stąd **można wyciągnąć wnioski przyczynowe** na temat skuteczności i szkód leczenia

Przeglądy systematyczne i meta-analizy

Niestety nieraz dzieje się tak, że wyniki poszczególnych badań obserwacyjnych czy randomizowanych badań klinicznych dają sprzeczne lub nie w pełni jednolite wyniki. Przyczyną jest to, że w różnych badaniach:

- Populacje osób badanych mogą różnić się między sobą.
- Stosowane interwencje lub badane ekspozycje mogą nie być identyczne. Mogą mieć zmienne nasilenie czy częstotliwość, a w przypadku leków różnić mogą się dawki, droga podania, czas trwania leczenia.
- Interwencje terapeutyczne mogą być porównywane z różnymi grupami kontrolnymi – np. placebo, brakiem leczenia, innym mniej skutecznym lekiem.
- Liczebności badanych populacji mogą różnić się, dając różną precyzję otrzymanych wyników. Przeważnie badania prowadzone na mniejszych grupach dają mniej precyzyjne wyniki oraz większą

niepewność co do skuteczności interwencji czy istotności badanej ekspozycji.

- Warunki przeprowadzenia badania mogą mieć wpływ na wyniki. Bardzo podobne badanie przeprowadzone w czołowym ośrodku klinicznym może wykazać inną skuteczność leku niż podobne badanie przeprowadzone w niedofinansowanym ośrodku powiatowym o gorszej infrastrukturze.
- Nieraz umiejętności badaczy mają znaczenie przy analizie wyników badania. Szczególnie w badaniach dot. wyników operacji chirurgicznych ważne jest to, czy pacjenci są operowani przez doświadczonych klinicystów, czy przez młodszych lekarzy, którzy jeszcze pracują nad swoimi umiejętnościami.
- Wyniki badania mogą być różnie mierzone. Np. w badaniach skuteczności leków przeciwdepresyjnych stosowane są różne skale do oceny nasilenia objawów, co może utrudniać interpretację porównawczą badań, używających różnych narzędzi.

W sytuacjach, w których istnieje duża niepewność co do skuteczności leczenia bądź istotności danej ekspozycji, pomocne jest wykonanie przeglądu systematycznego z lub bez meta-analizy.

Przegląd systematyczny (*systematic review*) to badanie przeprowadzone według określonego protokołu, którego zadaniem jest przeszukanie baz danych, selekcja oryginalnych artykułów naukowych oraz synteza **wszystkich** dostępnych dowodów naukowych na dany temat. Innymi słowy, autorzy przeglądu systematycznego w sposób uporządkowany przeszukują dostępne bazy danych w poszukiwaniu opublikowanych (oraz często nieopublikowanych) wyników badań, które krytycznie oceniają oraz przeprowadzają ich syntezę celem zmniejszenia niepewności wokół zadanego pytania klinicznego. **Meta-analiza**

(*meta-analysis*) z kolei to metoda statystyczna, używana do syntezy danych, która jest opcjonalnym elementem przeglądu systematycznego.

W większości przypadków przegląd systematyczny badań dot. sprecyzowanego pytania klinicznego daje odpowiedź wyższej jakości niż pojedyncze badanie. Nie jest to jednak nieomylna reguła. Czasem pojedyncze, nowsze badanie wysokiej jakości przeprowadzone na dużej grupie pacjentów da lepszą odpowiedź niż nieco starszy przegląd systematyczny, podsumowujący wcześniejsze badania gorszej jakości. Pamiętaj proszę, że jakość przeglądu systematycznego jest pochodną jakości badań, które zostały do niego włączone. Przeglądy systematyczne identyfikują luki w wiedzy oraz priorytety badawcze – wskazują obszary, gdzie brakuje danych lub gdzie wyniki są niejednoznaczne, co może nadać kierunek przyszłym badaniom.

Po zapoznaniu się z możliwymi rodzajami badań, należy zadać sobie pytanie, które z nich da nam najlepszą odpowiedź na zadane przez nas pytania.

Hierarchia medycznych dowodów naukowych

Wyobraź sobie, że jesteś w pracy i spotykasz się z wyzwaniem klinicznym, z którym nie masz doświadczenia. Musisz podjąć dość szybką decyzję kliniczną, lecz nie masz czasu na wielogodzinne przeszukiwanie baz danych ani tym bardziej na samodzielne przeprowadzanie przeglądu systematycznego. Potrzebujesz prostego sposobu, który pozwoli Ci sprawnie znaleźć najbardziej wiarygodną odpowiedź.

W tym właśnie pomaga Tabela Poziomów Dowodów Oxfordzkiego Centrum Medycyny Opartej na Dowodach Naukowych (*Oxford Center for Evidence-Based Medicine*).²¹ Jest to narzędzie zaprojektowane tak, by służyło jako skrót dla zapracowanych lekarzy (Tabela 1).

Tabela 1. Hierarchia dowodów Oxfordzkiego Centrum Medycyny Opartej na Dowodach

Oxfordzkie Centrum Medycyny Opartej na Dowodach 2011 Poziomy dowodów

Pytanie	Krok 1 (Poziom 1*)	Krok 2 (Poziom 2*)	Krok 3 (Poziom 3*)	Krok 4 (Poziom 4*)	Krok 5 (Poziom 5)
Jak powszechny jest problem?	Aktualna ankieta (lub spisy ludności) z lokalną, losowo wybraną próbą	Przegląd systematyczny ankiet pozwalających na dopasowanie do lokalnych warunków**	Lokalna nierandomizowana próba**	Opis serii przypadków**	n/d
Czy test diagnostyczny / metoda monitorowania jest dokładna? (Diagnoza)	Przegląd systematyczny badań przekrojowych z konsekwentnie stosowanym standardem referencyjnym i zaślepieniem	Pojedyncze badanie przekrojowe z konsekwentnie stosowanym standardem referencyjnym i zaślepieniem	Badanie na grupie pacjentów niewłączanych kolejno lub bez konsekwentnie stosowanego standardu referencyjnego**	Badanie kliniczno-kontrolne lub stały / inny niż niezależny standard referencyjny**	Rozumowanie oparte o mechanizmy
Co się stanie, jeśli nie zastosujemy leczenia? (Prognoza)	Przegląd systematyczny podłużnych badań obserwacyjnych kohort zrekrutowanych w momencie wystąpienia ekspozycji / zdarzenia (<i>Inception cohort studies</i>)	Podłużne badanie obserwacyjne pojedynczej kohorty zrekrutowanej w momencie wystąpienia ekspozycji / zdarzenia (<i>Inception cohort study</i>)	Badanie kohortowe lub ramię kontrolne badania randomizowanego*	Opis serii przypadków lub badanie kliniczno-kontrolne lub niskiej jakości kohortowe badanie prognostyczne**	n/d
Czy ta interwencja pomaga? (Korzyści leczenia)	Przegląd systematyczny badań randomizowanych lub badań typu <i>n-of-1</i>	Badanie randomizowane lub badanie obserwacyjne z dramatycznym efektem	Nierandomizowane kontrolowane badanie kohortowe / uzupełniające (<i>follow-up</i>)**	Opis serii przypadków, badanie kliniczno-kontrolne lub badanie z historyczną grupą kontrolną**	Rozumowanie oparte o mechanizmy
Jakie są CZĘSTE szkody? (Szkody spowodowane leczeniem)	Przegląd systematyczny badań randomizowanych, przegląd systematyczny zagnieżdżonych badań kliniczno-kontrolnych, badanie typu <i>n-of-1</i> wśród pacjentów, których dotyczy pytanie lub badanie obserwacyjne z dramatycznym efektem	Pojedyncze badanie randomizowane lub (wyjątkowo) badanie obserwacyjne z dramatycznym efektem	Nierandomizowane kontrolowane badanie kohortowe / uzupełniające (monitorowanie postmarketingowe) zakładając wystarczającą liczebność grupy, by wykluczyć częste szkody. (Dla długoterminowych szkód odpowiedni okres obserwacyjny jest niezbędny)**	Opis serii przypadków, badanie kliniczno-kontrolne lub badanie z historyczną grupą kontrolną**	Rozumowanie oparte o mechanizmy
Jakie są RZADKIE szkody? (Szkody spowodowane leczeniem)	Przegląd systematyczny badań randomizowanych lub badanie typu <i>n-of-1</i>	Badanie randomizowane lub (wyjątkowo) badanie obserwacyjne z dramatycznym efektem			
Czy test (służący do wczesnego wykrycia) jest wart zastosowania? (Badania przesiewowe)	Przegląd systematyczny badań randomizowanych	Badanie randomizowane	Nierandomizowane kontrolowane badanie kohortowe / uzupełniające (<i>follow-up</i>)**	Opis serii przypadków, badanie kliniczno-kontrolne lub badanie z historyczną grupą kontrolną**	Rozumowanie oparte o mechanizmy

* Poziom może być obniżony na podstawie jakości badania, nieprecyzyjności, pośredniości (PICO badania nie odzwierciedla PICO pytania klinicznego), z powodu niespójności między badaniami lub ze względu na bardzo mały rozmiar efektu; Poziom może być podwyższony, jeśli istnieje duży lub bardzo duży rozmiar efektu.

** Jak zawsze, przegląd systematyczny jest przeważnie lepszy niż pojedyncze badanie.

Jak czytać tę tabelę?

- 1. Kolumna „Pytanie” (pionowa oś):** Układa się w naturalny przebieg spotkania klinicznego. Najpierw pytasz: „Jak częsty to problem?” (*prevalence*), potem: „Czy mój test jest dokładny?” (*diagnosis*), następnie: „Co się stanie, jeśli nie podam leku?” (*prognosis*), aż wreszcie: „Czy to leczenie pomoże i jakie niesie szkody?” (*benefits and harms*)
- 2. Kroki od 1 do 5 (Pozioma oś):** To rekomendowana kolejność poszukiwań. Zaczynasz od **Kroku 1** (lewa strona), gdzie zazwyczaj znajdziesz najsilniejsze dowody, takie jak przeglądy systematyczne. Jeśli tam nie znajdziesz odpowiedzi, przesuwasz się w prawo,

do kolejnych „pudełek” z dowodami o potencjalnie mniejszej sile, aż do **Kroku 5**, czyli rozumowania opartego na mechanizmach.

Jak warto myśleć o tej hierarchii? Wyobraź sobie, że przed Tobą stoi pięć pudełek z etykietami: przegląd systematyczny, badanie randomizowane, badanie kohortowe, seria przypadków oraz rozumowanie mechanistyczne. Jeśli szukasz dowodów na korzyści z leczenia, zdrowy rozsądek podpowiada, by najpierw otworzyć pudełko z napisem „przegląd systematyczny”. Dopiero gdy jest ono puste, sięgasz po pojedyncze badania randomizowane, a w ostateczności po opisy przypadków. Przedstawiona tabela to właśnie graficzny zapis tej kolejności.

Zanim zaczniesz korzystać z tabeli, zapamiętaj trzy ważne zasady:

- **To nie jest „książka kucharska”:** Żaden system rankingowy nie zastąpi Twojego osądu klinicznego i sceptycyzmu. Tabela wskazuje, co jest prawdopodobnie najlepszym dowodem, ale zawsze musisz brać pod uwagę jakość konkretnego badania.
- **Perspektywa jakości:** Poziom dowodu może zostać obniżony, jeśli badanie jest słabe metodologicznie lub mało precyzyjne (np. ma niewielu uczestników). Z drugiej strony, badanie obserwacyjne może „wskoczyć” na wyższy poziom, jeśli wykazuje ogromny, wręcz dramatyczny efekt np. pomiędzy ekspozycją i badanym wynikiem. W medycynie przykładem jest leczenie ciężkiej hipoksemii tlenem – nie ma potrzeby randomizowania pacjentów do grupy bez podania tlenu, ponieważ skuteczność jest natychmiastowa i dramatyczna.
- **Czy to o moim pacjencie?** Nawet najlepszy dowód z poziomu 1 nie jest drogą do panaceum. Zanim go zastosujesz, musisz zadać sobie pytanie: „Czy mój pacjent jest wystarczająco podobny do tych z badania?” oraz „Czy jego wartości i sytuacja życiowa pozwalają na takie leczenie?”.

Ostatni punkt jest szczególnie ważny. W nomenklaturze EBM to czy nasz pacjent “pasuje” do populacji badania, które czytamy nazywa się **trafnością zewnętrzną** (*external validity*). Innymi słowy, gdyby mój pacjent przyszedł do ośrodka, gdzie prowadzono dane badanie, czy zostałby do niego włączony? Jeżeli tak, to doskonale. Wyniki (po krytycznej analizie) aplikują się do mojego pacjenta. Mogę z nich skorzystać. Nie należy jednak przekładać wyników badań prowadzonych na 30-latkach na 75-letniego pacjenta, który oczekuje akurat pomocy w Twoim gabinecie.

Przyjrzyjmy się teraz dwóm najważniejszym kategoriom:

1. **Leczenie (Czy ta interwencja pomaga?):**

- o **Krok 1 (Poziom najwyższy):** Przegląd systematyczny badań randomizowanych (RCT) lub badań typu *n-of-1* (szczególny rodzaj randomizowanego badania klinicznego).²² To tutaj szukamy odpowiedzi na pytania o najskuteczniejszy lek dla naszej pacjentki z psychozą poporodową.
- o **Krok 2:** Pojedyncze badanie randomizowane (RCT).
- o **Krok 5 (Poziom najniższy):** Rozumowanie oparte o mechanizmy (*mechanism-based reasoning*). To sytuacja, w której ekspert mówi: „Lek powinien działać, bo blokuje receptor X”, ale nie ma na to twardych dowodów z udziałem pacjentów.

2. **Prognoza (Co się stanie z pacjentem?):**

- o **Krok 1:** Przeglądy systematyczne badań kohortowych zrekrutowanych w momencie wystąpienia zdarzenia (tzw. *inception cohort studies*). Tylko śledzenie pacjentów od samego początku choroby daje nam rzetelny obraz ich przyszłych losów.

Pamiętaj, że przegląd systematyczny jest przeważnie lepszy niż pojedyncze badanie. Zawsze staraj się szukać dowodów z jak najwyższego

poziomu, ale bądź gotów obniżyć swoje zaufanie do nich, jeśli zauważysz błędy w ich wykonaniu. W EBM nie chodzi o ślepią wiarę w hierarchię dowodów, ale o krytyczne myślenie na każdym stopniu tej drabiny. W kolejnym rozdziale zmierzmy się z tym, w jaki sposób zacząć czytać artykuły naukowe oraz w jaki sposób czytać poszczególne rodzaje badań.

Jak czytać publikacje naukowe?

Jak zacząć czytać artykuły, czyli sztuka wyrzucania nauki do kosza

Jak wspomniałem na początku skryptu, większość publikowanych badań nie powinna mieć wpływu na praktykę medyczną i zdaniem niektórych powinna wręcz trafić prosto do kosza. Choć brzmi to brutalnie, dla studenta i lekarza jest to informacja wyzwalająca – nie musisz czytać wszystkiego, co opublikowano. Twoim zadaniem jest nauczyć się odróżniać ziarno od plew, skupiając się nie na tym, jak obiecujące są wyniki, ale na tym, jak solidna była **metodologia**.²³

Większość prac naukowych pisana jest w schemacie **IMRAD** (*Introduction, Methods, Results, and Discussion*). We wstępie autorzy opisują aktualny stan wiedzy, lukę w wiedzy, które ich badanie adresuje oraz główny cel badania. W metodach (kluczowe z punktu widzenia jakości badania) opisane jest szczegółowo pytanie kliniczne, rodzaj badania, liczba uczestników. Wyniki przedstawiają zebrane i przeanalizowane dane, zaś dyskusja umiejscawia wyniki w kontekście dotychczasowej wiedzy medycznej. Pamiętaj: decydując, czy praca jest warta Twojego czasu, oceniaj ją po rozdziale Metody, a nie po spekulacjach w Dyskusji.

Zanim zagłębisz się w szczegółową analizę, zadaj sobie trzy wstępne pytania, które pomogą Ci rozeznaczyć się w tekście:

1. **Dlaczego przeprowadzono to badanie i jakie pytanie kliniczne postawili autorzy?** Dobra praca powinna jasno określić tło problemu i sformułować pytanie kliniczne – najczęściej jest to ostatni akapit lub zdanie wstępu.
2. **Jaki to rodzaj badania?** Teraz już znając rodzaje badań możesz sprawdzić, jakie jest to, które czytasz oraz jakie miejsce zajmuje w hierarchii dowodów. Czy jest to badanie randomizowane, obserwacyjne czy może przegląd systematyczny przeprowadzonych już badań?
3. **Czy projekt badania był odpowiedni do postawionego pytania?** To kluczowy moment – nie każda metoda pasuje do każdego problemu.

Przypomnę, które badania najlepiej odpowiadają na poszczególne pytania:

- **Leczenie:** Tutaj złotym standardem jest RCT. Dzięki losowemu przydziałowi do grup (np. leku i placebo), możemy z dużym prawdopodobieństwem przypuszczać, że różnice w wynikach wynikają z samej interwencji.
- **Rokowanie (prognoza):** Tu najlepiej sprawdzają się badania kohortowe, preferencyjnie prospektywne. Śledzimy w nich grupę ludzi przez lata (czasem dekady!), aby sprawdzić, jak ekspozycja na dany czynnik (np. palenie tytoniu) wpływa na rozwój choroby. Klasycznym przykładem jest badanie 40 000 brytyjskich lekarzy, które dowiodło związku palenia z rakiem płuc (opisane dokładnie pod koniec skryptu).
- **Rzadkie choroby:** W ich przypadku stosujemy badania kliniczno-kontrolne (*case-control studies*), gdzie porównujemy osoby chore

z dopasowaną grupą zdrową, szukając w ich przeszłości czynników sprawczych. Jest to rodzaj badania obserwacyjnego.

- **Opisy przypadków:** Choć stoją nisko w hierarchii, są niezastąpione w szybkim ostrzeganiu przed nowymi zagrożeniami. To właśnie dzięki czujności lekarzy opisujących przypadki dzieci z wadami kończyn, świat dowiedział się o tragicznych skutkach stosowania talidomidu u kobiet w ciąży.^{24,25}

Warto wiedzieć, co dyskwalifikuje badanie w oczach recenzentów prestiżowych czasopism. Najczęstsze przyczyny to: zbyt mała grupa badana, brak odpowiedniej grupy kontrolnej, niewłaściwa analiza statystyczna, niejasny lub niepełny opis metod i analizy czy wyciąganie wniosków, które nie wynikają z zebranych danych.²³

Przeglądy systematyczne i meta-analizy: filtry w świecie informacyjnego szumu

Wyobraź sobie, że stoisz przed biblioteką, w której codziennie pojawia się 113 nowych artykułów na temat samej tylko depresji. Jako lekarz masz obowiązek leczyć zgodnie z „aktualną wiedzą medyczną”, ale przeczytanie wszystkiego jest fizycznie niemożliwe. Co gorsza, wiemy, że nawet 85% opublikowanych badań to prace źle zaprojektowane, błędnie raportowane lub po prostu nieprzydatne w praktyce. Jak w tym gąszczu znaleźć informacje, które realnie pomogą Twojemu pacjentowi? Pomocne są **przeglądy systematyczne** oraz **meta-analizy**.²⁶

Istnieje kilka podejść do pisania prac przeglądowych, czyli takich, które zbierają informacje z opublikowanych już artykułów. Pierwszy polega na selektywnym przeglądaniu baz danych, wybieraniu artykułów czy fragmentów badań, które pasują do tezy autora, który niewygodne fragmenty może chcieć pominąć. Tak powstają **przeglądy narracyjne**. Są one podatne na subiektywność i stronniczość autora,

zwykle nie wskazują kryteriów selekcji literatury, mają ograniczoną powtarzalność oraz mniejszą wiarygodność w porównaniu do przeglądów systematycznych. Przeglądy narracyjne mogą pomijać istotne badania, nie uwzględniać wszystkich dostępnych danych i zazwyczaj nie zawierają oceny jakości dowodów, co ogranicza ich wartość w podejmowaniu decyzji klinicznych.

Pisanie **przeгляdu systematycznego** ma strukturę, której zadaniem jest ograniczenie stroniczości oraz selektywności doboru informacji. Celem autora jest podsumowanie wszystkich dostępnych dowodów na dany temat. To badanie naukowe, w którym „pacjentami” są inne badania oryginalne. Jego cechą charakterystyczną jest **przejrzysta i powtarzalna metodologia**. Autor musi:

- Z góry określić cele i metody szukania informacji.
- Opisać, dlaczego konkretne badania włączył, a inne odrzucił (nie może odrzucić badania tylko dlatego, że ma „niepasujące” wyniki).
- Ocenić jakość metodologiczną każdego badania.

Zanim zmienisz sposób leczenia na podstawie przeglądu, przeprowadź mały test rzetelności, zadając cztery pytania:

4. **Czy pytanie było precyzyjne?** Zamiast ogólnego „Czy leki działają?”, szukamy konkretnego: „Czy lek X w porównaniu z interwencją Y (np. placebo) zmniejsza nasilenie objawów psychiatrycznych u kobiet po porodzie?”. Precyzyjne pytania dają też precyzyjne odpowiedzi, które można uwzględnić w swojej praktyce klinicznej.
5. **Czy autorzy gruntownie przeszukali literaturę?** Dobry badacz nie ogranicza się do jednej bazy danych. Szuka w rejestrach badań klinicznych (Cochrane, clinicaltrials.gov), w literaturze obcojęzycznej, a nawet w tzw. szarej literaturze (*grey literature*) – czyli w nieopublikowanych raportach badań czy pracach doktorskich.

6. **Czy oceniono jakość badań?** Nie wszystkie badania są równe. Autor musi sprawdzić, czy dany artykuł nie ma poważnych wad w projektowaniu, prowadzeniu lub analizie badania, które mogłyby sfalszować wynik.
7. **Analiza wrażliwości (*sensitivity analysis*), czyli „co by było, gdyby?”** To sprawdzanie czy wynik całego przeglądu nie zależy od jednego, małego i słabego badania. Jeśli po usunięciu podejrzanego badania optymistyczne wnioski znikają, to znak, że musisz być bardzo ostrożny.

Gdy autorzy przeglądu systematycznego postanowią połączyć wyniki wielu badań przy użyciu metod statystycznych, otrzymujemy **meta-analizę**. Pozwala ona na zsumowanie mniejszych grup pacjentów w jedną większą populację, co zwiększa precyzję ostatecznego wyniku. Statystycznie przejawia się to jako zawężenie 95% przedziału ufności – więcej na ten temat w rozdziale poświęconym statystyce.

Meta-analiza ułatwia integrację sprzecznych lub niejednoznacznych wyników badań, dostarczając jednoznacznej, syntetycznej odpowiedzi na pytanie kliniczne. Pozwala także na eksplorację różnic pomiędzy badaniami, odpowiadającymi na to samo pytanie. Wreszcie, meta-analiza może przyspieszyć wdrożenie skutecznych terapii i wykrycie szkodliwych interwencji w oczekiwaniu na duże, pojedyncze badania.

Badania jakościowe: Jak zrozumieć to, czego nie da się policzyć?

Do tej pory skupialiśmy się na badaniach, które produkują wyniki w postaci liczb. Istnieje jednak pewna pułapka: jako lekarze często czujemy się bezpieczniej, gdy widzimy cyfry, wierząc, że są one twardymi faktami. Ale czy samo stwierdzenie, że lek X zmniejsza nasilenie objawów psychiatrycznych, mówi nam wszystko o doświadczeniu naszej pacjentki

z pierwszego rozdziału? Nie do końca. Tutaj z pomocą przychodzą **badania jakościowe** (*qualitative research*).²⁷

Podczas gdy badania ilościowe (*quantitative*) pytają: „Ile?“, „Jaki procent?“ lub „Jak często?“, badania jakościowe szukają odpowiedzi na pytania: „**Dlaczego?**“, „**W jaki sposób?**“ oraz „**Co to oznacza dla pacjenta?**“. Jeśli chcemy wiedzieć, dlaczego matki z psychozą poporodową boją się przyjmować leki, nie potrzebujemy ankiety, ale szczerzej rozmowy, która pozwoli nam zrozumieć ich perspektywę.

Badania jakościowe nie są gorszą czy „miękką” wersją nauki. Mają po prostu inną logikę. Zamiast wyciągać wnioski poprzez **dedukcję** (sprawdzanie z góry założonej hipotezy), badacz jakościowy stosuje **rozumowanie indukcyjne** (*inductive reasoning*) – zaczyna od obserwacji i rozmów, na podstawie których poprzez analizę danych tworzy tezy tłumaczące dane zagadnienie. Siłą tych badań nie jest **niezawodność/powtarzalność** (*reliability*), ale **trafność** (*validity*), czyli bliskość prawdy o ludzkim doświadczeniu.

Oto lista pytań, które pomogą Ci ocenić wartość takiego artykułu:

1. **Czy pytanie badawcze jest jasne i ważne klinicznie?** Samo „postanowiliśmy porozmawiać z 20 pacjentami o ich chorobie” to za mało. Dobra praca powinna jasno uzasadnić: „Wiemy, że pacjenci nie biorą leków na padaczkę. Postanowiliśmy więc zbadać ich ukryte przekonania na temat tej choroby, aby zrozumieć przyczyny tego zjawiska”.
2. **Czy podejście jakościowe było właściwe?** Jeśli autor chce sprawdzić częstość występowania skutków ubocznych, powinien pozyskać dane z badań randomizowanych lub przeprowadzić badanie kohortowe. Jeśli jednak chce zinterpretować, jak pacjent radzi sobie z piętnem choroby psychicznej – metoda jakościowa jest idealna.

3. **W jaki sposób dobrano uczestników?** W badaniach ilościowych staramy się mieć porównywalne grupy zapewnione przez losowy przydział (randomizację). W badaniach jakościowych stosujemy **dobór teoretyczny** (*theoretical sampling*). Jeśli badamy doświadczenia porodowe kobiet, badacz powinien celowo szukać osób o skrajnie różnych przeżyciach: po cesarskim cięciu, po porodzie naturalnym czy po poronieniu. Nie szukamy średniej, ale pełnego spektrum ludzkich losów.
4. **Kim jest badacz i jakie ma uprzedzenia?** W EBM dążymy do obiektywizmu, ale w badaniach jakościowych badacz współtworzy dane poprzez wniesienie do procesu badawczego swojej własnej perspektywy. Jego poglądy, zawód (czy jest lekarzem z kliniki, czy pracownikiem socjalnym) oraz perspektywa kulturowa zawsze wpływają na to, jak interpretuje słowa pacjenta. Dobry autor opíše to wprost, abyś mógł ocenić wyniki w odpowiednim kontekście – takie transparentne przedstawienie własnego stanowiska w badaniu jakościowym nazywa się **refleksyjnością** (*reflexivity*).²⁸
5. **Czy analiza danych była rzetelna?** Badacz nie może po prostu wybrać ciekawych cytatów, które pasują do jego opinii. Szukamy dowodów na **systematyczną analizę**, np. analizę tematyczną (*thematic analysis*),²⁹ gdzie dane są kodowane i porównywane. Kluczowe jest też szukanie „przypadków sprzecznych”, czyli wypowiedzi pacjentów, które nie pasują do głównej teorii autora.

Dwa pojęcia, które są przydatne przy czytaniu prac jakościowych,

to:

- **Triangulacja** (*triangulation*):³⁰ To proces zwiększania rzetelności poprzez użycie więcej niż jednej metody (np. wywiady + obserwacja), analizę tych samych danych przez kilku niezależnych badaczy lub zbierania danych na ten sam temat od pacjentów i ich bliskich.

- **Cytaty dosłowne** (*verbatim quotes*): Są dla badania jakościowego tym, czym liczby dla meta-analizy – to surowy dowód. Twierdzenie autora, że „lekarze rzadko rozumieją pacjentów”, jest mało wiarygodne, dopóki nie przeczytasz konkretnego cytatu, który to ilustruje.

Częstym zarzutem jest to, że wyniki badań jakościowych dotyczą tylko małej grupy osób w konkretnym miejscu. Jednak dzięki rzetelnemu doborowi uczestników i głębokiej analizie, wnioski te są często **transferowalne** (*transferable*).³¹ Jeśli zrozumiesz mechanizm lęku matki z psychozą poporodową w Londynie, prawdopodobnie pomoże Ci to lepiej rozmawiać z pacjentką w Warszawie, nawet jeśli szczegóły ich życia się różnią.

Badania jakościowe nie są gorsze, lecz odpowiadają na inne pytania. To systematyczne narzędzie, które pozwala zrozumieć sens i znaczenie, jakie pacjenci przypisują swoim doświadczeniom. W świecie EBM liczby dają nam precyzję, ale to badania jakościowe dają nam zrozumienie.

Badania randomizowane: Jak oddzielić naukę od marketingu?

Jeśli w swojej przyszłej praktyce będziesz wypisywać leki, staniesz się celem zainteresowania przemysłu farmaceutycznego, który inwestuje zawrotne sumy, by wpłynąć na Twoje decyzje. Najskuteczniejszym narzędziem tej maszyny są przedstawiciele medyczni, którzy odwiedzą Cię z teczką pełną dowodów na skuteczność swoich produktów. Przedstawiciele medyczni pracują dla firm, których celem jest promowanie i sprzedaż leków, dlatego trudno oczekiwać pełnego obiektywizmu w przekazywanych przez nich informacjach. Ich komunikaty zwykle akcentują korzyści terapii, a kwestie ograniczeń przedstawiane są bardziej skrótowo. Jak zatem czytać o badaniach leków, by nie dać się uwieść marketingowi?³²

Zanim zdecydujesz się podać lek pacjentowi, powinieneś przejść przez trzy etapy:

1. **Określ cel leczenia:** Czy chcesz pacjenta wyleczyć, zapobiec nawrotowi, ograniczyć niepełnosprawność, czy jedynie złagodzić objawy?
2. **Wybierz leczenie na podstawie twardych dowodów:** Opieraj się na wynikach badań z randomizacją kontrolowanych placebo z odpowiednio dobraną grupą kontrolną.
3. **Ustal punkt docelowy terapii:** Musisz wiedzieć, kiedy przerwać leczenie lub zmienić jego intensywność (np. dążenie do konkretnej wartości ciśnienia rozkurczowego).

Wynik zastępczy lub zastępczy punkt końcowy (*surrogate outcome/endpoint*) jest bardzo ważnym pojęciem pojawiającym się w badaniach przedstawianych przez pracowników firm farmaceutycznych. **Wynik zastępczy** to zmienna, którą łatwo zmierzyć (np. poziom cholesterolu, ciśnienie krwi, obraz RTG), a która ma rzekomo przewidywać rzadki lub odległy wynik kliniczny (np. zawał serca czy zgon). Przemysł farmaceutyczny uwielbia punkty zastępcze, ponieważ pozwalają one na przeprowadzanie krótszych i tańszych badań, które jednak nie dają twardych danych nt. skuteczności leku.

Niestety, punkty zastępcze bywają zwodnicze. Klasycznym przykładem jest **badanie CONCORDE**, które dotyczyło leczenia pacjentów z HIV.³³ Wcześniejsze badania sugerowały, że wczesne podanie zydowudyny spowalnia spadek liczby komórek CD4 (wynik zastępczy). Zakładano, że wyższa liczba CD4 przełoży się na przeżycie. Badanie CONCORDE potwierdziło, że poziom CD4 faktycznie spadał wolniej, ale... wskaźniki przeżycia po trzech latach były identyczne w obu grupach. CD4 okazało się nieważne dla tego, co istotne dla pacjenta – czyli długości życia.

Pamiętaj: idealny punkt zastępczy w badaniach klinicznych powinien być łatwy do zmierzenia i mieć biologiczne uzasadnienie, czyli logicznie wiązać się z mechanizmem choroby oraz efektem leczenia. Kluczowe jest, aby zmiana tego punktu pod wpływem terapii przewidywała rzeczywistą poprawę istotnych klinicznych wyników, takich jak przeżycie lub jakość życia, co musi być potwierdzone w wielu opublikowanych badaniach. Takimi uznanymi punktami końcowymi są: **Hemoglobina glikowana (HbA1c)** w cukrzycy – obniżenie HbA1c przewiduje zmniejszenie ryzyka powikłań cukrzycowych. **Ciśnienie tętnicze** w nadciśnieniu – obniżenie ciśnienia przewiduje zmniejszenie ryzyka udaru i zawału serca. Zawsze pytaj przedstawiciela medycznego nie o to, czy lek poprawia wyniki badań laboratoryjnych, ale czy pacjenci dzięki niemu żyją dłużej i zdrowiej.

Nie musisz unikać przedstawicieli, ale to Ty powinieneś przejąć kontrolę nad rozmową. Proponuję podchodzić sceptycznie do dowodów typu „znany profesor to stosuje” czy do broszur reklamowych. Zamiast tego możesz poprosić o dowód w postaci publikacji z recenzowanych czasopism naukowych. Następnie sam oceń lek, używając akronimu **STEPS**:

- **S (Safety – Bezpieczeństwo)**: Jakie jest prawdopodobieństwo poważnych skutków ubocznych?
- **T (Tolerability – Tolerancja)**: Najlepiej mierzona poprzez porównanie odsetka pacjentów wycofujących się z badania w porównaniu z głównym konkurentem.
- **E (Efficacy – Skuteczność)**: Jak produkt wypada w porównaniu z lekiem, który obecnie uważasz za najlepszy?
- **P (Price – Cena)**: Uwzględnij koszty bezpośrednie i pośrednie.
- **S (Simplicity – Prostota)**: Jak łatwe będzie dla pacjenta zastosowanie interwencji? Chociażby wykonanie iniekcji podskórnej wymaga odpowiedniego instruktażu.

Zawsze sprawdzaj, czy wyniki badań są istotne nie tylko statystycznie, ale i klinicznie (o tym więcej w dalszej części skryptu). Nie daj się przekonać „nowością” produktu – często jest to argument przeciwko zmianie, dopóki bezpieczeństwo leku nie zostanie potwierdzone w dłuższej obserwacji. Twoim zadaniem jest odróżnienie nauki od wypracowanych taktyk marketingu i perswazji.

Jeszcze jedno – pamiętaj proszę, żeby sprawdzić czy badanie, o które oparte są materiały reklamowe firmy, zostały przeprowadzone w odpowiedniej populacji chorych (trafność zewnętrzna). Pamiętam, gdy pewna przedstawicielka farmaceutyczna polecała psychiatrom, pracującym w zamkniętym oddziale psychiatrycznym, lek zmniejszający poziom lęku u pacjentów. Zapomniała dodać, że badanie było prowadzone na zdrowych osobach...

Dlaczego nie możemy ufać wszystkiemu, co opublikowano?

Mogłoby się wydawać, że skoro praca naukowa przeszła przez proces recenzji i została opublikowana w prestiżowym czasopiśmie, to zawarte w niej wnioski są niepodważalne. Nic bardziej mylnego. Jak już wspominaliśmy, ogromna część doniesień medycznych ma ograniczone zastosowanie, a rzetelna **ocena jakości metodologicznej** jest Twoją główną linią obrony przed wprowadzeniem w błąd.³⁴ Weryfikowanie poprawności metodologicznej danej publikacji wymaga jednak dość specjalistycznej wiedzy. Nieraz ocena jakości uzależniona jest od rodzaju badania. Jak wspomniałem wcześniej, elementem przeglądu systematycznego jest krytyczna analiza włączonych do przeglądu prac. Autorzy przeglądu w teorii powinni zająć się tym aspektem za Ciebie, przedstawiając Ci możliwie transparentnie podsumowane wyniki, zwracając uwagę na ograniczenia metodologiczne wynikające z niedociągnięć

poszczególnych badań. Niemniej, warto mieć podstawowe zrozumienie procesu oceny metodologicznej publikacji.

Oto kluczowe powody, dla których warto samodzielnie oceniać jakość każdego artykułu. Część z nich była wspomniana w poprzednich rozdziałach:

1. Identyfikacja błędów systematycznych (bias)

Głównym celem oceny jakości jest sprawdzenie czy autorzy uniknęli **błędu systematycznego**, który mógłby zniekształcić porównania między grupami. Błąd systematyczny (*bias*) to systematyczne odchylenie wyników lub wniosków od prawdy. Może to być błąd w koncepcji i projekcie badania – lub w gromadzeniu, analizie, interpretacji, raportowaniu, publikacji lub przeglądzie danych – prowadzący do wyników lub wniosków, które są systematycznie różne od prawdy.³⁵ Innymi słowy, na etapie tworzenia, prowadzenia lub analizy badania można popełnić błąd, który sprawi, że uzyskane wyniki nie będą odzwierciedlać rzeczywistości klinicznej. Przykłady takich błędów zamieszczono poniżej, choć szerszy katalog błędów systematycznych dostępny jest [tutaj](#).³⁶

- **Błąd selekcji (*selection bias*):** Uczestnicy badań mogą się systematycznie różnić od populacji, którą jesteśmy zainteresowani. Na przykład uczestnikami badania nad szczepionką przeciw grypie mogą być zdrowi młodzi dorośli, podczas gdy osoby, które w praktyce najprawdopodobniej skorzystają z interwencji, mogą być w podeszłym wieku i cierpieć na wiele chorób współistniejących.
- **Błąd pamięciowy (*recall bias*):** systematyczny błąd, który pojawia się, gdy uczestnicy nie pamiętają dokładnie wcześniejszych wydarzeń lub doświadczeń albo pomijają szczegóły: na dokładność i ilość wspomnień mogą wpływać późniejsze wydarzenia i doświadczenia. Błąd pamięciowy to problem w badaniach, które

opierają się na samoocenie i zbieraniu danych z przeszłości, np. w retrospektywnych badaniach kohortowych.

- **Błąd związany z wypadnięciem z badania (*attrition bias*)** – występuje, gdy uczestnicy opuszczają badanie w trakcie jego trwania. Niemal zawsze ma to miejsce w pewnym stopniu. Na utratę uczestników mogą mieć wpływ takie czynniki, jak niezadowolająca skuteczność leczenia lub nietolerowane zdarzenia niepożądane. Kiedy uczestnicy opuszczają badanie, może nie być wiadomo, czy kontynuują oni interwencję, czy ją przerywają; mogą nie być dostępne dane dotyczące wyników dla tych uczestników po tym czasie. Jeśli autorzy po prostu ich zignorują i przeanalizują tylko tych, którzy wytrwali do końca, wynik będzie **sztucznie zawyżony** na korzyść leku.
- **Błąd wykonania (*performance bias*)** – wynika z wiedzy o przydziale interwencji posiadanej przez badacza lub uczestnika. Powoduje to różnice w opiece otrzymywanej przez grupę interwencyjną i kontrolną. Na przykład uczestnicy grupy kontrolnej mogą szukać innych metod leczenia, a badacze/lekarze mogą traktować uczestników w różny sposób w zależności od tego, do której grupy należą. Błąd ten może zawyżać szacowany efekt interwencji, szczególnie w badaniach z subiektywnymi wynikami. Sposobem zapobiegania temu błędowi jest zaślepienie (*blinding*), omówione wcześniej w rozdziale poświęconym randomizowanym badaniom klinicznym.
- **Zakłócanie (*confounding*)**, omówione wcześniej w kontekście badań obserwacyjnych, również jest rodzajem błędu systematycznego.

2. Problem trafności zewnętrznej (Czy to pasuje do Twojego pacjenta?)

Nawet metodologicznie prawidłowe badanie może mieć ograniczone zastosowanie w Twojej praktyce, jeśli grupa badana nie przypomina Twoich pacjentów.

- Badania często wykluczają osoby starsze, z wieloma chorobami lub kobiety w wieku rozrodczym.
- Wyniki uzyskane na 23-letnich zdrowych mężczyznach (częstych ochotnikach w badaniach farmakokinetycznych) nie mają przełożenia na leczenie starszej kobiety z licznymi obciążeniami.
- Warunki laboratoryjne badania (częste kontrole, darmowy sprzęt, dodatkowe infolinie dla pacjentów) rzadko odpowiadają realiom Twojego szpitala, co poddaje w wątpliwość to, czy lek zadziała tak samo.

3. Wiarygodność statystyczna i rozmiar grupy

Niektóre badania są po prostu **za małe**, by wykryć istotny efekt, nawet jeśli on istnieje. Im mniejszy efekt danej interwencji czy ekspozycji, tym większa grupa jest potrzebna do wykrycia tegoż efektu. Będzie to szerzej omawiane w rozdziale poświęconemu statystyce. Niewielkie badania często prowadzą do błędnego wniosku, że dana interwencja nie działa. Tzw. **błąd typu II** odnosi się do sytuacji, w której autorzy badania stwierdzają brak efektu (ze względu na małą liczebność grupy), podczas gdy efekt byłby wykrywalny w większej grupie.

Zapoznanie się z rodzajami badań, podstawowymi rodzajami błędów systematycznych oraz uważność na trafność zewnętrzną badania oraz rozmiar badanej grupy w dużej mierze pozwoli Ci odsiać ziarna od plew. Teraz czas dowiedzieć się, gdzie i w jaki sposób szukać badań naukowych, które odpowiedzą na nasze pytania.

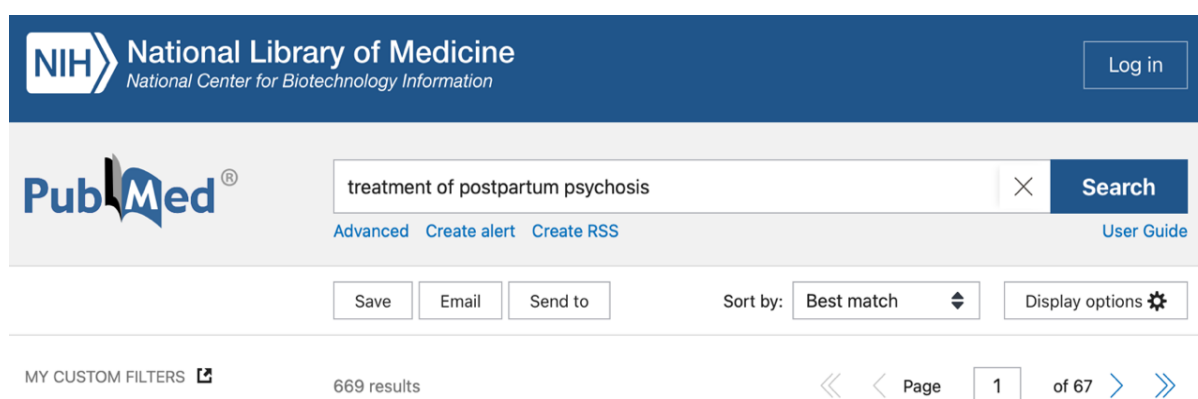
Gdzie znajdę dobrą odpowiedź na moje pytania?

Jak w wielu innych przypadkach – jest wiele dróg do tego samego celu. Lekarze jednak mają mało czasu. Dlatego ważna jest umiejętność szybkiego znajdowania odpowiedzi.

Żeby docenić prostotę rozwiązania, o którym piszę w dalszej części tego rozdziału, najpierw omówmy tradycyjną metodę. Pacjenci czerpią swoją wiedzę z Internetu, ChatuGPT, od znajomych, rodziny. Lekarze jednak powinni odwoływać się do sprawdzonych źródeł informacji, tj. do wysokiej jakości artykułów używających prawidłowego rodzaju badań, o których pisałem poprzednio. Jak je znaleźć?

Opublikowane raporty badań znajdują się w bazach danych, a jedną z najlepiej znanych jest PubMed (<https://pubmed.ncbi.nlm.nih.gov>). Ściślej mówiąc, PubMed jest nie tyle bazą, ile wyszukiwarką danych – warto o niej myśleć jak o Google’u dla naukowców czy lekarzy. Sam w sobie PubMed czerpie z bazy danych MEDLINE, czyli bazy danych Narodowej Biblioteki Medycznej (*National Library of Medicine*) Stanów Zjednoczonych oraz kilku innych źródeł. W ankiecie przeprowadzonej wśród pracowników ochrony zdrowia w Stanach Zjednoczonych pokazano, że PubMed lub bezpośrednio MEDLINE były najczęściej używanymi (ok. 60% respondentów) bazami danych medycznych.³⁷ Omawianie dalszych szczegółów innych medycznych baz danych znacznie wykracza poza zakres tematyczny tego skryptu.

Czy bazy danych są jednak proste w obsłudze i intuicyjne? Po wpisaniu w PubMed „treatment of postpartum psychosis” otrzymujemy ponad 600 wyników (Rycina 6). Wybranie jednego, dwóch czy nawet pięciu najważniejszych artykułów może stanowić wyzwanie. Żeby to zrobić należałoby znaleźć badanie odpowiedniego rodzaju (najlepiej przegląd systematyczny badań randomizowanych lub przynajmniej pojedyncze badanie randomizowane), a później zastanowić się, na ile zewnętrznie trafne oraz metodologicznie prawidłowe jest znaleziony przez nas artykuł. To oczywiście wymaga dużo czasu i skupienia.

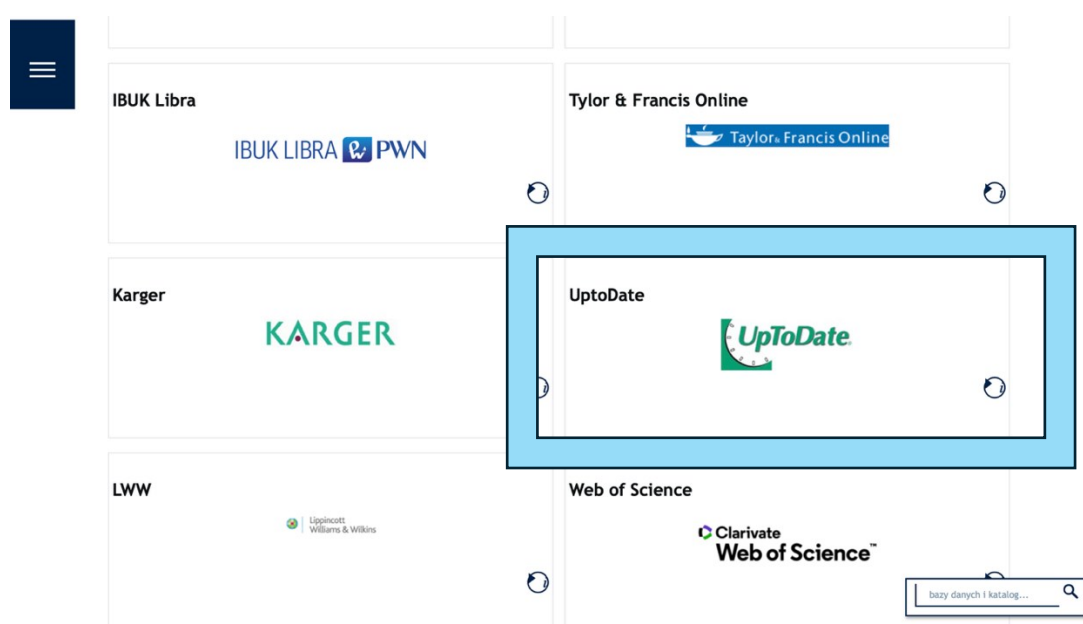


Rycina 6. Wynik wyszukiwania w PubMed (maj 2025)

Korzystanie z UpToDate

Na szczęście mamy alternatywy. Jedną z nich jest strona internetowa UpToDate, która zawiera opracowane przez ekspertów rozdziały dotyczące wielu chorób z różnych gałęzi medycyny. Założeniem tego serwisu jest bycie na bieżąco z aktualną wiedzą medyczną. Autorzy i edytorzy poszczególnych rozdziałów regularnie przeprowadzają przegląd literatury, by upewnić się, że przedstawione w UpToDate informacje są zgodne z aktualną wiedzą medyczną. Kolejną silną stroną tego serwisu jest weryfikacja merytoryczna badań, na które powołują się autorzy poszczególnych rozdziałów. Jak pisałem wcześniej, publikowane badania

medyczne (np. randomizowane badania kliniczne), mimo że stanowią „złoty standard” testowania przyczynowości w medycynie, nadal mogą być różnej jakości. Część z nich może być nadmiernie optymistyczna względem danej interwencji. Wiemy, że badania randomizowane, których jakość jest słaba (tzn. mają kilka lub wiele błędów systematycznych), zawyżają szacowane efekty testowanych interwencji.³⁸ Od długiego czasu wiemy także, że badania z małymi grupami pacjentów nieraz zawyżają skuteczność badanych interwencji, co może być powiązane z ich słabszą jakością metodologiczną.^{39,40} Zaletą UpToDate jest to, że badania, na podstawie których powstają rozdziały, są oceniane przez profesjonalnych edytorów, co daje nam większą pewność, że informacje pozyskane w ten sposób będą bardziej wiarygodne. Jak zyskać dostęp do UpToDate? Każdy student Warszawskiego Uniwersytetu Medycznego może zalogować się do tego serwisu przez stronę internetową naszej biblioteki (<https://biblioteka.wum.edu.pl/zasoby>). W zasobach należy wybrać UpToDate (Rycina 7) i zalogować się swoimi danymi.

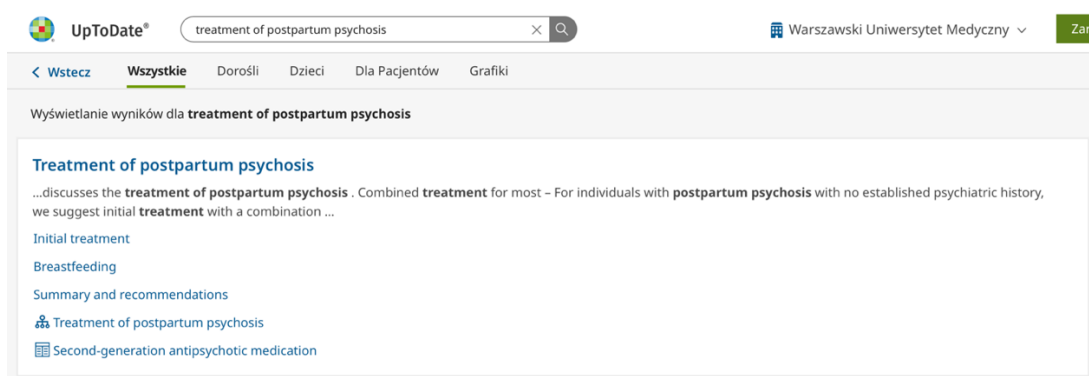


Rycina 7. Widok ze strony internetowej biblioteki uczelnianej, z której można uzyskać dostęp do UpToDate

W tym miejscu należy zaznaczyć, że czytanie literatury medycznej po angielsku może być wyzwaniem! Żeby sobie pomóc, proponuję używać tłumacza opartego o sztuczną inteligencję, czyli DeepL (<https://www.deepl.com/en/translator>). W ten sposób można jakiegokolwiek informacje szybko przetłumaczyć na język polski i nie frustrować się tym, że pewnych rzeczy nie rozumiemy.

Skuteczność leczenia naszej pacjentki – czego się dowiadujemy?

Wracamy do szukania informacji i używamy w UpToDate tego samego hasła, którego użyliśmy w PubMedzie: „treatment of postpartum psychosis”. Tym razem znajdujemy rozdział poświęcony dokładnie temu, czego szukamy (Rycina 8).



Rycina 8. Wynik wyszukiwania w UpToDate (maj 2025)

Przypomnę pytanie, które zadaliśmy w kontekście leczenia:

- **(P)**opulacja – kobiety w wieku rozrodczym z psychozą poporodową
- **(I)**nterwencja – lek (nie wiemy jeszcze jaki)
- **(C)**omparison, tj. porównanie – brak leku lub placebo (tj. interwencja porównawcza)
- **(O)**utcome, tj. wyniki leczenia – ustąpienie psychozy lub zmniejszenie dolegliwości psychotycznych; szkody leczenia

Teraz wiemy również, że najlepiej byłoby znaleźć odpowiedź w formie przeglądu systematycznego randomizowanych badań klinicznych z grupą kontrolną lub chociaż jednego randomizowanego badania klinicznego (na podstawie tabeli dowodów OCEBM). Niemniej, wiemy również, że jeśli nie znajdziemy tego rodzaju dowodu, przejdziemy do kolejnego „pudełka”, czyli będziemy chcieli chociaż dane z badań obserwacyjnych czy opisów przypadków (jedynie w przypadku braku dowodów z badań randomizowanych!).

W przypadku naszej pacjentki mamy do czynienia z poważnym stanem nagłym, zagrażającym potencjalnie życiu zarówno jej, jak i jej dziecka. Może się zatem okazać, że nie znajdziemy badań z randomizacją, gdyż ich przeprowadzenie (tj. porównanie leków z placebo) mogłoby być zwyczajnie zbyt niebezpieczne dla osób w grupie placebo. Pamiętaj proszę, że jeżeli nie znajdziesz randomizowanego badania klinicznego dot. skuteczności leczenia, nadal możesz szukać innego rodzaju badań – zgodnie z zasadą hierarchii dowodów, przedstawionych w rozdziale „Rodzaje odpowiedzi”. Zobaczmy, do czego odwołują się autorzy rozdziału z UpToDate (Rycina 9).

Initiating medication

Combined treatment for most — Our preferred pharmacologic treatment for individuals with postpartum psychosis who do not have an established psychiatric history is with a combination of an antipsychotic and [lithium](#) [4-6]. Our rationale is that individuals with postpartum psychosis have a high likelihood of subsequent diagnosis of bipolar disorder. This differs from our preferred treatment for individuals with psychosis in general (ie, nonpostpartum psychosis) in whom we typically do not begin lithium unless there are clear signs of mood dysregulation (eg, hypomania, mania). Postpartum treatment of individuals with an established psychiatric history is described elsewhere ([algorithm 1](#)).

Rycina 9. Zalecane leczenie psychozy poporodowej wg. UpToDate

Widzimy, że rekomendowanym leczeniem jest połączenie leku przeciwpsychotycznego (*antipsychotic*), inaczej neuroleptyku z litem, czyli lekiem z grupy stabilizatorów nastroju (*mood stabilizers*). Na ten moment nie skupiamy się na dokładnych mechanizmach działania tych

leków, na to jeszcze przyjdzie czas. Teraz skupmy się na tym, jakie dowody stoją za doborem leczenia (Rycina 10).

Evidence basis for treatment of postpartum psychosis

- **Mood stabilizers/lithium** – Our preference for using [lithium](#) in postpartum psychosis is based on clinical trials demonstrating efficacy of mood stabilizers in the treatment of bipolar disorder in nonpostpartum females, reports of successful treatment with lithium in postpartum psychosis, and our clinical experience [5,9,10].

In a retrospective report including 64 in-patients who received stepped care for postpartum psychosis, 47 of the 48 individuals who did not respond to acute treatment with benzodiazepines and antipsychotics remitted following the addition of [lithium](#) [5]. Of those that continued lithium monotherapy or augmentation for nine months, 83 percent showed sustained remission.

- **Antipsychotics** – Our preference for using antipsychotic medications in the treatment of females with postpartum psychosis is based on published reports that suggest their use is common and effective in treating the syndrome, and our clinical experience [3,10,11]. Evidence is limited to small uncontrolled studies or case reports [5,10]. Indirect evidence includes randomized trials of medications for the underlying psychotic disorders in nonpostpartum samples [12-15]. There are no known differences among antipsychotics in their efficacy for reducing psychosis. Efficacy of antipsychotics in the treatment of psychosis is discussed elsewhere. (See "[Psychosis in adults: Initial management](#)" and "[Schizophrenia in adults: Maintenance therapy and side effect management](#)".)

Rycina 10. Podstawa dowodowa leczenia psychozy poporodowej wg. UpToDate

Autorzy piszą, że preferencja co do litu opiera się na randomizowanych badaniach klinicznych dot. skuteczności leczenia choroby afektywnej dwubiegunowej u kobiet, jednak nie w okresie poporodowym. Odnoszą się do tej grupy pacjentów ze względu na to, że psychoza poporodowa wiąże się z ryzykiem wystąpienia choroby afektywnej dwubiegunowej w przyszłości (pamiętasz nasze drugie pytanie dotyczące prognozy?). Dane dot. skuteczności leczenia psychozy poporodowej u kobiet pochodzą z opisów przypadków i doświadczenia klinicznego autorów artykułu.

Jeśli zaś chodzi o leki przeciwpsychotyczne baza dowodów ma się podobnie – opisy przypadków oraz doświadczenie autorów, jeśli chodzi o populację kobiet z psychozą poporodową. Autorzy piszą, że dowody są ograniczone do „małych, niekontrolowanych badań i opisów przypadków”. Piszą też dalej, że „Dowody pośrednie obejmują randomizowane badania kliniczne leków stosowanych w leczeniu zaburzeń psychotycznych u pacjentów niebędących w okresie poporodowym. Nie stwierdzono

różnic między lekami przeciwpsychotycznymi pod względem skuteczności w zmniejszaniu objawów psychozy.”

Rozpatrzmy przedstawione dowody pod dwoma względami (Tabela 1) – ich siły (rodzaj badania) oraz trafności zewnętrznej (czy cytowane dowody odnoszą się bezpośrednio do naszej pacjentki?).

Tabela 2. Siła dowodu i trafność zewnętrzna dowodów na skuteczność leków na psychozę poporodową

Interwencja	Przedstawiony dowód	Siła dowodu	Trafność zewnętrzna
Lit	Badania randomizowane skuteczności litu u kobiet (nie po porodzie) z chorobą afektywną dwubiegunową	Wysoka (randomizowane badanie kliniczne)	Ograniczona – nasza pacjentka nie jest „jakakolwiek” kobietą. Wystąpienie psychozy po porodzie może mieć wpływ na leczenie i prognozę.
	Serie przypadków i doświadczenie kliniczne autorów	Niska (brak grupy kontrolnej, rygorystycznej metodologii)	Wysoka – przedstawione dowody odnoszą się bezpośrednio do naszej pacjentki
Leki p-psychotyczne	Serie przypadków i doświadczenie kliniczne autorów	Niska (brak grupy kontrolnej, rygorystycznej metodologii)	Wysoka – przedstawione dowody odnoszą się bezpośrednio do naszej pacjentki
	Badania randomizowane potwierdzające skuteczność leków w leczeniu psychoz niezwiązanych z porodem	Wysoka (randomizowane badanie kliniczne, a nawet przeglądy systematyczne takich badań)	Ograniczona – nasza pacjentka nie jest „jakakolwiek” kobietą. Wystąpienie psychozy po porodzie może mieć wpływ na leczenie i prognozę.

W idealnej sytuacji chcielibyśmy uzyskać dowody z przeglądów systematycznych randomizowanych badań klinicznych na temat skuteczności leków w danej jednostce chorobowej. Rzeczywistość bywa jednak inna – musimy podejmować decyzje na podstawie danych, które mamy. Niepewność jest wpisana w praktykowanie medycyny, a EBM

jej nie likwiduje – jedynie ją ogranicza. Autorzy w skrócie stwierdzili, że z opisów przypadków i ich własnego doświadczenia klinicznego przedstawione leki są skuteczne w leczeniu psychozy poporodowej **wśród pacjentek podobnych do naszej**. Ponadto zaznaczyli, że istnieją wysokiej jakości dowody na to, że te leki są skuteczne **w chorobach podobnych**, ale nieidentycznych do tej, której doświadcza nasza pacjentka. W swojej syntezie są **transparentni** i przedstawiają **możliwie najlepsze dostępne dowody**. Nie zatrzymują się jednak na podsumowaniu, ale podają też twarde dane, które mogą pomóc nam zrozumieć **jak bardzo** leczenie może być skuteczne: „W retrospektywnym raporcie obejmującym 64 pacjentki hospitalizowane, które otrzymały stopniową opiekę w związku z psychozą poporodową, 47 z 48 osób, które nie zareagowały na ostre leczenie benzodiazepinami i lekami przeciwpsychotycznymi, osiągnęło remisję po dodaniu litu. Spośród osób, które kontynuowały monoterapię litem lub leczenie wspomagające przez dziewięć miesięcy, 83 procent wykazało trwałą remisję.”

Działania niepożądane terapii – jak znaleźć dowody?

Wiemy już, że nie ma randomizowanych badań klinicznych, z których moglibyśmy czerpać informacje na temat skuteczności czy szkód leczenia naszej pacjentki. Dowiedzieliśmy się, że po integracji najlepszych możliwych dowodów z doświadczeniem klinicznym, rekomendowane jest używanie litu i leków przeciwpsychotycznych. Powstaje jednak pytanie – który lek powinniśmy wybrać? Autorzy sami wskazali, że nie ma ewidentnych różnic co do skuteczności działania tych leków. Mogę jednak dodać od siebie, że leki te różnią się, jeśli chodzi o profil potencjalnych działań niepożądanych, których jest całkiem sporo. Skupmy się jednak na jednym z nich, który może być istotny dla naszej pacjentki, a następnie spróbujmy znaleźć dowody na temat tego, **jak dużej** szkody należy się

spodziewać w zależności od leku, którego użyjemy. Działaniem niepożądanym, które rozważymy, będzie hiperprolaktynemia, czyli wzrost stężenie prolaktyny we krwi.

Prolaktyna w okresie poporodowym powinna podlegać fizjologicznej regulacji. Jej nieprawidłowo podwyższone stężenie może być powiązane z opóźnionym powrotem owulacji, nasileniem laktacji, pojawieniem się objawów depresyjnych, obniżeniem libido, a także potencjalnie zwiększeniem ryzyka zachorowania na raka piersi w przypadku wieloletniego przyjmowania leku, zwiększającego stężenie prolaktyny.⁴¹ Wkraczamy zatem poniekąd w terytorium wartości i preferencji pacjentki – warto byłoby dobrać taki lek, który pacjentka będzie dobrze tolerowała i nie odstawi go po wyjściu ze szpitala. Oczywiście należałoby przedyskutować z nią wiele potencjalnych działań niepożądanych leczenia, jednak tutaj ograniczymy się do jednego, żeby nie wprowadzać zbyt dużo informacji naraz.

Szukanie w UpToDate

Co autorzy UpToDate doradzają nam w zakresie doboru leku pod względem jego działań niepożądanych? Sugerują, żebyśmy wiedzę na ten temat oparli o dane zebrane z dziesiątek randomizowanych badań klinicznych pacjentów ze schizofrenią, wśród których te leki były używane. Po przejściu do sekcji „*Schizophrenia in adults: Maintenance therapy and side effect management*” („Schizofrenia u dorosłych: terapia podtrzymująca i zarządzanie skutkami ubocznymi”) oraz do podrozdziału „Skutki uboczne endokrynologiczne i metaboliczne” widzimy następującą tabelę (Rycina 11).

Selected adverse effects of antipsychotic medications for schizophrenia^[1-3]

	Weight gain	Glucose abnormalities	Hyperlipidemia	Akathisia	Parkinsonism	Dystonia	Tardive dyskinesia	Prolactin elevation	Sedation	Anticholinergic	Orthostatic hypotension
Second-generation agents											
Aripiprazole	+	+	+	++	+	+	+	0		+	+
Asenapine	++	++	++	++	+	++	++	++		+	+ / ++
Brexpiprazole [§]	+	+	++	++	+	+	+	0		+	+
Cariprazine [¶]	++	+	+	++	+	+	+	+		++	+
Clozapine ^Δ	+++	+++	+++	+	+	+	+	+	+	+++	+++
Iloperidone	++	++	+	+	+	+	+	++		+	+++
Lumateperone [¶]	+	+	+	+	+	+	+	0		+	+
Lurasidone	+	++	++	++	++	++	++	++		+	+
Olanzapine	+++	+++	+++	++	++	+	+	++	+	++	++
Paliperidone	++	+	++	++	++	++	++	+++		+	++
Pimavanserin	-	+	+	+	+	+	+	+		+	++
Quetiapine	++	++	+++	+	+	+	+	+	+	++	++
Risperidone	++	++	+	+++	+++	++	++	+++		+	++
Ziprasidone	+	+	+	++	+	+	+	++		+	++

Rycina 11. Tabela działań niepożądanych leków przeciwpsychotycznych na podstawie UpToDate

Na podstawie tej tabeli widzimy, że najmniejszy wpływ na prolaktynę miałyby aripiprazol, brexpiprazol oraz lumateperon (ten ostatni niedostępny w Polsce). Zawęża nam to pulę do dwóch potencjalnych kandydatów. Autorzy tabeli powołują się między innymi na bardzo ciekawe badanie – przegląd systematyczny badań randomizowanych z meta-analizą, w której badany był wpływ poszczególnych leków na samą tylko prolaktynę.⁴² Mamy zatem najwyższej jakości dowód, do jakiego moglibyśmy się odnieść, zadając pytanie o potencjalne szkody leczenia. Nie zawsze jednak będziemy mieć dostęp do UpToDate. Pytanie brzmi – jak wówczas poradzić sobie ze zidentyfikowaniem badania, odpowiadającego na istotne pytanie kliniczne? Z odpowiedzią przychodzi nam PubMed.

Szukanie w PubMedzie

Tutaj sprawa jest bardziej skomplikowana – musimy trafnie zidentyfikować ważny artykuł. Żeby to zrobić, najpierw przypomnijmy sobie nasze pytanie

w formacie PICO. Skupmy się na wyniku, jakim jest podwyższony poziom prolaktyny po włączeniu leku przeciwpsychotycznego.

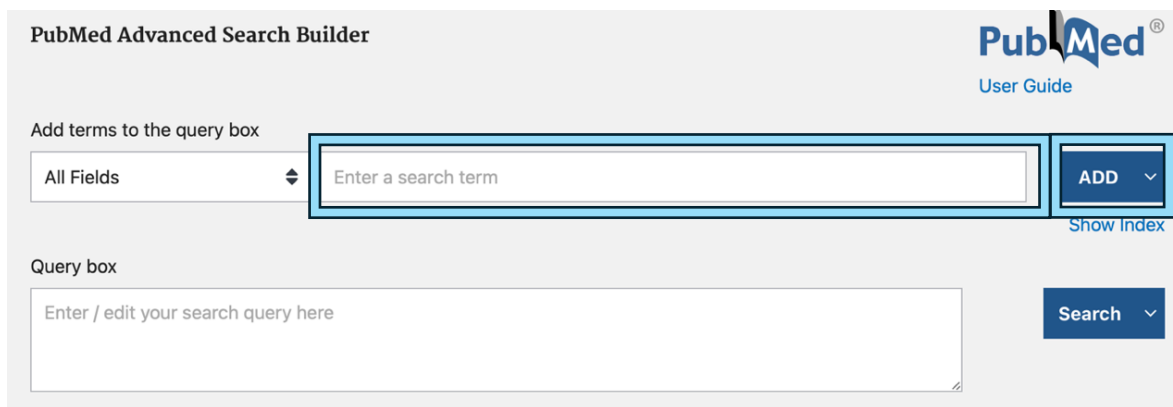
- **(P)**opulacja – kobiety w wieku rozrodczym z psychozą poporodową
- **(I)**nterwencja – leki przeciwpsychotyczne
- **(C)**omparison – placebo lub inne leki przeciwpsychotyczne
- **(O)**utcome – wzrost stężenie prolaktyny

Do tego dodajmy **(S)tudy design**, czyli rodzaj badania, który najlepiej odpowiedziałby na nasze pytanie – w tym wypadku przegląd systematyczny z meta-analizą. Mając wszystkie elementy naszego PICO(S), możemy udać się do PubMedu i rozpocząć poszukiwania!



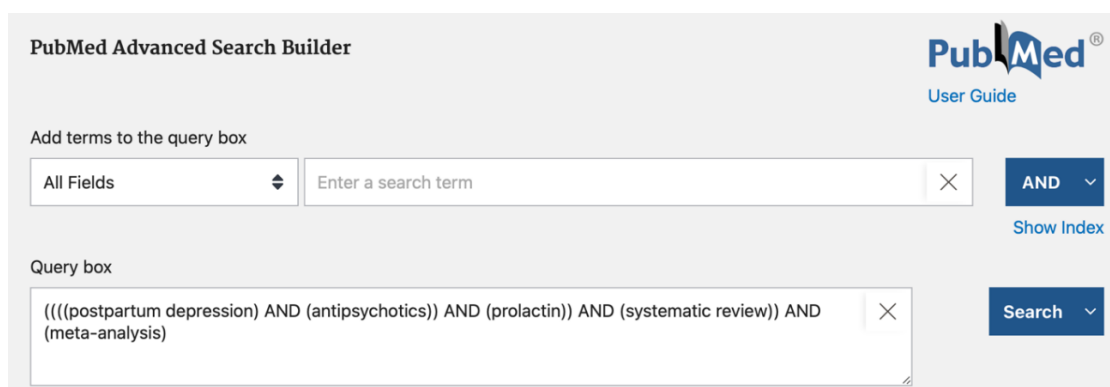
Rycina 12. Ekran startowy PubMed

Tak wygląda ekran startowy. Skorzystajmy z funkcji advanced (Rycina 12). Teraz w polu „Enter search term” wpiszmy elementy naszego PICO(S), łącząc je spójnikiem „AND” po prawej stronie (Rycina 13). W tym celu użyjemy słów kluczowych.



Rycina 13. Zaawansowane opcje wyszukiwania PubMed

Dla (P) – „postpartum psychosis”, dla (I) – „antipsychotics” (leki p-psychotyczne), dla (C) sugeruję nie wpisywać nic (to nie pomoże w szukaniu, a może ograniczyć nasze wyniki), dla (O) – wpiszemy „prolactin”, czyli parametr, którym jesteśmy zainteresowani. W końcu dla (S) wpiszemy dwa słowa kluczowe – „systematic review” oraz „meta-analysis”. Powinno to wyglądać jak na Rycinie 14.



Rycina 14. Komenda do wyszukiwania badania

Niestety, kiedy klikamy „Search” nie znajdujemy żadnych badań. Dlaczego? Jak dowiedzieliśmy się wcześniej, nie ma badań randomizowanych, które były prowadzone wśród kobiet z psychozą poporodową. W takiej sytuacji możemy zrezygnować z tego elementu naszej komendy (*postpartum psychosis*) i ponownie wyszukujemy.

Tym razem wyświetla się lista artykułów naukowych, spośród których należy wybrać to, czego szukamy. Najbardziej trafnym wynikiem wydaje się wynik numer 4 (Rycina 15).

10 years

Custom Range

TEXT AVAILABILITY

Abstract

Free full text

Full text

ARTICLE ATTRIBUTE

Associated data

ARTICLE TYPE

Books and Documents

Clinical Trial

Meta-Analysis

Randomized Controlled Trial

Review

Systematic Review

[See all article type filters](#)

Additional filters +

Clear applied filters

Reset filters menu

Antipsychotic therapy and breast cancer risk: a comprehensive **systematic review and meta-analysis.**

1

Cite Revelou MT, Filippatos C, Malandrakis P, Terpos E, Zagouri F, Gavriatopoulou M, Ntanasis-Stathopoulos I.

Breast Cancer. 2026 Jan;33(1):38-49. doi: 10.1007/s12282-025-01804-5. Epub 2025 Nov 27. PMID: 41307631

Primary analysis involved an overall hazard ratio **meta-analysis**, with secondary odds ratio analysis for complementary evidence. Further subgroup and meta-regression analyses were conducted. ...Furthermore, **prolactin** (PRL)-increasing **antipsychotics** pose ...

Correction: Antipsychotic-Related Prolactin Changes: A Systematic Review and Dose-Response Meta-analysis.

2

Cite Lin X, Sifias S, Tian J, Wu H, Qin M, Correll CU, Schneider-Thoma J, Leucht S.

CNS Drugs. 2026 Feb;40(2):271-273. doi: 10.1007/s40263-025-01243-y. PMID: 41217737 **Free PMC article.** No abstract available.

Sexual dysfunctions related to use of antipsychotics: A protocol for a systematic review and meta-analysis.

3

Cite Herder T, Spoelstra SK, Niemeijer AS, Knegtering H.

PLoS One. 2025 Aug 21;20(8):e0329559. doi: 10.1371/journal.pone.0329559. eCollection 2025. PMID: 40839626 **Free PMC article.**

INTRODUCTION: Sexual dysfunctions are a challenging side effect associated with **antipsychotic** treatment. This protocol outlines a **systematic review** and **meta-analysis** to assess the prevalence of overall sexual dysfunction, as well as the specific ...

Antipsychotic-Related Prolactin Changes: A Systematic Review and Dose-Response Meta-analysis.

4

Cite Lin X, Sifias S, Tian J, Wu H, Qin M, Correll CU, Schneider-Thoma J, Leucht S.

CNS Drugs. 2025 Oct;39(10):937-947. doi: 10.1007/s40263-025-01218-z. Epub 2025 Aug 20. PMID: 40830715 **Free PMC article.**

BACKGROUND: **Prolactin** increase is a common and potentially problematic adverse event of **antipsychotics**. We aimed to discover the relationship between **antipsychotic** doses and changes in **prolactin** levels. OBJECTIVE: To examine the relationship between ...

Rycina 15. Wyniki wyszukiwania w PubMed

Po wejściu w link wyświetla nam się tytuł artykułu wraz z linkiem do pełnego tekstu po prawej stronie – w ramce (Rycina 16).

Meta-Analysis > CNS Drugs. 2025 Oct;39(10):937-947. doi: 10.1007/s40263-025-01218-z.
Epub 2025 Aug 20.

Antipsychotic-Related Prolactin Changes: A Systematic Review and Dose-Response Meta-analysis

Xiao Lin ^{1 2}, Spyridon Sifis ^{1 2}, Jing Tian ^{1 2}, Hui Wu ^{1 2}, Mengchang Qin ^{1 2},
Christoph U Correll ^{3 4 5 6}, Johannes Schneider-Thoma ^{# 1 2}, Stefan Leucht ^{# 7 8}

Affiliations + expand

PMID: 40830715 PMCID: PMC12423237 DOI: 10.1007/s40263-025-01218-z

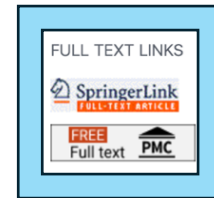
Erratum in

[Correction: Antipsychotic-Related Prolactin Changes: A Systematic Review and Dose-Response Meta-analysis.](#)

Lin X, Sifis S, Tian J, Wu H, Qin M, Correll CU, Schneider-Thoma J, Leucht S.

CNS Drugs. 2026 Feb;40(2):271-273. doi: 10.1007/s40263-025-01243-y.

PMID: 41217737 [Free PMC article.](#) No abstract available.



ACTIONS

“ Cite

📖 Collections

↪ Permalink

PAGE NAVIGATION

< Title & authors

Erratum in

Abstract

Rycina 16. Widok artykułu w PubMed

Poniżej znajduje się abstrakt, czyli streszczenie artykułu – składa się ze wstępu, celu, metod, wyników oraz konkluzji (Rycina 17). Pełen tekst ma analogiczne sekcje, choć zazwyczaj cel zawiera się we wstępie, a między wynikami a konkluzjami mamy jeszcze dyskusję, czyli omówienie wyników w kontekście dostępnej literatury. W metodach czytamy, że artykuł jest przeglądem systematycznym, w którym zbierano dane z randomizowanych badań klinicznych na temat wpływu leków przeciwpsychotycznych na poziom prolaktyny wśród pacjentów z zaostrzeniem psychozy w przebiegu schizofrenii. Dane były systematycznie zebrane i poddane meta-analizie (czyli obróbce statystycznej). Artykuł ten niemal idealnie wpasowuje się w zadane przez nas pytanie – niestety nie zawęża się do grupy kobiet z psychozą poporodową, zatem nie jest idealnie zewnętrznie trafny. Niemniej, możemy założyć, że biologiczny efekt leków przeciwpsychotycznych na mózg (w tym wpływ na produkcję prolaktyny) może być podobny niezależnie od przyczyny psychozy. Jest to ten sam artykuł, na który powołują się autorzy artykułu w UpToDate.

Abstract

Background: Prolactin increase is a common and potentially problematic adverse event of antipsychotics. We aimed to discover the relationship between antipsychotic doses and changes in prolactin levels.

Objective: To examine the relationship between antipsychotic doses and changes in prolactin levels in adults with acutely exacerbated schizophrenia.

Methods: We searched the Cochrane Schizophrenia Group register (last search 26 July 2024) and previous reviews for fixed-dose, randomized controlled trials (RCTs) that investigated monotherapy of 21 antipsychotics in adults with acutely exacerbated schizophrenia. The primary outcome was mean prolactin change from baseline to study endpoint adopting mean differences (MD) in ng/mL as the effect size measure. The dose-response curves were estimated by conducting random-effects dose-response meta-analyses using the restricted cubic spline method.

Results: Among 165 eligible studies, 68 studies with 238 dose arms (23,128 participants, 35% female) reported on prolactin and were meta-analyzed. Of these, 94% lasted \leq 3 months, and 90% of the studies used oral formulations. Participants in one study experienced their first episode, while all other studies also included multiepisode participants. The dose-response curves indicated that with aripiprazole, higher doses were significantly associated with lower prolactin levels than lower doses. Brexpiprazole, cariprazine, lumateperone, and quetiapine carried negligible risks for prolactin increase across examined doses. During treatment with most other antipsychotics, i.e., asenapine, haloperidol, iloperidone, lurasidone, olanzapine, paliperidone, risperidone, and ziprasidone, prolactin levels rose with increasing doses and then continued to increase or plateaued. The shape of the dose-response curves was similar in males and females, with generally larger amplitudes of the curves in females.

Conclusions: The prolactin-increasing property varies among antipsychotics, is dose-related, and is greater in females. These findings in adults with acutely exacerbated schizophrenia can help clinicians titrate and adapt antipsychotic doses and consider patients' sex in treatment decisions. The protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO); registration no. CRD42020181467.

Rycina 17. Abstrakt artykułu

Po otworzeniu pełnego tekstu, z wyników oraz figury 2. dowiadujemy się, że autorzy dokonali szczegółowej analizy wpływu każdego leku (w zależności od dawki) na stężenie prolaktyny. Rozważmy w tym kontekście dwa leki, żeby zrozumieć, co oznaczają „+” lub „0” w tabeli autorów UpToDate (Rycina 11). Na podstawie danych z tej publikacji możemy wyczytać, że w porównaniu z placebo, wpływ na stężenie prolaktyny wśród kobiet jest następujący dla dwóch wybranych przeze mnie leków w porównywalnych dawkach:

- Aripiprazol: -7.5 ng/mL (95% CI -40 ng/mL do 25 ng/mL)
- Risperidon: 34 ng/mL (95% CI 13 ng/mL do 55 ng/mL).

Na razie to może wyglądać przerażająco, ale zrobmy tu pauzę. W rozdziale poświęconym statystyce omówię, co oznaczają te liczby oraz w jaki sposób można z nich wywnioskować, że aripiprazol nie ma istotnego wpływu na stężenie prolaktyny, podczas gdy rysperidon powoduje wzrost stężenia prolaktyny u kobiet.

Po wstępnym omówieniu danych, dotyczących skuteczności leczenia oraz jego działań niepożądanych, przejdźmy do drugiego pytania, które zadaliśmy w kontekście naszej pacjentki.

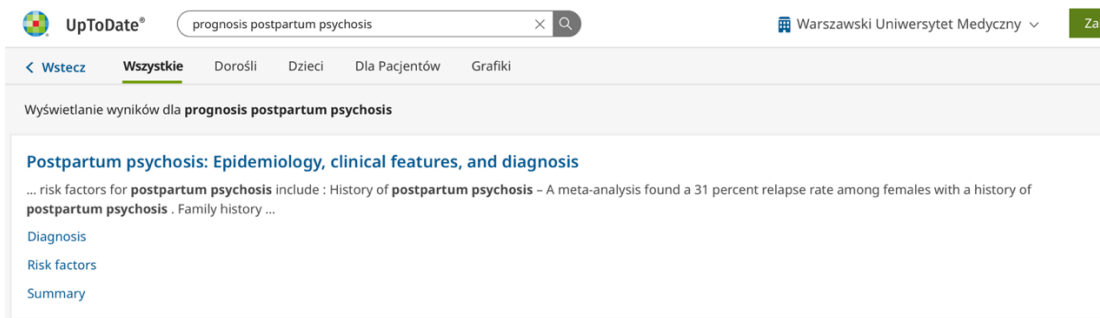
Jak znaleźć informacje nt. prognozy?

Przypomnę drugie pytanie, które zadaliśmy w formacie PECO, dotyczące prognozy.

- **(P)**opulacja – kobiety w wieku rozrodczym
- **(E)**xposition, tj. ekspozycja – zachorowanie na psychozę poporodową
- **(C)**omparison, tj. porównanie – brak zachorowania na psychozę poporodową
- **(O)**utcome, tj. wynik kliniczny – zachorowanie w przyszłości na przewlekłą chorobą psychiczną (np. schizofrenię lub chorobę afektywną dwubiegunową)

Rozważamy zatem, czy zachorowanie na psychozę poporodową wiąże się z zachorowaniem na przewlekłą chorobę psychiczną, taką jak choroba afektywna dwubiegunowa czy schizofrenia. Jest to istotne pytanie, gdyż w przypadku nawrotu dolegliwości rodzina, bliscy oraz sama pacjentka powinni być przygotowani i gotowi adekwatnie zareagować, by możliwie jak najszybciej pomóc chorej.

Ponownie udajemy się do UpToDate'a w poszukiwaniu danych. Tym razem szukamy informacji na temat prognozy, wpisując w wyszukiwarce „*prognosis postpartum psychosis*” (Rycina 18).



Rycina 18. Wyszukiwanie danych dot. prognozy w UpToDate

Identyfikujemy odpowiedni artykuł, a następnie w nim znajdujemy podrozdział „*Course*”, czyli przebieg choroby (Rycina 19).

COURSE

Postpartum psychosis most commonly presents within two weeks of childbirth. Episodes of postpartum psychosis can be severe and prolonged. Postpartum psychosis may interfere with maternal-infant bonding, which is also disrupted by inpatient hospitalization of the mother.

Postpartum psychosis has been strongly and consistently associated with bipolar disorder. Its onset in the immediate postpartum period has been shown to significantly predict a subsequent diagnosis of bipolar disorder [33]. Postpartum psychosis can also occur in individuals with a major depression with psychosis, schizophrenia, or schizoaffective disorder, as discussed elsewhere (see 'Risk factors' above). However, 50 percent or more of women with postpartum psychosis have no prior psychiatric history [18,19].

Women experiencing first-episode postpartum psychosis appear to have a high risk of recurrence outside of the postpartum time-period. A systematic review and meta-analysis analyzed the outcomes of 645 patients with a first psychosis episode postpartum with follow-up periods of 11 to 26 years [29]. Three hundred sixty-six (weighted estimate 56.5 percent) experienced one or more subsequent episodes outside of the postpartum period. Forty-six women (weighted estimate 6.1 percent) experienced one or more subsequent episodes exclusively limited to the postpartum time period. Two hundred seventy-nine women (weighted estimate 43.5 percent) did not experience subsequent episodes and were classified as having “isolated postpartum psychosis”.

Rycina 19. Informacje nt. przebiegu psychozy poporodowej (UpToDate)

Dowiadujemy się dwóch bardzo istotnych informacji. Pierwsza jest taka, że wystąpienie psychozy poporodowej jest powiązane ze zwiększonym ryzykiem rozpoznania w przyszłości choroby afektywnej dwubiegunowej. Druga informacja pochodzi z przeglądu systematycznego badań obserwacyjnych z meta-analizą.⁴³ Wynika z niej, że w okresach obserwacyjnych, trwających między 11 a 26 lat, aż 56.5% kobiet miało co najmniej jeden dodatkowy epizod psychotyczny poza okresem połogowym. Ta informacja jest bardzo ważna – oznacza, że nawet

po wielu latach może dojść do wystąpienia kolejnego epizody psychotycznego. Z drugiej strony, nie dotyczy to wszystkich pacjentek. Około 44% nie doświadczyło już żadnego epizodu chorobowego w opisywanym wcześniej wieloletnim okresie obserwacyjnym. Tę informację w odpowiedni, empatyczny sposób należałoby przekazać rodzinie wraz z zaleceniami dot. postępowania w przypadku nawrotu. Oczywiście pojawia się kolejne pytanie – jak zmniejszyć prawdopodobieństwo nawrotu? Zachęcam do sformułowania pytania w postaci PICO oraz poszukiwania odpowiedzi. W tym momencie to zagadnienie nie będzie przeze mnie omawiane.

Skupmy się na pierwszej informacji. Psychoza poporodowa jest powiązana z większym ryzykiem rozwoju choroby afektywnej dwubiegunowej. **Jak bardzo?** Żeby odpowiedzieć na to pytanie, musimy wczytać się w tekst artykułu, który cytują autorzy UpToDate'a. Przy każdej informacji znajduje się referencja w kwadratowym nawiasie, co widać na Figurze 18. W tym przypadku interesuje nas numer 33 – link przekierowuje nas na stronę PubMedu, którego strukturę już poznaliśmy (Rycina 20).

> Arch Gen Psychiatry. 2012 Apr;69(4):428-34. doi: 10.1001/archgenpsychiatry.2011.157.
Epub 2011 Dec 5.

Psychiatric disorders with postpartum onset: possible early manifestations of bipolar affective disorders

Trine Munk-Olsen ¹, Thomas Munk Laursen, Samantha Meltzer-Brody, Preben Bo Mortensen,
Ian Jones

Affiliations + expand

PMID: 22147807 DOI: 10.1001/archgenpsychiatry.2011.157

FULL TEXT LINKS

FULL TEXT
JAMA Psychiatry

ACTIONS

“ Cite

📖 Collections

SHARE

Rycina 20. Artykuł dot. prognozy pacjentek z psychozą poporodową

Poniżej znajduje się abstrakt, z którego dowiadujemy się, że rodzaj badania (*design*) tu zastosowany to badanie kohortowe oparte na duńskich rejestrach. Celem badania było sprawdzenie w jakim stopniu wystąpienie

zaburzeń psychicznych, związanych z porodem, było powiązane z wystąpieniem choroby afektywnej dwubiegunowej w przyszłości (Rycina 21). Autorzy w wynikach przedstawiają, że kobiety doświadczające zaburzeń psychicznych w okresie od 0 do 14 dni po porodzie miały większą szansę konwersji do choroby afektywnej dwubiegunowej w 15-letnim okresie obserwacyjnym (relative risk = 4.25, 95% CI 3.11-5.85).

Abstract

Context: Childbirth has an important influence on the onset and course of bipolar affective disorder, and it is well established that there may be a delay of many years before receiving a diagnosis of bipolar disorder following an initial episode of psychiatric illness.

Objective: To study to what extent psychiatric disorders with postpartum onset are early manifestations of an underlying bipolar affective disorder.

Design: Survival analyses were performed in a register-based cohort study linking information from the Danish Civil Registration System and the Danish Psychiatric Central Register.

Setting: Denmark.

Participants: A total of 120,378 women with a first-time psychiatric inpatient or outpatient contact with any type of mental disorder excluding bipolar affective disorder.

Main outcome measures: Each woman was followed up individually from the day of discharge, with the outcome of interest being an inpatient or outpatient contact during the follow-up period with a first-time diagnosis of bipolar affective disorder.

Results: A total of 3062 women were readmitted or had an outpatient contact with bipolar affective disorder diagnoses. A postpartum onset of symptoms within 0 to 14 days after delivery predicted subsequent conversion to bipolar disorder (relative risk = 4.26; 95% CI =3.11-5.85). Approximately 14% of women with first-time psychiatric contacts during the first postpartum month converted to a bipolar diagnosis within the 15-year follow-up period compared with 4% of women with a first psychiatric contact not related to childbirth. Postpartum inpatient admissions were also associated with higher conversion rates to bipolar disorder than outpatient contacts (relative risk = 2.16; 95% CI = 1.27-3.66).

Conclusions: A psychiatric episode in the immediate postpartum period significantly predicted conversion to bipolar affective disorder during the follow-up period. Results indicate that the presentation of mental illness in the early postpartum period is a marker of possible underlying bipolarity.

Rycina 21. Abstrakt badania kohortowego, badającego związek między poporodowymi zaburzeniami psychicznymi z wystąpieniem choroby afektywnej dwubiegunowej

Tutaj także zrobimy pauzę, gdyż wkraczamy w grząskie tereny statystyki. Wytłumaczenie tych wyników wymaga pewnego kontekstu i temu będzie poświęcony kolejny rozdział. W nim wrócimy do znaczenia wyników, które znaleźliśmy w tej części skryptu.

Dlaczego AI zawodzi, gdy szukamy wysokiej jakości danych?

Modele sztucznej inteligencji halucynują.⁴⁴ Oznacza to, że mogą podać wiarygodnie wyglądające badanie, kiedy spytamy o pomoc. Jest jednak pewien problem – to badanie nigdy nie zostało przeprowadzone. Ostatecznie to zawsze na nas spoczywa odpowiedzialność za znalezienie badania, jego interpretację oraz wyciąganie z niego wniosków. ChatGPT, choć jest wygodny i prosty w obsłudze, zawodzi. Poniżej zresztą sam opisuje w jaki sposób, w odpowiedzi na moje pytanie.

Jestem lekarzem, który zadaje sobie pytanie kliniczne dot. skuteczności i szkód leczenia oraz prognozy mojego pacjenta. Chcę polegać na możliwie jak najwyższej jakości realnie przeprowadzonych badań naukowych, które będą zewnętrznie trafne (tzn. będą dotyczyć mojego pacjenta). Ponadto chcę dostać informacje nt. leczenia z wysokiej jakości randomizowanych badań klinicznych lub przeglądów systematycznych. W zakresie prognozy z wysokiej jakości badań kohortowych lub przeglądów systematycznych takich badań. Jeżeli spytam Ciebie o pomoc, jakie jest ryzyko, że zmyślisz dane, źle je przedstawiś, źle zinterpretujesz lub nie będziesz miał do nich dostępu?

1. Ryzyko „zmyślenia” (hallucinations)

Na czym polega problem

Jako model językowy:

- **nie mam świadomości**, czy dana praca *faktycznie istnieje*,
- generuję odpowiedzi na podstawie wzorców językowych, a nie „bazy publikacji” w sensie epistemicznym.

Dlatego mogę:

- podać **wiarygodnie brzmiące, ale nieistniejące badanie**,
- pomylić autorów, rok, czasopismo,
- przypisać wyniki do złego badania.

Skala ryzyka

- najwyższa: gdy prosisz o **konkretne cytowania bez weryfikacji**,
- niższa: gdy mówimy o **ogólnych wnioskach zgodnych z konsensusem**,
- najniższa: gdy **pracujemy na badaniach, które sam wskazujesz** (np. DOI, PMID).

2. Ryzyko błędnej prezentacji danych

Mogę:

- nieprawidłowo streścić punkt końcowy (np. myląc surrogate vs clinical outcome),
- nie podkreślić **ryzyka systematycznego błędu** (bias),
- nie odróżnić **analizy ITT od per-protocol**,
- uprościć heterogeniczność populacji.

To ryzyko rośnie:

- przy złożonych analizach statystycznych,
- przy subgrupach i analizach post hoc,
- przy badaniach o słabej jakości metodologicznej.

3. Ryzyko błędnej interpretacji (to najważniejsze)

Nawet przy poprawnych danych mogę:

- **przecenić istotność kliniczną** (np. mały ARR przy istotnym p),
- nie uwzględnić **external validity** (Twój pacjent ≠ populacja badania),
- nie odróżnić **dowodu braku efektu od braku dowodu**,
- nie uwzględnić competing risks, adherence, crossover.

Interpretacja prognozy (cohort studies) jest tu szczególnie podatna na błąd.

4. Ograniczenia dostępu do danych

- Nie mam bezpośredniego dostępu do baz takich jak PubMed, Embase, Cochrane (chyba że wyraźnie użyję wyszukiwania online).
- Nie widzę pełnych tekstów paywalled.
- Operuję głównie na:
 - wiedzy ogólnej,
 - streszczeniach,
 - ugruntowanych przeglądach,
 - standardach klinicznych.

To oznacza, że:

- mogę nie znać najnowszego badania,
- mogę nie znać szczegółów metodologii dostępnych tylko w full text.

Innymi słowy, sztuczna inteligencja może być pomocna, jednak ryzyko wprowadzenia w błąd (zwłaszcza przy interpretacji słabych metodologicznie badań) jest duże. Nie mówię, żeby nie korzystać z tego narzędzia. Proponuję jednak, żeby, formułując tzw. prompty dawać AI możliwie jak najwięcej obostrzeń, eliminując ryzyko błędnego wniosku. Nie da się tych obostrzeń stworzyć, nie znając podstaw EBM. Jeśli poprosimy chatGPT o informację w stylu „Jaki lek na psychozę byłby okej?” nie spodziewajmy się rzetelnej odpowiedzi, która może kierować naszą praktyką kliniczną.

Jeżeli już planujesz korzystać ze sztucznej inteligencji, wykorzystaj swoją wiedzę, by:

- Określić rodzaj badania, który Cię interesuje
- Wymagać linku do oryginalnego źródła
- Samodzielnie określić trafność zewnętrzną na podstawie przeczytanego artykułu
- Samodzielnie zinterpretować wyniki, biorąc pod uwagę słabości metodologiczne, takie jak błędy systematyczne czy słaby rodzaj badania.

Jak rozgryźć dane statystyczne?

W jaki sposób można przedstawić dane z badania klinicznego?

Kiedy będziesz czytać wyniki badań, spotkasz się najczęściej z jedną z dwóch sytuacji. W pierwszej możemy być zainteresowani tym, czy coś się zdarzy czy się nie wydarzy. Zgon pacjenta, nawrót, remisja, wystąpienie choroby – albo się wydarzy, albo nie. W przypadku naszej pacjentki albo nastąpi progresja do przewlekłej choroby psychicznej (np. choroby afektywnej dwubiegunowej), albo nie. Taki rodzaj danych nazywa się **danymi binarnymi (*binary*)**. Dla takich danych określamy ich prawdopodobieństwo, które najczęściej przedstawiane jest w procentach [%]. Np. w jednej grupie 10% pacjentów zmarło, w innej 14%. W jednej 60% miało progresję, w drugiej 78% itd.

W innych przypadkach dane będą mierzone liczbowo. Będziemy wtedy zainteresowani nasileniem danego zjawiska, zmianą stężenia. Ocena bólu, nasilenie depresji, przyrost masy ciała w kilogramach, poprawa w zakresie dolegliwości chorobowych. Tego typu zjawiska są mierzone przy użyciu liczb, by wyrazić zmianę. W przypadku naszej pacjentki leki, których używamy, mogą zwiększyć stężenie prolaktyny w surowicy, co może mieć konsekwencje dla jej funkcjonowania. Taki rodzaj danych nazywa się **danymi ciągłymi (*continuous*)**. Ból w jednej grupie spadł z 7/10 do 4/10, czyli średnio o 3 punkty. Depresja zmniejszyła nasilenie w jednej grupie z 35/40 do 10/40,

a w drugiej z 33/40 do 25/40. Dla takich danych określamy różnicę. O ile coś się zmieniło?

Zrozumienie tego to połowa sukcesu. Na co patrzeć? Czy to prawdopodobieństwo czy różnica? Kiedy już nabierzesz wprawy w rozumieniu na co patrzysz, kolejne kroki staną się dużo prostsze. Istnieją standardy raportowania zarówno danych „albo-albo” (binarnych), jak i „o ile się zmieniło?” (ciągłych). Teraz omówię na przykładach, w jaki sposób te dwa scenariusze są przedstawiane w badaniach klinicznych.

Liczenie ryzyka, szans, relatywnego ryzyka (RR) i relatywnych szans (OR)

W przypadku danych binarnych mówimy o ryzyku, relatywnej zmianie ryzyka (*relative risk reduction/increase*) lub absolutnej zmianie ryzyka (*absolute risk reduction/increase*). Jak to działa? Wyobraźmy sobie hipotetyczne badanie, trwające rok, w którym sprawdzano czy lek X zmniejszy śmiertelność wśród osób z chorobą A.

	Przeżycie	Zgon
<i>Grupa kontrolna</i> (<i>n=100</i>)	90	10
<i>Grupa leczona</i> (<i>n=100</i>)	95	5

$$\text{Ryzyko zgonu w grupie kontrolnej} = \frac{10}{100} \times 100\% = 10\%$$

$$\text{Ryzyko zgonu w grupie leczonej} = \frac{5}{100} \times 100\% = 5\%$$

Relatywne ryzyko liczy się poprzez dzielenie ryzyka w grupie badanej przez ryzyko w grupie kontrolnej.

$$\text{Relatywne ryzyko (RR)} = \frac{\text{ryzyko w grupie leczonej}}{\text{ryzyko w grupie kontrolnej}} = \frac{5\%}{10\%} = 0.5$$

Interpretacja: Na koniec rocznego okresu obserwacyjnego lek X zmniejszył relatywne ryzyko zgonu z powodu choroby A o **połowę** lub o **50%** w porównaniu z placebo.

Absolutną redukcję ryzyka z kolei liczy się poprzez odejmowanie ryzyka w grupie leczonej od ryzyka w grupie kontrolnej.

Absolutna redukcja ryzyka = ryzyko w grupie kontrolnej – ryzyko w grupie leczonej = 10% - 5% = 5%.

Interpretacja: Na koniec rocznego okresu obserwacyjnego na 100 osób leczonych lekiem X o 5 mniej umrze z powodu choroby A w porównaniu ze 100 osobami leczonymi placebo.

Pochodną absolutnej redukcji ryzyka jest liczba osób potrzebnych do wyleczenia (*number needed to treat*). Liczy się go poprzez dzielenie 100% przez absolutną redukcję ryzyka. NNT mówi nam o tym, ile osób należy leczyć, żeby wyleczyć jedną dodatkową osobę względem interwencji stosowanej w grupie kontrolnej (placebo/nierobienie niczego).

$$\text{NNT (number needed to treat)} = \frac{100\%}{\text{absolutna redukcja ryzyka}} = \frac{100\%}{5\%} = 20.$$

Interpretacja: Jeżeli będziemy leczyć **20 osób** lekiem X przez rok, **zapobiegniemy jednemu dodatkowemu zgonowi** względem osób przyjmujących placebo.

Statystycznie **szanse** mają się nieco inaczej. Szansa to prawdopodobieństwo wystąpienia danego zdarzenia podzielone przez

prawdopodobieństwo niewystąpienia tego zdarzenia. Zatem zamiast 10% ryzyka (10/100) będziemy mieli nieco inne szanse (10/90), gdzie 90 to nie całkowita liczba osób w grupie, ale liczba osób, która nie zmarła.

$$\text{Szanse zgonu w grupie kontrolnej} = \frac{10\%}{90\%} = 0.11$$

$$\text{Szanse zgonu w grupie leczonej} = \frac{5\%}{95\%} = 0.053$$

$$\text{Relatywne szanse (OR)} = \frac{\text{szanse w grupie leczonej}}{\text{szanse w grupie kontrolnej}} = \frac{0.053}{0.11} = 0.48$$

Interpretacja: Na koniec rocznego okresu obserwacyjnego lek X zmniejszył relatywne szanse zgonu z powodu choroby A o **nieco ponad połowę** lub o **52%** w porównaniu z placebo.

RR ≠ OR

Rozważmy poniższą sytuację.

	Przeżycie	Zgon
Grupa kontrolna (n=100)	55	45
Grupa leczona (n=100)	70	30

W tym przypadku **szanse** zgonu są następujące:

$$\text{Szansa zgonu w grupie kontrolnej} = \frac{45}{55} = 0.82$$

$$\text{Szansa zgonu w grupie leczonej} = \frac{30}{70} = 0.43$$

$$\text{Wówczas OR} = \frac{0.43}{0.82} = 0.52$$

$$\text{RR zaś wynosi} \frac{30\%}{45\%} = 0.67$$

Wartości OR zawsze wskazują na bardziej dramatyczny efekt niż RR, a jest to szczególnie widoczne w przypadku zdarzeń, które występują z częstością >10%. W przypadku zaś zdarzeń rzadszych $OR \approx RR$.

Interpretacja RR, OR i HR

Wszystkie wyniki przedstawiane w sposób relatywny – relatywne ryzyko, szanse i szkody (*HR – hazard ratio*) interpretuje się podobnie.

RR/HR – mówimy, że **ryzyko** jest większe/mniejsze/takie samo

OR – mówimy, że **szanse** są większe/mniejsze/takie same

Jeżeli $RR/OR/HR = 1$, wówczas ryzyko/szanse w obu badanych grupach są identyczne.

Jeżeli $RR/OR/HR > 1$, wówczas ryzyko/szanse są wyższe w pierwszej grupie (najczęściej badanej) względem drugiej grupy (najczęściej kontrolnej). Można to zinterpretować na dwa sposoby:

$RR = 1.6$ (np. $\frac{16\%}{10\%}$), tzn. relatywne ryzyko zdarzenia w pierwszej grupie **jest o 60% większe** niż w drugiej grupie ($1.6 - 1 = 0.6$; $0.6 \times 100\% = 60\%$) LUB ryzyko zdarzenia w pierwszej grupie jest **relatywnie większe o ponad połowę** niż w grupie drugiej.

$OR = 3.5$ (np. $\frac{7\%}{2\%}$), tzn. szanse zdarzenia w pierwszej grupie są **relatywnie o 250% wyższe** niż w drugiej grupie ($3.5 - 1 = 2.5$; $2.5 \times 100\% = 250\%$) LUB szanse zdarzenia w pierwszej grupie są **relatywnie 3.5-krotnie wyższe** niż w drugiej grupie.

Jeżeli RR/OR/HR < 1, wówczas ryzyko/szansa są niższe w pierwszej grupie (najczęściej badanej) względem drugiej grupy (najczęściej kontrolnej).

HR = 0.87, tzn. ryzyko zdarzenia w pierwszej grupie jest **relatywnie o 13% mniejsze** niż w drugiej grupie ($1 - 0.87 = 0.13$; $0.13 \times 100\% = 13\%$).

OR = 0.25 (np. $\frac{4\%}{16\%}$), tzn. szansa zdarzenia w pierwszej grupie są **relatywnie o 75% niższe** niż w drugiej grupie ($1 - 0.75 = 0.25$; $0.25 \times 100\% = 25\%$) LUB szansa zdarzenia w pierwszej grupie są **relatywnie czterokrotnie niższe** niż w drugiej grupie.

Pomiędzy relatywnym ryzykiem (RR) a ilorazem szkód/hazardu (HR – *hazard ratio*) istnieją pewne różnice:

1. RR jest (zazwyczaj) wynikiem prostego dzielenia, np. $10\%/5\% = 2$. To jest relatywne ryzyko. Mówimy o relatywnym ryzyku **na koniec** danego czasu, np. rocznej obserwacji.
2. HR jest wynikiem modelu statystycznego, nazywanego modelem proporcjonalnego hazardu Coxa.⁴⁵ Metoda ta jest podobna do wyliczania RR, ale bierze pod uwagę czas do wystąpienia zdarzenia (np. śmierci). Jest używana w analizach m.in. śmiertelności. Interpretując HR, mówimy o relatywnym ryzyku **w trakcie okresu obserwacyjnego**, np. rocznej obserwacji. Czyli innymi słowy HR pozwala ocenić względny efekt leczenia na przebieg zdarzeń w czasie, a nie tylko na ich całkowitą częstość.
3. Zarówno RR jak i HR mówią zatem o relatywnym ryzyku, ale trochę inaczej. Interpretacja jednak pozostaje podobna: ryzyko relatywnie zwiększa się lub maleje.

Czy uzyskane wyniki są przypadkowe, czyli dwa słowa o 95% przedziale ufności.

95% przedział ufności (95% confidence interval) jest wyliczany na podstawie danych zebranych w danym badaniu. Wyobraź sobie grupę 1000 ludzi, którym mierzymy masę ciała – średnia wartość w tej grupie wynosi 70kg. Następnie wyobraź sobie, że z tej grupy 100 razy losujemy po 100 przypadkowych osób i mierzymy ponownie ich masę ciała. W tak wykonanym, powtarzanym eksperymencie, formalnie, około 19 na 20 (95%) wyliczonych 95% przedziałów ufności dla każdej jednej grupy 100 osób będzie zawierało prawdziwą średnią populacyjną, czyli 70 kg. Około 1 na 20 (czyli 5%) wyliczonych 95% przedziałów ufności nie będzie zawierało tej średniej. Tak formalnie definiuje się 95% przedział ufności. Jest to jednak dość skomplikowane, nie sądzisz?

Proponuję nieco uproszczone, choć nie w pełni poprawne myślenie na temat tego parametru. 95% przedział ufności wyznacza przedział wartości, w którym na 95% znajduje się prawdziwa wartość dla całej populacji, którą jesteśmy zainteresowani. Im badana grupa jest większa, tym przedział ufności jest węższy. Im więcej osób zbadamy, tym bardziej pewni jesteśmy, że jesteśmy blisko prawdziwej odpowiedzi na nasze pytanie.

Rozważmy przedstawiony wcześniej przykład:

Relatywne ryzyko (RR) wynosiło 0.5, a przedział ufności wyliczony odpowiednią metodą: (95% CI 0.18–1.41).

Interpretacja: W badanej grupie osób z chorobą A uzyskaliśmy relatywną redukcję śmiertelności o połowę na koniec rocznego okresu obserwacyjnego wśród pacjentów przyjmujących lek X, **ALE** gdybyśmy przebadali całą populację osób z chorobą A, mamy 95% pewności,

że efekt naszego leku plasowałby się między relatywnym **zmniejszeniem śmiertelności o 82%** ($1-0.18$) a relatywnym **zwiększeniem śmiertelności o 41%** ($1.41-1$) w porównaniu z grupą kontrolną.

W związku z tym, że przedział ufności wskazuje na możliwe przeciwne efekty, tj. zwiększenie lub zmniejszenie śmiertelności, uznajemy ten wynik za **nieistotny statystycznie**. Innymi słowy uzyskany w badaniu **efekt** redukcji śmiertelności **mógł być przypadkowy**. Nie wiemy czy ten lek tak naprawdę pomaga, czy szkodzi, więc nie będziemy go używać.

Rozważmy tę samą chorobę i ten sam lek, ale znacznie większą populację:

	PRZEŻYCIE	ZGON
GRUPA KONTROLNA (N=1000)	900	100
GRUPA LECZONA (N=1000)	950	50

Ryzyko zgonu w obu grupach pozostaje takie samo jak w poprzednim badaniu. Relatywne ryzyko również. Przedział ufności uległ jednak zmianie. Relatywne ryzyko (RR) wynosi 0.5, a przedział ufności: (95% CI 0.36–0.69).

Interpretacja: W badanej grupie osób z chorobą A uzyskaliśmy relatywną redukcję śmiertelności o połowę na koniec rocznego okresu obserwacyjnego wśród pacjentów przyjmujących lek X **ALE** gdybyśmy przebadali całą populację osób z chorobą A, mamy 95% pewności, że nasz lek relatywnie zmniejszyłby śmiertelność **o gdzieś między 31%** ($1-0.69$) **a 64%** ($1 - 0.36$) w porównaniu z grupą kontrolną.

Przedział ufności wskazuje jednoznacznie, że lek zmniejsza śmiertelność, w związku z czym uznajemy wynik za **istotny statystycznie**. Innymi słowy, uzyskany efekt redukcji śmiertelności o połowę nie jest przypadkowy. Przebadawszy większą grupę, mamy pewność co do kierunku i siły działania leku. Możemy użyć go w praktyce. Pierwsze badanie miało mniej pacjentów, a to zawsze wiąże się z większą niepewnością. Pomimo że pierwsze badanie słusznie wykazało korzystny wpływ leku, efekt ten mógł być przypadkowy ze względu na małą grupę.

UWAGA! Jeżeli przedział ufności dla relatywnego ryzyka/szans zawiera w sobie „1”, wynik również uznajemy za nieistotny statystycznie. Wskazuje to wówczas na niepewny efekt, który nie upoważnia nas do tego, żeby powiedzieć, że znaleźliśmy nieprzypadkową różnicę.

Nie tylko dla relatywnych szans czy ryzyka wyznacza się przedział ufności. Można go też wyznaczyć dla innych danych. Rozważmy przykład wpływu leku na ból w badaniu, w którym nasilenie bólu określa się przy pomocy 11-punktowej skali, gdzie 0 to brak bólu, a 10 – najsilniejszy wyobrażalny ból (zmienna ciągła). W tym hipotetycznym badaniu średnio lek zmniejszył nasilenie bólu o 2 punkty. Jeżeli przedział ufności **nie będzie zawierał braku efektu**, czyli 0, uznamy wynik za **istotny statystycznie**. Tym samym będziemy mogli stwierdzić, że lek jest skuteczny. Np.:

Nasilenie bólu zmieniło się średnio o -2 punkty, tj. zmalało o 2 punkty (95% CI -1 do -3), tym samym mamy 95% pewności, że w ogólnej populacji badanej lek zmniejszałby ból średnio o gdzieś między 1 a 3 punkty.

Jeżeli zaś przedział ufności będzie zawierał 0 (**brak efektu**), wynik uznamy za **nieistotny statystycznie**, a lek za nieskuteczny. Np.:

Nasilenie bólu zmieniło się średnio o -2 punkty, czyli zmniejszyło się o 2 punkty (95% CI: -4 do 1). Oznacza to, że z 95% pewnością w populacji badanej lek mógłby średnio zmniejszać ból nawet o 4 punkty, ale też potencjalnie zwiększać jego nasilenie maksymalnie o 1 punkt.

Odpowiedzi na pytania dot. naszej pacjentki

Wróćmy teraz do odpowiedzi na pytania, które zadaliśmy w kontekście naszej pacjentki. Najpierw rozważmy wpływ leków na stężenie prolaktyny we krwi:

- Aripiprazol: -7.5 ng/mL (95% CI -40 ng/mL do 25 ng/mL)
- Risperidon: 34 ng/mL (95% CI 13 ng/mL do 55 ng/mL)

Z tych danych widzimy, że aripiprazol średnio zmniejsza stężenie prolaktyny o 7.5 ng/mL, jednak na 95% efekt leku zawiera się gdzieś między zmniejszeniem stężenia o do 40 ng/mL a zwiększeniem tegoż stężenia do 25 ng/mL. Ze względu na to, że 95% przedział ufności zawiera wartości o przeciwnych znakach, uznajemy efekt za nieistotny statystycznie. Innymi słowy, nie możemy powiedzieć, żeby aripiprazol wpływał istotnie na stężenie prolaktyny u kobiet z psychozą.

Z drugiej strony mamy risperidon, który średnio zwiększa stężenie prolaktyny o 34 ng/ml, z czego na 95% średni efekt zawiera się gdzieś między zwiększeniem stężenia o 13 ng/mL a 55 ng/mL. W związku z tym, że wartości zawarte w 95% przedziale ufności jednoznacznie wskazują na zwiększenie stężenie prolaktyny, wynik uznajemy za istotny statystycznie. Możemy powiedzieć, że risperidon powoduje zwiększenie stężenia prolaktyny we krwi kobiet z psychozą.

Rozważmy teraz drugą omawianą sytuację dot. prognozy. Kobiety doświadczające zaburzeń psychicznych w okresie od 0 do 14 dni

po porodzie miały większą szansę konwersji do choroby afektywnej dwubiegunowej w 15-letnim okresie obserwacyjnym (relative risk = 4.25, 95% CI 3.11-5.85). Wiedząc już, w jaki sposób analizować relatywne ryzyko, możemy powiedzieć, że wystąpienie zaburzeń psychicznych we wczesnym okresie po porodzie jest powiązane z ponad 4-krotnie zwiększonym ryzykiem zachorowania na chorobę afektywną dwubiegunową w 15-letnim okresie obserwacyjnym względem kobiet, które nie miały poporodowych zaburzeń psychicznych. Wynik jest nieprzypadkowy (tj. istotny statystycznie), ponieważ wartości zawarte w 95% przedziale ufności jednoznacznie wskazują na wzrost ryzyka.

Zapoznaliśmy się z interpretacją danych statystycznych oraz odczytaliśmy wyniki znalezionych wcześniej badań. Teraz warto zwrócić uwagę na potencjalne pułapki, w które można wpaść, analizując dane.

Statystyka dla klinicystów: Jak nie dać się zwieść liczbom?

Ponieważ współczesna medycyna w dużym stopniu opiera się na matematyce, warto rozumieć podstawowe zagadnienia statystyczne. Ten rozdział wykracza nieco poza podstawowe rozumienie statystyki. Ma stanowić listę punktów, na które zwrócić uwagę podczas czytania artykułu. Wzięcie tych aspektów pod uwagę pomoże Ci ocenić jakość czytanej pracy.^{46,47}

1. Czy grupy są porównywalne?

Pierwszym krokiem jest sprawdzenie czy autorzy poprawnie zaprojektowali badanie. Większość badań klinicznych zawiera tabelę lub akapit opisujący **charakterystykę wyjściową** (*baseline characteristics*) badanych grup. W przypadku badań randomizowanych tabela ta powinna pokazywać, że grupa interwencyjna i kontrolna są do siebie podobne pod względem wieku, płci oraz kluczowych zmiennych (np. średniej wielkości

guza nowotworowego). W badaniach obserwacyjnych, jeśli różnice między grupami nie zostaną zaadresowane, znacznie podważa to wartość przedstawionych wyników.

2. Rodzaje danych a dobór testów.

Liczby w medycynie często służą do etykietowania właściwości przedmiotów. O ile wzrost czy masę ciała opisujemy używając liczb i wyciągamy z nich średnią, o tyle w innych przypadkach jest to niemożliwe. Wyobraź sobie badanie, w którym miasto pochodzenia oznaczono liczbami: 1 = Londyn, 2 = Manchester, 3 = Birmingham. Obliczenie średniej w tym wypadku byłoby niezasadne. Podobnie jest ze skalą upodobań (np. 1 = wcale, 2 = trochę, 3 = bardzo) – wynik będzie nieinterpretowalny, chyba że wiemy, że różnica między „wcale” a „trochę” jest dokładnie taka sama, jak między „trochę” a „bardzo”.

Wszystkie testy statystyczne dzielimy na dwie grupy:

- **Testy parametryczne (*parametric*):** Zakładają one, że dane pochodzą z populacji o konkretnym rozkładzie, zazwyczaj **normalnym** (*normal distribution*), czyli przypominającym kształtem dzwon.
- **Testy nieparametryczne (*non-parametric*):** Nie czynią takich założeń; patrzą jedynie na kolejność wartości (która jest najmniejsza, która kolejna itd.), ignorując bezwzględne różnice między nimi.

Testy parametryczne łatwiej wykazują istotne statystycznie różnice. Dlatego badacze czasem ulegają pokusie ich niewłaściwego użycia (np. współczynnika *r* Pearsona), aby uzyskać (pozornie) istotne wyniki. Jeśli dane mają rozkład **skośny** (*skewed*), tj. nienormalny, używanie testów dla rozkładu normalnego jest nadużyciem.

3. Korelacja i regresja: To nie są synonimy!

Bywają używane zamiennie, ale spełniają inne funkcje i używa się ich w różnych kontekstach.

- **Współczynnik korelacji (r) Pearsona:** Mówi nam, jak silny jest związek między dwiema zmiennymi. Aby był wiarygodny, dane muszą mieć rozkład normalny, być od siebie niezależne, a na każdym badanym można wykonać tylko jeden pomiar. Każdemu r powinna towarzyszyć wartość p lub przedział ufności, żeby można było ocenić istotność statystyczną.
- **Regresja (*regression*):** To równanie matematyczne, które pozwala **przewidzieć** wartość jednej zmiennej na podstawie innej. Sugeruje ona kierunek wpływu, ale – co ważne – nadal nie dowodzi, że jedno zjawisko powoduje drugie.

Na przykład Twoja masa ciała zależy od wysokości, ale zależność ta nie jest liniowa — lepiej opisuje ją funkcja kwadratowa. A i tak sama informacja o wzroście to za mało, by precyzyjnie przewidzieć masę ciała – musielibyśmy dołożyć do równania wiek, płeć czy aktywność fizyczną.

4. Pułapka przyczynowości, czyli błąd ekologiczny

To, że w mieście z dużym bezrobociem jest wysoka przestępczość, nie oznacza automatycznie, że to bezrobotni kradną. To klasyczny **błąd ekologiczny** (*ecological fallacy*) – obecność związku między A i B nie mówi nam nic o kierunku przyczynowości.

Zanim uwierzysz, że A powoduje B, odpowiedz sobie na te pytania:

- Czy związek jest silny i spójny w różnych badaniach?
- Czy przyczyna występuje przed skutkiem?
- Czy istnieje „zależność od dawki” (im więcej A, tym więcej B)?
- Czy ma to sens biologiczny i epidemiologiczny?

5. Wartość p: prawda czy przypadek?

Umownie przyjmujemy, że $p < 0,05$ oznacza istotność statystyczną. Oznacza to, że szansa na to, iż wynik jest dziełem czystego przypadku, wynosi mniej niż 1 do 20 (więcej na ten temat w e-learningu).

Pamiętaj jednak o dwóch problemach:

1. **Statystyczny fart:** Skoro $P < 0,05$ to 1 szansa na 20, to statystycznie co dwudziesty opublikowany “statystycznie istotny” wynik może być po prostu wynikiem przypadkowym, który wygląda na ważny. Jeśli badacze sprawdzają wiele różnych wyników naraz, powinni użyć **korekty Bonferroniego**, by uniknąć takich pomyłek (ewentualnie innych poprawek statystycznych celem uniknięcia generowania fałszywie istotnych wyników). Korekta Bonferroniego polega na tym, że gdy wykonuje się wiele testów statystycznych jednocześnie, poziom istotności dzieli się przez liczbę testów, aby ograniczyć ryzyko fałszywie pozytywnych wyników (**błąd typu I**). Dzięki temu każdy test musi spełnić bardziej rygorystyczny próg istotności, co zmniejsza szansę przypadkowego uznania wyniku za istotny.
2. **Brak różnicy czy mała grupa?** Jeśli p jest wysokie (nieistotne), może to oznaczać brak efektu, ale równie dobrze może oznaczać, że uczestników było zbyt mało, by ten efekt wykazać.

Podczas gdy wartość p mówi jedynie o istotności statystycznej (tak/nie), 95% przedział ufności daje więcej informacji i warto z niego korzystać, jeśli jest dostępny.

6. Pułapka interpretacji przedziału ufności

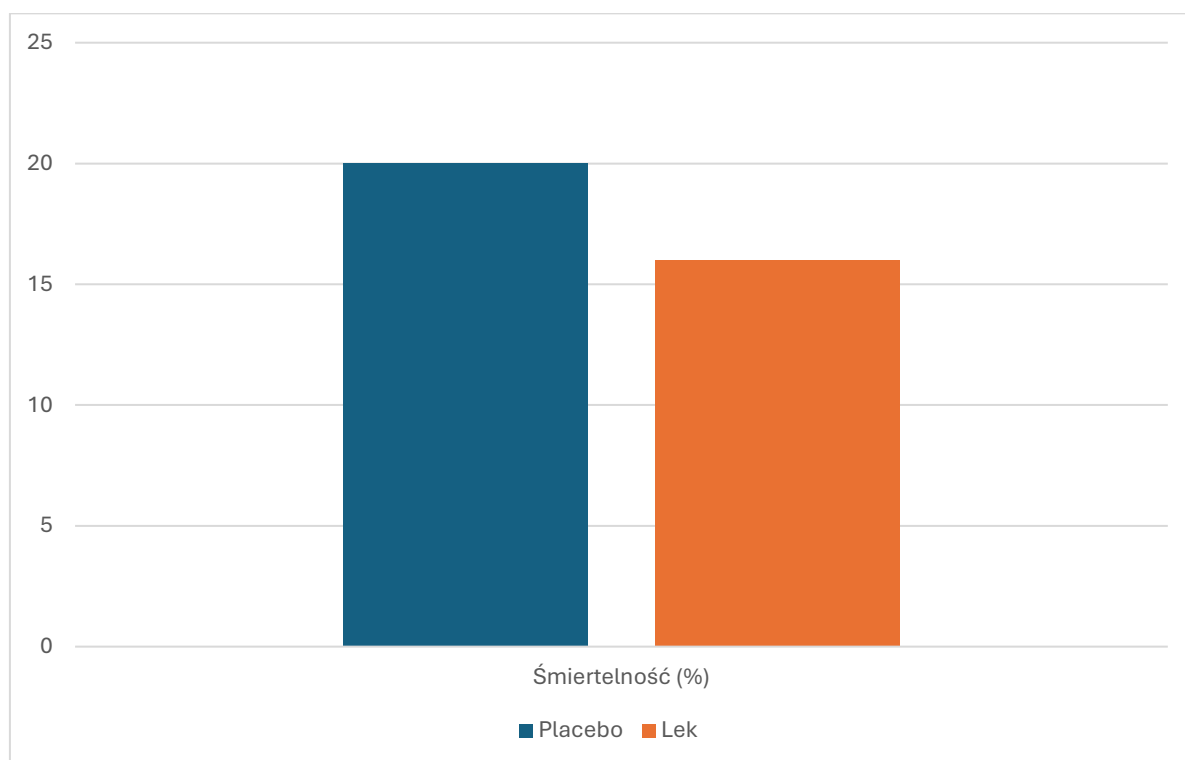
95% przedział ufności pozwala oszacować, czy efekt jest silny i czy uzyskany wynik jest nieprzypadkowy (tj. wynik istotny statystycznie). Jeśli przedział ufności dla różnicy między grupami zawiera brak efektu,

uznajemy badanie za nieistotne statystycznie (negatywne). Ale uwaga! Jeśli badanie jest „negatywne”, ale górna granica przedziału ufności nadal obejmuje klinicznie ważną korzyść, to badanie jest **niedefinitywne** – po prostu potrzebujemy większej grupy pacjentów, by mieć większą pewność. Im większe badanie, tym węższy przedział ufności i bardziej precyzyjny wynik.

7. Lek zmniejsza ryzyko – ale które?

Zawsze, gdy mowa o ryzyku, zadaj sobie pytanie, czy mowa o absolutnym, czy relatywnym ryzyku. Dlaczego? Spójrz na ten przykład:

- Lek A zmniejsza śmiertelność o 20% (**relatywnie**).
- Lek B zmniejsza śmiertelność o 4% (**absolutnie**).
- Jeżeli zastosujemy lek C u 25 pacjentów, uratujemy jedną dodatkową osobę (**NNT**).



Większość decydentów wybierze lek A, nie wiedząc, że wszystkie te liczby opisują dokładnie ten sam wynik. Marketing farmaceutyczny uwielbia relatywną redukcję ryzyka (RRR), bo 20% wygląda lepiej niż 3%. Pamiętaj, żeby zwracać uwagę na absolutną redukcję ryzyka, relatywną redukcję ryzyka i NNT, by wiedzieć, jakiej korzyści należy się spodziewać.

Istotność statystyczna kontra kliniczna: Czy każda „gwiazdka” w badaniu ma znaczenie dla Twojego pacjenta?

To najważniejszy punkt dla Ciebie jako przyszłego lekarza / lekarki. Wynik czasem może być „wysoce istotny statystycznie” (np. $P < 0,001$), ale dla pacjenta zupełnie bez znaczenia (np. redukcja ryzyka zgonu o 0.05%).

Wielu badaczy i studentów medycyny przypisuje zbyt duże znaczenie wartości P . Przyzwyczailiśmy się myśleć, że jeśli wynik jest istotny statystycznie, to automatycznie staje się ważny. Nic bardziej mylnego. W świecie medycyny opartej na dowodach musisz umieć odróżnić matematyczne prawdopodobieństwo od realnej korzyści dla chorego, czyli **istotności klinicznej** (*clinical significance*).⁴⁸

Czym różnią się te dwa pojęcia?

1. **Istotność statystyczna** to czysta matematyka. Mówi nam ona jedynie, czy obserwowany efekt jest dziełem przypadku, czy nie. Wartość $p < 0,05$ oznacza, że prawdopodobieństwo tego, że obserwowany efekt jest dziełem przypadku, wynosi mniej niż 5%. Nie mówi nam jednak absolutnie nic o tym, jak duży jest ten efekt.
2. **Istotność kliniczna** to pytanie o sens: „Czy ten efekt jest na tyle duży, aby był ważny dla mojego pacjenta?”. Podczas gdy istotność statystyczna daje czarno-białą odpowiedź „tak/nie”, istotność kliniczna zależy od kontekstu i często jest kwestią subiektywnej oceny klinicysty oraz samego pacjenta.

Można by się zastanowić, jak to możliwe, że wynik jest istotny statystycznie, a jednocześnie nieprzydatny klinicznie. Odpowiedzią jest **rozmiar grupy badanej** (*sample size*). Przy ogromnych grupach pacjentów (liczonych w dziesiątkach lub setkach tysięcy) statystyka nabiera nieco innej rozdzielczości. Pozwala wykryć nawet najmniejsze, wręcz mikroskopijne różnice między grupami, które nie wynikają z przypadku, ale nie mają żadnego praktycznego znaczenia.

Spójrzmy na dwa przykłady:

1. **Ćwiczenia a masa ciała.** Wielkie, 13-letnie badanie na grupie ponad 34 000 kobiet wykazało, że intensywnie ćwiczące panie tyją istotnie mniej niż te, które ćwiczą mało ($P < 0,003$).⁴⁹ Brzmi świetnie? Przyjrzyjmy się twardym danym: różnica w przyroście masy ciała wyniosła w tym badaniu zaledwie **0,12 kg w ciągu 3 lat**. W ciągu całej dekady intensywniej ćwiczące pacjentki schudły mniej niż pół kilograma wagi względem osoby niećwiczącej. Statystycznie? Sukces. Klinicznie? Trudno nazwać to motywującym wynikiem.
2. **Alkohol i sport.** Analiza narodowych ankiet (ponad 230 000 osób) wykazała „silnie istotny” związek między piciem alkoholu a aktywnością fizyczną ($P < 0,01$).⁵⁰ Okazało się, że osoby pijące ćwiczą o całe... **7,2 minuty więcej w tygodniu** niż abstynenci. To klasyczny przykład wyniku, który jest statystycznie istotny, ale klinicznie nieważny.

Jako lekarz musisz patrzeć poza wartość p. Nawiązując do wcześniej opisanych informacji, warto pamiętać o następujących zasadach:

- **Skup się na rozmiarze efektu (*effect size*):** Zawsze sprawdzaj bezwzględne liczby. Czy spadek ciśnienia o 1 mmHg, nawet z $p < 0,001$, jest wart włączania leku ze skutkami ubocznymi?

- **Analizuj 95% przedział ufności:** To Twoje najważniejsze narzędzie. Istotność statystyczna mówi tylko, czy CI nie zawiera zera (lub jedynki w przypadku RR/OR/HR). Istotność kliniczna każe Ci zapytać: „**Czy nawet najniższa wartość w tym przedziale jest dla mnie ważna?**”. Jeśli CI dla różnicy w wadze wynosi 0,04 do 0,20 kg, to nawet najbardziej optymistyczny scenariusz (0,20 kg w 3 lata) nie zmienia Twojego postępowania klinicznego.
- **Zachowaj sceptycyzm wobec ogromnych badań:** Jeśli widzisz setki tysięcy uczestników i imponujące wartości p, bądź potrójnie czujny na kliniczną nieistotność wyników.

Prawidłowe i błędne użycie dowodów naukowych

Na EBM składają się trzy elementy, jak pisałem wcześniej. Ten skrypt w większości był poświęcony tylko jednemu – dowodom naukowym. W dzisiejszym świecie bywają one podważane. Istnieje grupa ludzi silnie sceptycznie nastawiona do konwencjonalnych dowodów naukowych, zarzucając nieraz fałszowanie danych czy ukrywanie informacji. Padają zarzuty pod adresem firm farmaceutycznych, które są obwiniane o nieetyczne i nietransparentne prowadzenie badań. Z drugiej strony istnieje też spora grupa entuzjastów nauki, którzy w badaniach i danych doszukują się pewności, stabilności oraz rzetelnych odpowiedzi. Rzeczywistość nie zawsze jest jednak czarno-biała, a ostatnie dekady dostarczają licznych przykładów na dobre i złe praktykowanie nauki. Poniżej przedstawiam dwie zachęcające historie, ilustrujące siłę solidnie wygenerowanych dowodów oraz dwie historie, które pokazały, w jaki sposób można nadużyć naukowych informacji w celu osiągnięcia własnych celów. Najwięcej błędów w EBM nie bierze się ze złych danych, tylko z ich złej interpretacji.

Kiedy dowody zmieniły świat na lepsze

Historia tytoniowej epidemii: Jak Doll i Hill udowodnili, że palenie zabija
Wyobraź sobie gabinet lekarski w latach 40. XX wieku. Twój profesor siedzi przy biurku, wypisuje receptę i jednocześnie spokojnie dopala

papierosa, strzepując popiół do kryształowej popielniczki. W tamtych czasach palenie było normą, a lekarze – paradoksalnie – palili równie dużo, co ich pacjenci. Nikt nie podejrzewał, że ten niewinny nawyk stoi za gwałtownym wzrostem liczby zgonów na raka płuca, który w Anglii i Walii wzrósł piętnastokrotnie w ciągu zaledwie ćwierćwiecza. Poniżej opiszę historię dwóch badaczy, Richarda Dolla i Austina Bradforda Hilla, którzy dzięki ustrukturyzowanemu podejściu i cierpliwości zmienili bieg historii medycyny.⁵¹

Wszystko zaczęło się od pytania (jak to zwykle bywa): dlaczego rak płuca, wcześniej rzadki, staje się plagą? W 1950 roku Doll i Hill przeprowadzili **badanie kliniczno-kontrolne** (*case-control study*).⁵² Polegało ono na spoglądaniu wstecz – badacze przepytali pacjentów z rakiem płuca oraz zdrowe osoby z grupy kontrolnej o ich przeszłe nawyki.

Wyniki były szokujące: wśród 649 mężczyzn z rakiem płuca tylko dwóch (0,3%) nigdy nie paliło. Ryzyko zachorowania rosło proporcjonalnie do liczby wypalanych papierosów – u osób palących ponad 25 sztuk dziennie było ono aż 50-krotnie wyższe niż u niepalących. Choć badanie to było mocnym sygnałem, wielu naukowców (i przedstawicieli koncernów tytoniowych) twierdziło, że to tylko korelacja, a nie dowód przyczynowy.

Doll i Hill wiedzieli, że potrzebują silniejszego dowodu. W 1951 roku rozpoczęli jedno z najważniejszych badań w historii EBM – prospektywne badanie kohortowe, znane jako *British Doctors Study*.⁵³ Wybór grupy badanej był genialny. Zamiast przypadkowych ludzi, zaprosili do badania ponad 34 000 brytyjskich lekarzy. Dlaczego? Bo lekarze są łatwi do śledzenia (muszą być w rejestrze medycznym, by pracować) i zazwyczaj rzetelniej opisują swoje nawyki. Od tego momentu badacze nie pytali już „na co chorowałeś?”, ale obserwowali: „kto z was zachoruje

lub umrze w przyszłości?”. Badanie to trwało aż 50 lat. Oto co przyniosły kolejne raporty:

- **Po 10 latach (1964):** Potwierdzono, że palenie wiąże się nie tylko z rakiem płuca, ale też z przewlekłym zapaleniem oskrzeli, gruźlicą i chorobą wieńcową. Roczny współczynnik zgonów na raka płuca wynosił 0,07 na 1000 u niepalących, ale wzrastał do 3,15 u palących 35 i więcej papierosów dziennie.⁵⁴
- **Po 20 latach (1976):** Wykazano, że palenie powoduje śmierć głównie poprzez choroby serca, raka płuca i choroby naczyniowe. Oszacowano, że od 1/3 do 1/2 palaczy umrze z powodu swojego nałogu.⁵⁵
- **Po 40 i 50 latach (1994–2004):** Ostatnie raporty (współtworzone przez Richarda Peto) pokazały, że wcześniejsze analizy wręcz nie doceniały szkodliwości tytoniu. Okazało się, że osoby palące przez całe życie umierają średnio **10 lat wcześniej** niż osoby niepalące.^{56,57}

Badanie to dostarczyło nam też niezwykle ważnych danych na temat rzucenia palenia. Doll i Hill wykazali, że rzucenie palenia w dowolnym wieku przynosi korzyść, ale im wcześniej to zrobisz, tym lepiej:

- Rzucenie palenia w wieku **30 lat** pozwala uniknąć niemal całości nadmiarowego ryzyka zgonu.
- Rzucenie w wieku **40 lat** daje dodatkowe **9 lat życia**.
- Nawet rzucenie w wieku **60 lat** zyskuje dla pacjenta dodatkowe **3 lata życia**.

Historia *British Doctors Study* to lekcja o tym, jak rzetelna nauka potrafi przebić się przez mur niedowierzania i marketingu. Dzięki temu, że Doll i Hill nie poprzestali na dowodach anegdotycznych, ale zastosowali

rygorystyczne metody badawcze, Ty możesz dzisiaj z pełnym przekonaniem powiedzieć swojemu pacjentowi: „Rzucenie palenia realnie wydłuży Panu życie”. Czasem jedno, rzetelnie prowadzone przez 50 lat badanie, jest warte więcej niż tysiące szybkich doniesień.

Richard Doll sam rzucił palenie w trakcie prowadzenia tych badań, co prawdopodobnie pozwoliło mu dożyć 92 lat i osobiście zaprezentować wyniki 50-letniej obserwacji.

Sterydy w COVID-19 – od teorii do twardych dowodów

Początek pandemii COVID-19 był bardzo chaotyczny. Na początku II fali w październiku 2020 roku pracowałem jako lekarz-stażysta na oddziale internistycznym, gdzie przyjmowaliśmy tylko pacjentów z COVID-19. Pamiętam jak pewnego dnia moja koleżanka rezydentka załamała ręce, mówiąc „To przecież nie ma znaczenie co zrobimy, oni [pacjenci] i tak umrą”. Historia leczenia COVID-19 to jednak dość ciekawa lekcja medycyny opartej na dowodach naukowych. Pokazuje, jak w obliczu globalnego kryzysu przeszliśmy drogę od ostrożnych hipotez i mechanistycznego rozumowania do precyzyjnych wytycznych opartych na wielkich badaniach klinicznych. To historia o tym, jak to, co wydawało się słuszne, zostało zweryfikowane przez rygorystycznie przeprowadzone badania.

Na początku 2020 roku, gdy świat zmagał się z pierwszą falą pandemii, dominowało podejście zachowawcze. Eksperci z ramienia WHO odradzali rutynowe stosowanie glikokortykosteroidów.⁵⁸ Dlaczego? Ich argumentacja opierała się na dwóch filarach: **rozumowaniu opartym o mechanizmy** oraz **danych historycznych** z poprzednich epidemii. Sterydy to „broń obosieczna”. Z jednej strony hamują stan zapalny w płucach, co teoretycznie mogłoby ograniczyć groźne powikłanie, jakim jest ARDS (*acute respiratory distress syndrome*), ale z drugiej strony

– hamują odpowiedź immunologiczną i mogą utrudniać organizmowi usuwanie wirusa. Dane z epidemii SARS i MERS sugerowały, że sterydy opóźniały eliminację wirusa z organizmu. Co gorsza, w przypadku grypy sezonowej meta-analizy wykazywały **zwiększoną śmiertelność** u pacjentów leczonych sterydami. W tamtym momencie, nie mając twardych dowodów dla COVID-19, lekarze słusznie obawiali się powtórki tych tragicznych wyników.

Sytuacja zmieniła się diametralnie dzięki brytyjskiemu badaniu RECOVERY, które jest przykładem randomizowanego badania klinicznego. Badacze podeszli do problemu bez uprzedzeń: losowo przydzielili tysiące pacjentów do grup otrzymujących różne leczenie (w tym deksametazon – steryd) lub standardową opiekę.⁵⁹ Wyniki tego badania pokazały coś, czego nie dało się przewidzieć samym rozumowaniem teoretycznym: skuteczność sterydów zależała od stanu pacjenta.

- **Najciężej chorzy (pod respiratorami):** Śmiertelność spadła o około **jedną trzecią** (29,3% vs 41,4%; RR 0.64; 95% CI, 0.51–0.81).
- **Pacjenci wymagający tlenu (ale nie respiratora):** Śmiertelność spadła o około **jedną piątą** (23,3% vs 26,2%; RR 0.82; 95% CI, 0.72–0.94).
- **Lekkie przypadki (brak konieczności tlenoterapii):** Tutaj sterydy nie przyniosły korzyści, a wręcz istniał sygnał o możliwej szkodliwości (śmiertelność 17,8% vs 14,0%; RR 1.19; 95% CI, 0.92–1.55).

To badanie nauczyło nas, że w COVID-19 kluczowy jest moment podania leku. W późnej fazie choroby, gdy dominuje odpowiedź zapalna organizmu, sterydy ratują życie, mimo że na początku infekcji mogłyby przeszkadzać.

W podobnym czasie zespół WHO przeprowadził **meta-analizę**, łącząc dane z 7 różnych badań klinicznych przeprowadzonych w 12 krajach.⁶⁰

Dzięki połączeniu danych 1703 krytycznie chorych pacjentów, otrzymaliśmy potężny dowód: połączony **iloraz szans (OR) wyniósł 0,66**, a wynik był nieprzypadkowy. Oznacza to, że u pacjentów w stanie krytycznym podanie sterydów redukuje szansę na zgon średnio o 34%. Dzięki temu lekarze na całym świecie zyskali pewność, że sterydoterapia powinna być standardem opieki u najciężiej chorych na COVID-19.

Ta historia pokazuje istotę hierarchii dowodów EBM. Zaczęliśmy od uzasadnionego strachu opartego na patofizjologii (Krok 5 w hierarchii dowodów), ale dopiero rzetelne RCT (Krok 2) i meta-analiza (Krok 1) dały nam prawo do zmiany praktyki klinicznej.

Kiedy EBM jest źle używany

Polityka kontra nauka: Jak Donald Trump ogłosił, że paracetamol powoduje autyzmem

W dzisiejszych czasach lekarze muszą mierzyć się z nowym wyzwaniem – „epidemią dezinformacji”. Często nasi pacjenci przychodzą do gabinetu nie z pytaniem, ale z gotową tezą, którą usłyszeli w mediach społecznościowych lub od znanych polityków. Klasycznym przykładem tego zjawiska jest historia Donalda Trumpa i jego twierdzeń dotyczących paracetamolu (w USA znanego jako acetaminofen lub Tylenol – nazwa handlowa).⁶¹ Ten przykład dobrze ilustruje, jak nieostrożne konkluzje przekazane z dużą pewnością mogą przyczyniać się do szerzenia dezinformacji.

22. września 2025 roku Donald Trump, wspierany przez swojego sekretarza zdrowia Roberta F. Kennedy'ego Jr., publicznie odradził kobietom w ciąży przyjmowanie paracetamolu. Twierdził on, że lek ten wiąże się z „bardzo zwiększonym ryzykiem autyzmu”. Trump posunął się nawet do stwierdzenia, że Tylenol „nie jest dobry” i że nie powinno się go podawać dzieciom po urodzeniu ani po szczepieniach. Jako dowód

administracja Trumpa przytoczyła badanie obserwacyjne z 2020 roku opublikowane w *JAMA Psychiatry*, które wykazało wyższe prawdopodobieństwo wystąpienia ADHD i autyzmu u dzieci z wyższym stężeniem metabolitów paracetamolu w krwi pępowinowej przy porodzie.⁶² Autorzy badania wskazywali jednak na to, że czas półtrwania paracetamolu wynosi 3 godziny, więc stężenie w krwi pępowinowej odzwierciedla co najwyżej użycie paracetamolu w okresie okołoporodowym. Wskazali na brak możliwości kontroli niektórych czynników zakłócających – zwłaszcza genetycznych i środowiskowych, co ogranicza wnioskowanie przyczynowe. Administracja Trumpa odwołała się także do przeglądu systematycznego,⁶³ który wskazywał na możliwe powiązanie ekspozycji na paracetamol w życiu płodowym i wystąpienie zaburzeń ze spektrum autyzmu u dzieci. Niemniej, autorzy podkreślili w konkluzjach, że obecnie związek przyczynowy nie jest możliwy do ustalenia, a dalsze badania o bardziej rygorystycznej metodologii są potrzebne.

To bywa kuszące, żeby przeskakiwać do pewnych konkluzji na podstawie dowodów o ograniczonej jakości. Problem w tym, że politycy popełnili błąd, o którym wspominałem we wcześniejszych fragmentach skryptu. Na podstawie badania obserwacyjnego uznali związek między ekspozycją (paracetamolem) a wynikiem (autyzmem u dzieci) za przyczynowy i pewny, ignorując fakt, że badania, na które się powoływali, nie mogły potwierdzić takiego związku.

Eksperci natychmiast wytknęli błąd w tym rozumowaniu. Badania, na które powołał się Trump, były przeglądami systematycznymi badań obserwacyjnych lub pojedynczymi badaniami obserwacyjnymi. Ich interpretacja wymaga ostrożności ze względu na występowanie potencjalnych czynników zakłócających. Przykładowo, kobiety cierpiące na hipermobilność stawów częściej potrzebują paracetamolu na ból,

a jednocześnie to właśnie te uwarunkowania genetyczne są powiązane z częstszym występowaniem autyzmu u dzieci. Innymi słowy: nie lek musi być odpowiedzialny za rozwój autyzmu, ale powód, dla którego matka brała lek (uwarunkowania genetyczne, infekcje, przewlekły ból) mógł być prawdziwym czynnikiem ryzyka.

Aby uciąć spekulacje, badacze ze Szwecji przeprowadzili gigantyczne, wysokiej jakości badanie na populacji **prawie 2,5 miliona dzieci** urodzonych w latach 1995–2019.⁶⁴ Zastosowali oni finezyjny rodzaj badania: **analizę kontrolną rodzeństwa** (*sibling control analysis*).⁶⁵ Co to znaczy? Porównując dwoje dzieci tej samej matki – z których jedno było narażone na paracetamol w ciąży, a drugie nie – badacze automatycznie „wyeliminowali” wpływ wspólnych genów i środowiska domowego.

Wyniki były jednoznaczne:

- W zwykłych modelach (bez kontroli rodzeństwa) paracetamol wydawał się nieznacznie zwiększać ryzyko autyzmu (HR 1.05).
- Jednak, gdy naukowcy spojrzeli na pary rodzeństwa, związek ten całkowicie zniknął.
- Badanie nie wykazało również żadnej zależności od dawki – niezależnie od tego, czy matka brała małe, czy duże ilości leku, ryzyko autyzmu u dzieci było takie samo.

Badanie to dostarczyło silnego dowodu, że wcześniejsze doniesienia o szkodliwości paracetamolu były wynikiem błędów systematycznych, a nie rzeczywistego działania leku. Straszanie pacjentek „toksycznym paracetamolem” niesie realne zagrożenie. Nieleczony ból i gorączka w trakcie ciąży mogą być znacznie groźniejsze dla płodu niż bezpieczne dawki leku przeciwbólowego. WHO, NHS oraz inne agencje medyczne nadal uznają paracetamol za **lek pierwszego wyboru** u ciężarnych. Pamiętaj, by uważnie przyjrzeć

się dowodom, kiedy politycy powołują się na dane naukowe i doprowadzają do burzy medialnej.

Skandal Vioxx: Lekcja o tym, czego nie widać na pierwszy rzut oka

Wyobraź sobie, że w 1999 roku na rynek wchodzi „cudowny lek” – rofecoxib, znany szerzej jako Vioxx. Lek miał uśmierzać ból tak skutecznie jak klasyczne niesteroidowe leki przeciwzapalne (np. naproksen czy ibuprofen), ale bez ich największej wady – uszkodzenia błony śluzowej żołądka dzięki wybiórczemu działaniu na cyklooksygenazę typu 2 (COX-2). Przez kilka lat wielu pacjentów przyjmowały ten preparat, nie wiedząc, że biorą udział w jednym z największych skandali w historii współczesnej medycyny. Historia Vioxx to przypomnienie, że badanie randomizowane, choć jest złotym standardem, może zostać zaprojektowane tak, aby wyeksponować korzyści, jednocześnie ukrywając rzadkie, ale śmiertelne zagrożenia.⁶⁶

Kluczowym dowodem na skuteczność Vioxx miało być badanie **VIGOR**, opublikowane w prestiżowym czasopiśmie *New England Journal of Medicine*.⁶⁷ Objęło ono ponad 8000 pacjentów z reumatoidalnym zapaleniem stawów. Wyniki w zakresie przewodu pokarmowego były imponujące. Vioxx zmniejszył ryzyko poważnych zdarzeń żołądkowo-jelitowych (takich jak perforacje czy krwawienia) o **połowę** (RR 0,5) w porównaniu z naproksenem, czyli dotychczas rutynowo stosowanym lekiem.

Zauważono jednak, że zawały mięśnia sercowego występowały rzadko, ale częściej u pacjentów biorących Vioxx (0,4%) niż naproksen (0,1%). Zamiast uczciwie przyznać, że Vioxx może szkodzić sercu, autorzy i producent (firma Merck) sformułowali tzw. „hipotezę naproksenu”. Twierdzili, że to nie Vioxx jest niebezpieczny, ale naproksen posiada wyjątkowe właściwości ochronne dla serca (podobnie jak aspiryna),

co miało rzekomo tłumaczyć różnicę w wynikach. Nie było na to jednak żadnych twardych dowodów.

Dopiero po latach, dzięki procesom sądowym, światło dzienne ujrzały wewnętrzne dokumenty, które pokazały, jak manipulowano danymi:

- W publikacji VIGOR zliczano zdarzenia żołądkowo-jelitowe o miesiąc dłużej niż zdarzenia sercowo-naczyniowe. Ten „zabieg” pozwolił pominąć trzy dodatkowe zawały serca w grupie Vioxx, które wystąpiły pod koniec badania. Gdyby te trzy przypadki zostały uwzględnione, ryzyko sercowo-naczyniowe stałoby się jeszcze bardziej ewidentne.
- Komisja monitorująca bezpieczeństwo danych powinna być niezależna, tymczasem jej przewodniczący podpisał kontrakt doradczy z firmą Merck na dwa tygodnie przed zakończeniem badania.

Badania randomizowane są projektowane, by wykazać skuteczność leczenia. W związku z tym informacje dotyczące rzadkich szkód mogą nie ujrzeć od razu światła dziennego. W badaniu VIGOR wyłapano sygnał o szkodliwości Vioxxu, jednak liczby były na tyle małe, że nie zwrócono na nie początkowo należytej uwagi. Aby udowodnić, że lek pomaga na ból, wystarczy kilkuset pacjentów. Jednak aby wykryć rzadkie działanie niepożądane (np. zawał serca u 1 na 200 osób), potrzeba tysięcy badanych. Dodatkowym wyzwaniem jest to, że okres obserwacyjny w badaniu z randomizacją nie może trwać w nieskończoność, a w praktyce rzadko trwa dłużej niż rok. W związku z tym szkody oddalone w czasie mogą być niewykryte lub tylko częściowo uwzględnione.

Vioxx został wycofany z rynku w 2004 roku, po tym jak oszacowano, że mógł przyczynić się do dziesiątek tysięcy nadmiarowych zgonów z przyczyn sercowych. Historia ta uczy nas uważności w analizie raportowanych działań niepożądanych. Jako lekarz, zawsze pytaj

nie tylko o to, czy lek działa, ale też: „Jak długo trwało badanie?” oraz „Czy grupa była wystarczająco duża, by wykryć rzadkie powikłania?”. W medycynie opartej na dowodach, brak dowodu na szkodliwość nie jest tym samym, co dowód na bezpieczeństwo.

Więcej o EBM

W tym skrypcie omówiłem podstawowe informacje dotyczące EBM. Pomogą Ci one, drogi Czytelniku, znajdować oraz interpretować badania, mogące mieć wpływ na Twoją praktykę kliniczną. Niemniej, EBM to znacznie więcej niż to, co przedstawiłem dotychczas. Omówienie bardziej zaawansowanych zagadnień wykracza poza zakres tego skryptu, jednak poniżej zarysuję kilka istotnych elementów.

Ocena jakości i raportowanie badań

Nie byłoby EBM-u, gdyby nie naukowcy, którzy prowadzą badania. Każdy z rodzajów badań powinien być raportowany w standardowy sposób. Pionierem w tym zakresie był wspomniany wcześniej Doug Altman, statystyk z Oxfordu, który stworzył [Equator Network](#) w 2006 roku.⁶⁸ Jego celem było ujednoczenie prezentacji wyników poszczególnych badań, co umożliwiłoby ich należytą ocenę jakościową oraz systematyczne porównywanie wyników. Jeżeli kiedyś postanowisz, drogi Czytelniku, sam przeprowadzić badanie, pamiętaj, żeby skorzystać z tych wytycznych raportowania.

Jak wspomniałem wcześniej, kluczowym elementem przeglądu systematycznego jest ocena jakości (*risk of bias assessment*) poszczególnych badań, włączonych do przeglądu. W tym celu używa się konkretnych narzędzi metodologicznych. Są one przydatne również poza

kontekstem pisania przeglądu systematycznego – jako pomoce we wstępnej ocenie jakościowej danego badania. W tym zakresie polecam strony internetowe [Joanna Briggs Institute](#) oraz [Oxford Centre for Evidence-Based Medicine](#), gdzie znajdziesz podstawowe narzędzia do oceny jakości poszczególnych rodzajów badań. Obie te strony są też dobrymi źródłami do poszerzania wiedzy na temat rodzajów badań czy EBM w ogóle.

Nowe kierunki rozwoju

Dotychczas omawiałem klasyczne modele badań, które używane są w EBM. Niemniej, w ostatnich latach rozwój technologii oraz usprawniony dostęp do rejestrów chorych umożliwiły prowadzenie badań na szerszą skalę i w sposób do tej pory niepraktykowany.

Pierwszym przykładem są badania oparte na wielkich ilościach danych (*big data*).⁶⁹ Polegają one na wykorzystywaniu gigantycznych zbiorów danych medycznych do wspierania decyzji klinicznych, wykrywania zależności i tworzenia bardziej precyzyjnych zaleceń terapeutycznych. Doskonałym przykładem jest badanie, w którym analizowano ogromny zbiór danych celem znalezienia zależności pomiędzy allelami poszczególnych genów a podatnością na zachorowanie na poszczególne choroby psychiatryczne.⁷⁰ Na podstawie ponad miliona przypadków wyróżniono 5 klastrów chorób czy zaburzeń psychicznych, dla których zidentyfikowano grupy alleli, biorących udział w podatności na zachorowanie.

Drugim przykładem są badania z rzeczywistej praktyki klinicznej (*real-world evidence*).⁷¹ Mają one na celu ocenę, jak leczenie, diagnostyka lub interwencje medyczne działają w rzeczywistej praktyce klinicznej, a nie tylko w ściśle kontrolowanych warunkach badań klinicznych. Nieraz okazuje się, że do badań randomizowanych włączona może być tylko

niewielka pula pacjentów, którzy mają kontakt z ochroną zdrowia.⁷² Poza tym badania te są ograniczone w czasie i mogą mieć mało elastyczne schematy dawkowania leków. Żeby zyskać szerszą perspektywę na skuteczność czy bezpieczeństwo interwencji w warunkach rzeczywistej praktyki klinicznej można użyć w sposób systematyczny danych z rejestrów medycznych. Przykładem jest praca analizująca dane ponad 2000 pacjentów, żeby określić, czy stosowanie leków na ADHD jest powiązane z pogorszeniem stanu psychicznego czy rehospitalizacją u pacjentów z zaburzeniami psychiatrycznymi.⁷³ Badanie to wykorzystywało dane z rejestrów zdrowotnych i obserwację nieselekcjonowanej populacji pacjentów leczonych w normalnych warunkach systemu ochrony zdrowia, a nie w kontrolowanym badaniu randomizowanym. Dzięki niemu pokazano, że leki na ADHD mogą być bezpieczniejsze wśród pacjentów z wcześniejszym rozpoznaniem psychozy, niż uprzednio sądzono. Należy jednak pamiętać, że badania opierające się na danych rejestrowych są w istocie badaniami obserwacyjnymi, a w interpretacji ich wyników należy brać pod uwagę ograniczenia tego rodzaju badań, w tym możliwości występowania niezmiierzonych zmiennych zakłócających.

Ostatnim przykładem (choć jest ich znacznie więcej) jest zastosowanie metod mieszanych (*mixed-methods research*).⁷⁴ Podejście to polega na zbieraniu danych ilościowych i jakościowych w taki sposób, by dane uzyskane w różny sposób wzajemnie się dopełniały. Przykładem jest badanie zaburzeń pamięci autobiograficznej (wspomnień ze swojego życia) wśród pacjentów poddanych terapii elektrowstrząsowej z powodu depresji lekoopornej.⁷⁵ Istnieją wprawdzie skale służące do oceny pamięci, jednak są one ograniczone do konkretnych wydarzeń, o które pacjent pytany jest przed i po terapii. Okazuje się jednak, że pacjenci mogą tracić również wspomnienia nieuwzględnione w kwestionariuszach ilościowych. Ponadto utrata pamięci może wpływać na życie pacjentów

i ich bliskich w sposób, którego testy oparte wyłącznie na wynikach liczbowych nie są w stanie uchwycić. Możliwe jest zatem połączenie danych pochodzących z kwestionariuszy ilościowych z jakościową analizą doświadczeń pacjentów, aby ocenić, czy ich przeżycia odpowiadają uzyskanym wynikom liczbowym. Takie podejście pozwala lepiej zrozumieć doświadczenia pacjentów oraz może wpływać na organizację opieki po hospitalizacji i sposób informowania pacjentów o potencjalnych działaniach niepożądanych leczenia.

Podziękowania

Serdecznie dziękuję dr hab. Piotrowi Dziechciarzowi, dr. Wiktorowi Pascalowi, lek. Łukaszowi Banaszкови, lek. Piotrowi Krysiakowi, lek. Agnieszce Kuc, lek. Herbertowi Phamowi, Julii Stachurze i Marcie Wojtkiewicz za cenne uwagi redakcyjne i merytoryczne oraz wsparcie w przygotowaniu skryptu. Dziękuję również prof. Hannie Szajewskiej oraz prof. Mariuszowi Panczykowi za recenzję skryptu. Notebook LM był użyty do generowania części tekstu skryptu, jednak koncepcja, dobór artykułów źródłowych oraz ostateczna redakcja tekstu w pełni należą do autora.

Piśmiennictwo

- 1 Chalmers, I. & Glasziou, P. Avoidable waste in the production and reporting of research evidence. *Lancet* **374**, 86-89 (2009). [https://doi.org/10.1016/s0140-6736\(09\)60329-9](https://doi.org/10.1016/s0140-6736(09)60329-9)
- 2 Altman, D.G. The scandal of poor medical research. *Bmj* **308**, 283-284 (1994). <https://doi.org/10.1136/bmj.308.6924.283>
- 3 Ustawa o zawodach lekarza i lekarza dentystry. Dostępna pod adresem: <https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20240001287/U/D20241287Lj.pdf>.
- 4 Lind, J. *A treatise on the scurvy: in three parts. Containing an inquiry into the nature, causes, and cure, of that disease. Together with a critical and chronological view of what has been published on the subject.* (S. Crowder, 1772).
- 5 Evidence-based medicine. A new approach to teaching the practice of medicine. *Jama* **268**, 2420-2425 (1992). <https://doi.org/10.1001/jama.1992.03490170092032>
- 6 Garratt, C., Ward, D.E. & Camm, A.J. Lessons from the cardiac arrhythmia suppression trial. *Bmj* **299**, 805-806 (1989). <https://doi.org/10.1136/bmj.299.6703.805>
- 7 Anderson, J.L. *et al.* Oral flecainide acetate for the treatment of ventricular arrhythmias. *N Engl J Med* **305**, 473-477 (1981). <https://doi.org/10.1056/nejm198108273050901>
- 8 Echt, D.S. *et al.* Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* **324**, 781-788 (1991). <https://doi.org/10.1056/nejm199103213241201>
- 9 Friedman, S.H., Reed, E. & Ross, N.E. Postpartum Psychosis. *Curr Psychiatry Rep* **25**, 65-72 (2023). <https://doi.org/10.1007/s11920-022-01406-4>

- 10 Forde, R., Peters, S. & Wittkowski, A. Recovery from postpartum psychosis: a systematic review and metasynthesis of women's and families' experiences. *Archives of Women's Mental Health* **23**, 597-612 (2020). <https://doi.org/10.1007/s00737-020-01025-z>
- 11 Hopewell, S., Loudon, K., Clarke, M.J., Oxman, A.D. & Dickersin, K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* **2009**, Mr000006 (2009). <https://doi.org/10.1002/14651858.MR000006.pub3>
- 12 Bschor, T., Nagel, L., Unger, J., Schwarzer, G. & Baethge, C. Differential Outcomes of Placebo Treatment Across 9 Psychiatric Disorders: A Systematic Review and Meta-Analysis. *JAMA Psychiatry* **81**, 757-768 (2024). <https://doi.org/10.1001/jamapsychiatry.2024.0994>
- 13 Cipriani, A. *et al.* Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* **391**, 1357-1366 (2018). [https://doi.org/10.1016/s0140-6736\(17\)32802-7](https://doi.org/10.1016/s0140-6736(17)32802-7)
- 14 Furukawa, T.A. Review: placebo is better than no treatment for subjective continuous outcomes and for treatment of pain. *Evid Based Ment Health* **5**, 15 (2002). <https://doi.org/10.1136/ebmh.5.1.15>
- 15 Murad, M.H., Ramachandran, N., Lin, L. & Li, F. Regression to the mean: a primer for evidence-based medicine practitioners. *BMJ Evidence-Based Medicine*, bmjebm-2025-113948 (2025). <https://doi.org/10.1136/bmjebm-2025-113948>
- 16 Hymes, K.B. *et al.* Kaposi's sarcoma in homosexual men—a report of eight cases. *Lancet* **2**, 598-600 (1981). [https://doi.org/10.1016/s0140-6736\(81\)92740-9](https://doi.org/10.1016/s0140-6736(81)92740-9)
- 17 McNamee, R. Confounding and confounders. *Occupational and Environmental Medicine* **60**, 227 (2003). <https://doi.org/10.1136/oem.60.3.227>
- 18 Pourhoseingholi, M.A., Baghestani, A.R. & Vahedi, M. How to control confounding effects by statistical analysis. *Gastroenterol Hepatol Bed Bench* **5**, 79-83 (2012).
- 19 Roberts, C. & Torgerson, D. Randomisation methods in controlled trials. *BMJ* **317**, 1301-1310 (1998). <https://doi.org/10.1136/bmj.317.7168.1301>
- 20 Hróbjartsson, A., Emanuelsson, F., Skou Thomsen, A.S., Hilden, J. & Brorson, S. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. *Int J Epidemiol* **43**, 1272-1283 (2014). <https://doi.org/10.1093/ije/dyu115>

- 21 OCEBM Levels of Evidence Working Group*. “The Oxford Levels of Evidence 2”. Oxford Centre for Evidence-Based Medicine.
<https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebmllevels-of-evidence> * OCEBM Levels of Evidence Working Group = Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary Hodgkinson.
- 22 Sedgwick, P. What is an “n-of-1” trial? *BMJ : British Medical Journal* **348**, g2674 (2014). <https://doi.org/10.1136/bmj.g2674>
- 23 Greenhalgh, T. How to read a paper : getting your bearings (deciding what the paper is about). *BMJ* **315**, 243-246 (1997).
<https://doi.org/10.1136/bmj.315.7102.243>
- 24 McBride, W.G. Thalidomide and congenital abnormalities. *Lancet* **2**, 90927-90928 (1961).
- 25 Lenz, W. in *Problems of Birth Defects: From Hippocrates to Thalidomide and After 199-199* (Springer, 1962).
- 26 Greenhalgh, T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* **315**, 672-675 (1997).
<https://doi.org/10.1136/bmj.315.7109.672>
- 27 Greenhalgh, T. & Taylor, R. How to read a paper: Papers that go beyond numbers (qualitative research). *BMJ* **315**, 740-743 (1997).
<https://doi.org/10.1136/bmj.315.7110.740>
- 28 Dodgson, J.E. Reflexivity in Qualitative Research. *J Hum Lact* **35**, 220-222 (2019). <https://doi.org/10.1177/0890334419830990>
- 29 Braun, V. & Clarke, V. Using thematic analysis in psychology. *Qualitative research in psychology* **3**, 77-101 (2006).
- 30 Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J. & Neville, A.J. The use of triangulation in qualitative research. *Oncol Nurs Forum* **41**, 545-547 (2014).
<https://doi.org/10.1188/14.Onf.545-547>
- 31 Drisko, J.W. Transferability and Generalization in Qualitative Research. *Research on Social Work Practice* **35**, 102-110 (2025).
<https://doi.org/10.1177/10497315241256560>
- 32 Greenhalgh, T. How to read a paper: Papers that report drug trials. *BMJ* **315**, 480-483 (1997). <https://doi.org/10.1136/bmj.315.7106.480>
- 33 Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. Concorde Coordinating Committee. *Lancet* **343**, 871-881 (1994).

- 34 Greenhalgh, T. How to read a paper: Assessing the methodological quality of published papers. *BMJ* **315**, 305-308 (1997).
<https://doi.org/10.1136/bmj.315.7103.305>
- 35 Porta, M.S., Greenland, S., Hernán, M., dos Santos Silva, I. & Last, J.M. *A dictionary of epidemiology*. (Oxford university press, 2014).
- 36 <https://catalogofbias.org>.
- 37 Hermes-DeSantis, E.R. *et al.* Preferences for Accessing Medical Information in the Digital Age: Health Care Professional Survey. *J Med Internet Res* **23**, e25868 (2021). <https://doi.org/10.2196/25868>
- 38 Savović, J. *et al.* Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* **157**, 429-438 (2012). <https://doi.org/10.7326/0003-4819-157-6-201209180-00537>
- 39 Sterne, J.A. & Egger, M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* **54**, 1046-1055 (2001).
[https://doi.org/10.1016/s0895-4356\(01\)00377-8](https://doi.org/10.1016/s0895-4356(01)00377-8)
- 40 Zhang, Z., Xu, X. & Ni, H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. *Crit Care* **17**, R2 (2013). <https://doi.org/10.1186/cc11919>
- 41 Solmi, M. *et al.* Antipsychotic Use and Risk of Breast Cancer in Women With Severe Mental Illness: Replication of a Nationwide Nested Case-Control Database Study. *Schizophr Bull* **50**, 1471-1481 (2024).
<https://doi.org/10.1093/schbul/sbae058>
- 42 Lin, X. *et al.* Antipsychotic-Related Prolactin Changes: A Systematic Review and Dose-Response Meta-analysis. *CNS Drugs* **39**, 937-947 (2025).
<https://doi.org/10.1007/s40263-025-01218-z>
- 43 Gilden, J. *et al.* Long-Term Outcomes of Postpartum Psychosis: A Systematic Review and Meta-Analysis. *J Clin Psychiatry* **81** (2020).
<https://doi.org/10.4088/JCP.19r12906>
- 44 Naddaf, M. & Quill, E. Hallucinated citations are polluting the scientific literature. What can be done? *Nature* **652**, 26-29 (2026).
- 45 Deo, S.V., Deo, V. & Sundaram, V. Survival analysis-part 2: Cox proportional hazards model. *Indian J Thorac Cardiovasc Surg* **37**, 229-233 (2021).
<https://doi.org/10.1007/s12055-020-01108-7>
- 46 Greenhalgh, T. How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests. *BMJ* **315**, 364-366 (1997).
<https://doi.org/10.1136/bmj.315.7104.364>

- 47 Greenhalgh, T. How to read a paper: Statistics for the non-statistician. II: "Significant" relations and their pitfalls. *BMJ* **315**, 422-425 (1997). <https://doi.org/10.1136/bmj.315.7105.422>
- 48 Sainani, K.L. Clinical versus statistical significance. *Pm r* **4**, 442-445 (2012). <https://doi.org/10.1016/j.pmrj.2012.04.014>
- 49 Lee, I.M., Djoussé, L., Sesso, H.D., Wang, L. & Buring, J.E. Physical activity and weight gain prevention. *Jama* **303**, 1173-1179 (2010). <https://doi.org/10.1001/jama.2010.312>
- 50 French, M. T., Popovici, I. & Maclean, J.C. Do alcohol consumers exercise more? Findings from a national survey. *Am J Health Promot* **24**, 2-10 (2009). <https://doi.org/10.4278/ajhp.0801104>
- 51 Di Ciccio, M.E., Ragazzo, V. & Jacinto, T. Mortality in relation to smoking: the British Doctors Study. *Breathe (Sheff)* **12**, 275-276 (2016). <https://doi.org/10.1183/20734735.013416>
- 52 Doll, R. & Hill, A.B. Smoking and carcinoma of the lung; preliminary report. *Br Med J* **2**, 739-748 (1950). <https://doi.org/10.1136/bmj.2.4682.739>
- 53 Doll, R. & Hill, A.B. The mortality of doctors in relation to their smoking habits; a preliminary report. *Br Med J* **1**, 1451-1455 (1954). <https://doi.org/10.1136/bmj.1.4877.1451>
- 54 Doll, R. & Hill, A. B. MORTALITY IN RELATION TO SMOKING: TEN YEARS' OBSERVATIONS OF BRITISH DOCTORS. *Br Med J* **1**, 1460-1467 concl (1964). <https://doi.org/10.1136/bmj.1.5396.1460>
- 55 Doll, R. & Peto, R. Mortality in relation to smoking: 20 years' observations on male British doctors. *Br Med J* **2**, 1525-1536 (1976). <https://doi.org/10.1136/bmj.2.6051.1525>
- 56 Doll, R., Peto, R., Wheatley, K., Gray, R. & Sutherland, I. Mortality in relation to smoking: 40 years' observations on male British doctors. *Bmj* **309**, 901-911 (1994). <https://doi.org/10.1136/bmj.309.6959.901>
- 57 Doll, R., Peto, R., Boreham, J. & Sutherland, I. Mortality in relation to smoking: 50 years' observations on male British doctors. *Bmj* **328**, 1519 (2004). <https://doi.org/10.1136/bmj.38142.554479.AE>
- 58 Russell, C.D., Millar, J.E. & Baillie, J.K. Clinical evidence does not support corticosteroid treatment for 2019-nCoV lung injury. *Lancet* **395**, 473-475 (2020). [https://doi.org/10.1016/s0140-6736\(20\)30317-2](https://doi.org/10.1016/s0140-6736(20)30317-2)
- 59 Horby, P. *et al.* Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* **384**, 693-704 (2021). <https://doi.org/10.1056/NEJMoa2021436>

- 60 Sterne, J.A.C. *et al.* Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. *Jama* **324**, 1330-1341 (2020). <https://doi.org/10.1001/jama.2020.17023>
- 61 Lang, K. Trump's claims on Tylenol (paracetamol), vaccines, and autism-what's the truth? *Bmj* **390**, r2025 (2025). <https://doi.org/10.1136/bmj.r2025>
- 62 Ji, Y. *et al.* Association of Cord Plasma Biomarkers of In Utero Acetaminophen Exposure With Risk of Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in Childhood. *JAMA Psychiatry* **77**, 180-189 (2020). <https://doi.org/10.1001/jamapsychiatry.2019.3259>
- 63 Prada, D., Ritz, B., Bauer, A.Z. & Baccarelli, A.A. Evaluation of the evidence on acetaminophen use and neurodevelopmental disorders using the Navigation Guide methodology. *Environ Health* **24**, 56 (2025). <https://doi.org/10.1186/s12940-025-01208-0>
- 64 Ahlqvist, V.H. *et al.* Acetaminophen Use During Pregnancy and Children's Risk of Autism, ADHD, and Intellectual Disability. *Jama* **331**, 1205-1214 (2024). <https://doi.org/10.1001/jama.2024.3172>
- 65 Gustavson, K., Torvik, F.A., Davey Smith, G., Røysamb, E. & Eilertsen, E.M. Familial confounding or measurement error? How to interpret findings from sibling and co-twin control studies. *Eur J Epidemiol* **39**, 587-603 (2024). <https://doi.org/10.1007/s10654-024-01132-6>
- 66 Krumholz, H.M., Ross, J.S., Presler, A.H. & Egilman, D.S. What have we learnt from Vioxx? *BMJ* **334**, 120-123 (2007). <https://doi.org/10.1136/bmj.39024.487720.68>
- 67 Bombardier, C. *et al.* Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N Engl J Med* **343**, 1520-1528, 1522 p following 1528 (2000). <https://doi.org/10.1056/nejm200011233432103>
- 68 Altman, D.G. & Simera, I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J R Soc Med* **109**, 67-77 (2016). <https://doi.org/10.1177/0141076815625599>
- 69 Batko, K. & Ślęzak, A. The use of Big Data Analytics in healthcare. *J Big Data* **9**, 3 (2022). <https://doi.org/10.1186/s40537-021-00553-4>
- 70 Grotzinger, A.D. *et al.* Mapping the genetic landscape across 14 psychiatric disorders. *Nature* **649**, 406-415 (2026). <https://doi.org/10.1038/s41586-025-09820-3>
- 71 Liu, M., Qi, Y., Wang, W. & Sun, X. Toward a better understanding about real-world evidence. *Eur J Hosp Pharm* **29**, 8-11 (2022). <https://doi.org/10.1136/ejhpharm-2021-003081>

- 72 Azoulay, L. Rationale, Strengths, and Limitations of Real-World Evidence in Oncology: A Canadian Review and Perspective. *Oncologist* **27**, e731-e738 (2022). <https://doi.org/10.1093/oncolo/oyac114>
- 73 Corbeil, O. *et al.* Treatment with psychostimulants and atomoxetine in people with psychotic disorders: reassessing the risk of clinical deterioration in a real-world setting. *The British Journal of Psychiatry* **224**, 98-105 (2024). <https://doi.org/10.1192/bjp.2023.149>
- 74 Tariq, S. & Woodman, J. Using mixed methods in health research. *JRSM Short Rep* **4**, 2042533313479197 (2013). <https://doi.org/10.1177/2042533313479197>
- 75 Stachura, A. Retrograde autobiographical amnesia following electroconvulsive therapy (ECT) in patients treated for depression-a mixed-methods systematic review and meta-analysis. *Master's Thesis* (2024).

Spis rycin i tabel

Ryciny:

Rycina 1. Liczba opublikowanych badań (627 229) na temat depresji w bazie danych PubMed w latach 1980 - 2025 (stan na 26.03.2025 r.).....	6
Rycina 2. Elementy EBM.....	8
Rycina 3. 5 kroków medycyny opartej na dowodach naukowych	19
Rycina 4. Schemat badania obserwacyjnego (kohortowego)	27
Rycina 5. Schemat randomizowanego badania klinicznego.....	32
Rycina 6. Wynik wyszukiwania w PubMed (maj 2025).....	56
Rycina 7. Widok ze strony internetowej biblioteki uczelnianej, z której można uzyskać dostęp do UpToDate.....	57
Rycina 8. Wynik wyszukiwania w UpToDate (maj 2025)	58
Rycina 9. Zalecane leczenie psychozy poporodowej wg. UpToDate	59
Rycina 10. Podstawa dowodowa leczenia psychozy poporodowej wg. UpToDate ...	60
Rycina 11. Tabela działań niepożądanych leków przeciwpsychotycznych na podstawie UpToDate.	64
Rycina 12. Ekran startowy PubMed	65
Rycina 13. Zaawansowane opcje wyszukiwania PubMed.....	66
Rycina 14. Komenda do wyszukiwania badania.	66
Rycina 15. Wyniki wyszukiwania w PubMed	67
Rycina 16. Widok artykułu w PubMed	68
Rycina 17. Abstrakt artykułu.	69
Rycina 18. Wyszukiwanie danych dot. prognozy w UpToDate.	71
Rycina 19. Informacje nt. przebiegu psychozy poporodowej (UpToDate).	71
Rycina 20. Artykuł dot. prognozy pacjentek z psychozą poporodową.....	72
Rycina 21. Abstrakt badania kohortowego, badającego związek między poporodowymi zaburzeniami psychicznymi z wystąpieniem choroby afektywnej dwubiegunowej.....	73

Tabele:

Tabela 1. Hierarchia dowodów Oxfordzkiego Centrum Medycyny Opartej na Dowodach	37
Tabela 2. Siła dowodu i trafność zewnętrzna dowodów na skuteczność leków na psychozę poporodową.	61