



dr hab. Paweł Zagożdżon

Kierownik Zakładu Higieny i Epidemiologii GUMed. W pracy naukowej zajmuje się zagadnieniami epidemiologii klinicznej i społecznej. Tytuł doktora habilitowanego nauk medycznych uzyskał w 2011 r. na podstawie badań nad przeżywalnością i jakością życia osób bezrobotnych. Posiada specjalizację z psychiatrii i chorób wewnętrznych.

Jak czytać teksty naukowe?

Zdrowy rozsądek i czytanie ze zrozumieniem

Wagę tekstu naukowego należy oceniać w zależności od tego, jaką przyjmujemy perspektywę w takiej ocenie. Są dwie możliwości oceny tekstu naukowego w medycynie. Pierwsza to ta, dla której podstawowym kryterium jest zdolność do udowodnienia postulowanej hipotezy. Dla drugiej kryterium podstawowym oceny badania jest jego wkład w odkrywanie nowych mechanizmów, zjawisk i etiologii chorób. W tabeli 1 zaprezentowano (za Vandenbrouke, 2008) hierarchię ważności różnych typów badań w zależności od zdolności schematu badawczego do udowodniania związku przyczynowo-skutkowego lub do odkrywania nowych zjawisk.

Dlaczego warto sięgać po tekst, który ma charakter naukowego doniesienia? Pierwszym i jedynym powodem lektury powinno być dla Ciebie to czy tytuł wydaje Ci się interesujący i czy dotyczy istotnych dla Ciebie zagadnień.

Jak widać w kontekście badań ukierunkowanych na odkrywanie nowych mechanizmów i przyczyn chorób, hierarchia schematów badawczych jest odwrotna od tej uszeregowanej pod względem zdolności dowodzenia związków przyczynowo-skutkowych. W niniejszym rozdziale będziemy się zajmować jakością artykułów z punktu widzenia ich zdolności do dowodzenia związków przyczynowo-skutkowych i ich przydatności w praktyce klinicznej. W związku z tym badaniami o najwyższej wiarygodności naukowej są randomizowane badania kliniczne. Należy sobie jed-

Tab. 1. Hierarchia schematów badawczych

I. Hierarchia schematów badawczych ze względu na możliwość udowodnienia.

1. Randomizowane badania kliniczne;
2. Prospektywne badania kohortowe;
3. Retrospektywne badania kohortowe;
4. Badania kliniczno-kontrolne (*case-control*);
5. Opisy przypadków, badania ekologiczne i przekrojowe.

Dotyczy sytuacji, w której prawdopodobieństwo *a priori* weryfikowanej hipotezy jest duże (50:50).

II. Hierarchia schematów badawczych ze względu na możliwość odkrywania nowych zjawisk.

1. Opisy przypadków, przegląd danych literaturowych;
2. Badania kliniczno-kontrolne (*case-control*);
3. Retrospektywne badania kohortowe;
4. Prospektywne badania kohortowe;
5. Randomizowane badania kliniczne.

Dotyczy sytuacji, w której prawdopodobieństwo *a priori* weryfikowanej hipotezy jest bardzo małe (1:100-1:100 000?).

nak zdawać sprawę, że duża część literatury naukowej funkcjonuje w obrębie zagadnień, które opisują nowe zjawiska, skupiając się na próbach odkrycia nowych możliwości leczenia i niepoznanych jeszcze przyczyn i mechanizmach chorób. Początkiem programu badawczego zatem są badania będące opisami przypadków, następne w kolejności są badania obserwacyjne (np. typu *case-control* i kohortowe) i dopiero na końcu następuje weryfikacja badanej zależności w trybie badania eksperymentalnego u pacjentów, o ile oczywiście jest to możliwe z punktu widzenia etycznego. Badanie potencjalnie korzystnych efektów interwencji musi być zweryfikowane eksperymentalnie randomizowanymi kontrolowanymi badaniami klinicznymi. Badanie narażeń o potencjalnie szkodliwym charakterze nie może być realizowane eksperymentalnie i pozostaje nam jedynie ocena badań obserwacyjnych.

Odkrycie nowego genu związanego przyczynowo z ważną chorobą jest z pewnością sukcesem badawczym. Jednocześnie jednak praktyczne znaczenie takiego faktu w postępowaniu z pacjentami będzie niewielkie do momentu kiedy uda się zebrać więcej dowodów na ewentualne skuteczne modyfikowanie przebiegu choroby dzięki takiej wiedzy. Twoje zainte-

resowanie problemem jako czytelnika będzie się różnić w zależności od tego czy uznasz, że warto wykorzystać wnioski z artykułu w praktyce klinicznej, czy też stwierdzisz, że warto poszerzyć swoją wiedzę o danym problemie badawczym. A może po prostu artykuł wyda Ci się ciekawy przez to, że podejmuje kontrowersyjne zagadnienie albo sam tytuł został sformułowany w sposób intrygujący?

Formułowanie pytań klinicznych

Wskazówka kluczowa:

Na jakie pytanie autorzy artykułu starają się odpowiedzieć?

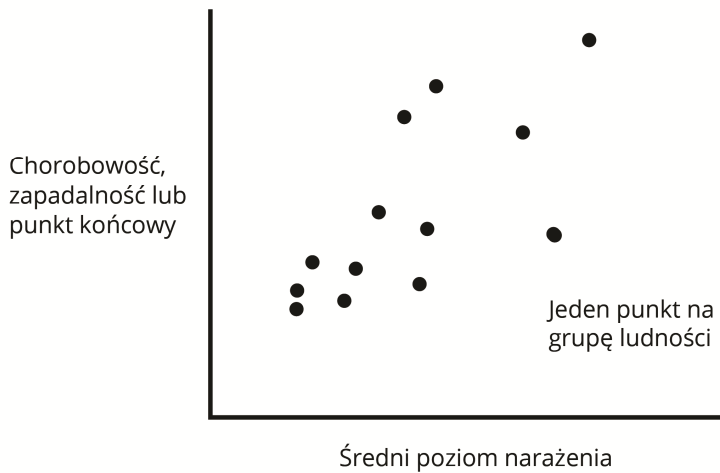
Początkiem lektury naukowego tekstu jest umiejscowienie tematyki na mapie naszej wiedzy. Wstęp i część wprowadzająca artykułu powinna pomóc czytelnikowi zrozumieć potrzebę badania i krótko opisać stan wiedzy w danej dziedzinie. Problem badawczy w artykule powinien być sformułowany w taki sposób, aby dało się łatwo ustalić cztery elementy, które prosto identyfikuje akronim w języku angielskim – PICO (*Patient Intervention Comparison Outcome*).

- *PATIENT* – w jakiej populacji pacjentów prowadzono badanie?
- *INTERVENTION/EXPOSURE* – jaka była interwencja/narażenie?
- *COMPARISON* – jaka była grupa kontrolna?
- *OUTCOME* – jakie stany kliniczne (punkty końcowe) służyły do oceny skutków?

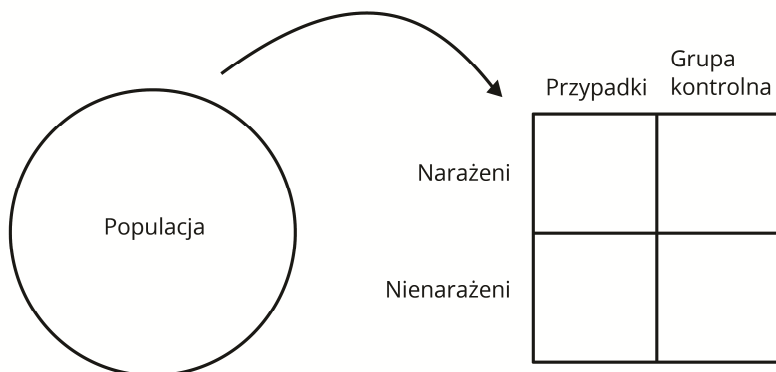
Cel badania powinien zostać opisany z uwzględnieniem wskazanych powyżej elementów: u jakich pacjentów, za pomocą jakiej interwencji, co chce się wykazać i w porównaniu do czego? Następnie należy ustalić w jakim schemacie badawczym autorzy dokonują analizy problemu badawczego. Przypomnijmy z zajęć z epidemiologii pięć podstawowych typów badań:

1. badanie ekologiczne;
2. badanie przekrojowe;
3. badanie kliniczno-kontrolne (*case-control*);
4. badanie kohortowe;
5. badanie eksperymentalne, randomizowane z grupą kontrolną.

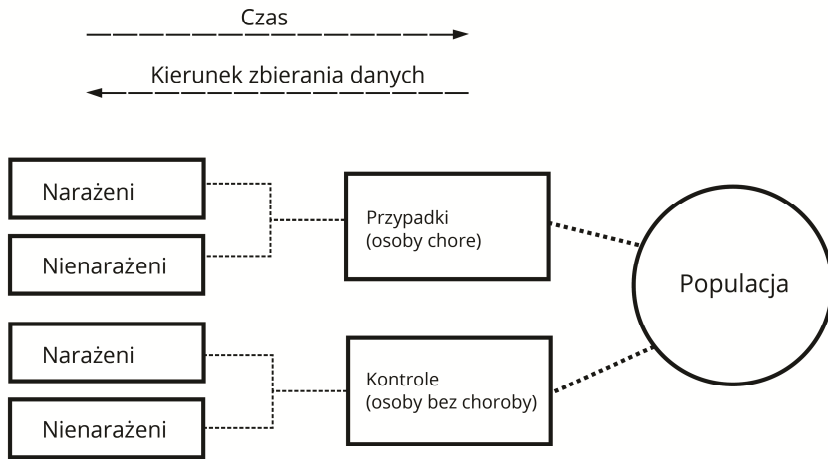
Poniżej zamieszczono ryciny ze schematami głównych typów badań.



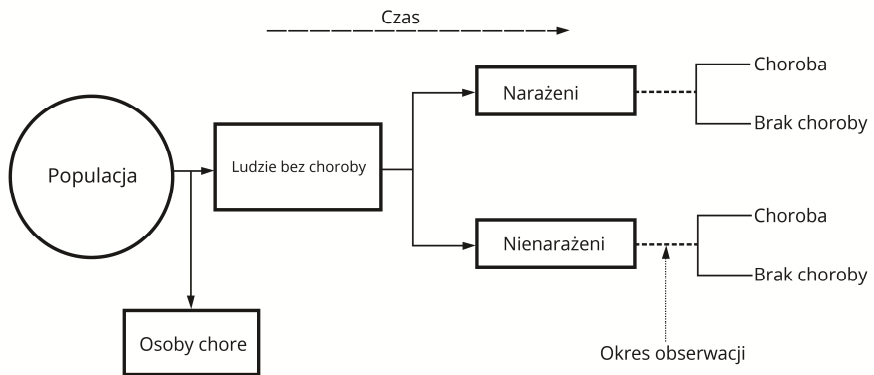
Ryc. 1. Schemat badania ekologicznego



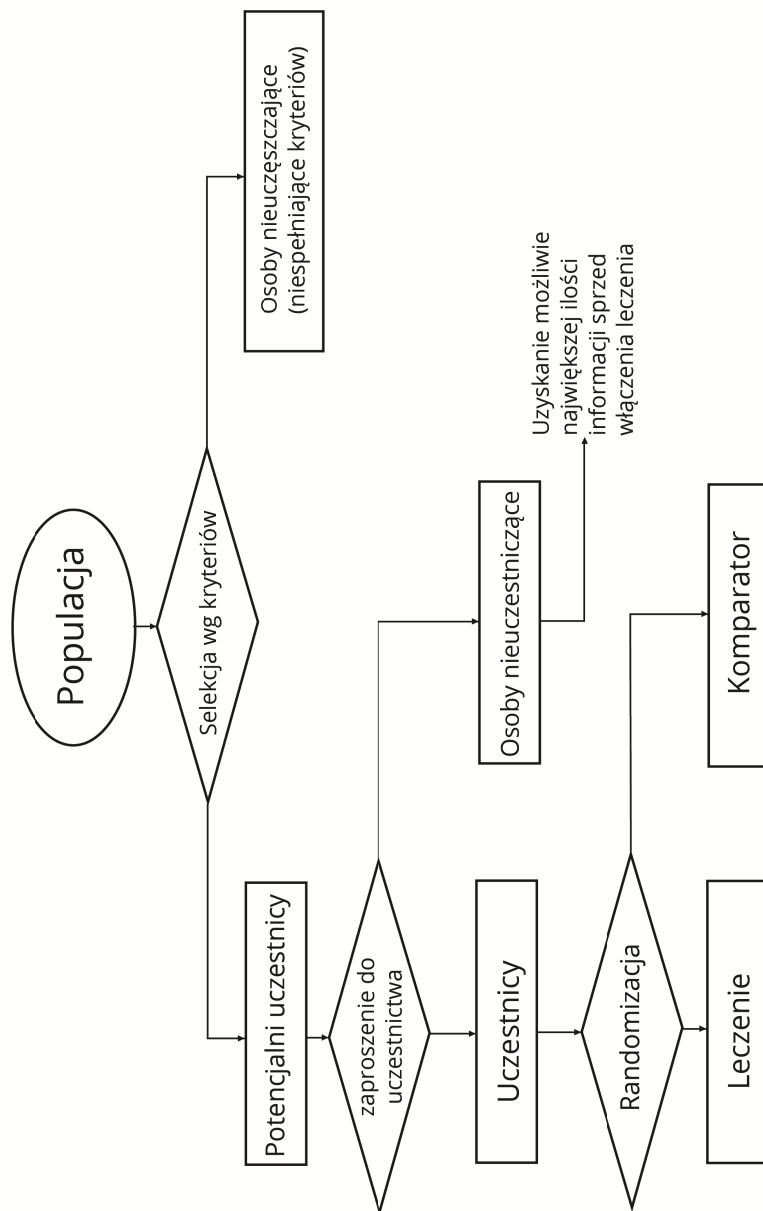
Ryc. 2. Schemat badania przekrojowego



Ryc. 3. Schemat badania kliniczno-kontrolnego



Ryc. 4. Schemat badania kohortowego



Ryc. 5. Schemat badania eksperymentalnego

Jak systematycznie ocenić wiarygodność artykułu?

Następnym etapem oceny jest ustalenie czy autorzy odpowiadają na zadane pytanie w sposób wiarygodny. Historia analitycznej oceny jakości naukowego piśmiennictwa w medycynie nie jest zbyt długa. Wynalazcami terminu *Evidence Based Medicine* byli na początku lat 90. XX wieku lekarze z McMaster University w Kanadzie: David L. Sackett, Gordon H. Guyatt i R. Brian Haynes. Uznali oni, że praktykę kliniczną należy w znacznie większym stopniu oprzeć na dowodach z dobrze zaprojektowanych badań randomizowanych niż polegać na samej opinii ekspertów (*eminence-based medicine*) czy wręcz tradycji. Zgodnie z klasycznymi zasadami EBM sformułowanymi przez ich autorów główne wskazówki zamieszczono w tabeli 2.

Proszę zwrócić uwagę, że mówiąc o wiarygodności artykułów dotyczących leczenia od razu zakładamy, że oceniamy badanie eksperymentalne. Badania obserwacyjne nie służą do oceny skuteczności leczenia! W badaniach obserwacyjnych możemy oceniać niezamierzone efekty leczenia przede wszystkim w kontekście działań niepożądanych leków czy interwencji. Takie badania są przeprowadzane dopiero po rejestracji leku na rynku i ich dłuższym stosowaniu w normalnej praktyce leczniczej. Zauważ, że pierwszym kryterium wiarygodności badania eksperymentalnego jest zastosowanie losowego przydziału interwencji, czyli randomizacji.

Dlaczego się stosuje randomizację? Aby zapewnić równy rozkład czynników ryzyka dla leczonej choroby w grupie interwencji i grupie kontrolnej. Dzięki randomizacji zapewniamy zrównoważony rozkład również wszystkich innych czynników zakłócających, o których niekoniecznie musimy wiedzieć, że są czynnikami ryzyka choroby. To czy randomizacja była skuteczna, można stwierdzić porównując charakterystyki pacjentów na początku badania. Takie dane muszą być zaprezentowane w artykule i zwykle są pokazywane w 1 tabeli.

Na drugim miejscu oceny wiarygodności jest zbadanie zgodności kryteriów włączenia z profilem klinicznym pacjentów faktycznie włączonych do badania. Zgodność ta skutkuje poprawnością wnioskowania o skuteczności leczenia w populacji pacjentów, którą opisano jako uczestniczącą w badaniu (*internal validity*). Jeśli wyniki badania były pozytywne, możemy przyjąć, że lek jest skuteczny w leczeniu choroby u pacjentów podobnych do tych, którzy wzięli udział w badaniu. Nie powinniśmy

Tab. 2. Wskazówki dla czytelnika artykułów dotyczących leczenia

Czy wyniki są wiarygodne?

1. Wskazówki podstawowe

- Czy dobór chorych do grup był przeprowadzony losowo?
- Czy właściwie przeanalizowano dane wszystkich chorych zakwalifikowanych do badania oraz czy zostali oni odpowiednio przydzieleni do grup w momencie jego zakończenia?
- Czy zakończono obserwację odległą?
- Czy analizy końcowe dokonywano w tych grupach, do których pacjenci byli zakwalifikowani losowo na początku? Czy nie zmieniono grup pacjentów?

2. Wskazówki drugorzędne

- Czy pacjenci, personel medyczny i członkowie zespołu badawczego pozostawali nieświadomi istoty metody terapeutycznej?
- Czy w momencie rozpoczęcia badania grupy były porównywalne?
- Czy poza terapią eksperymentalną pacjenci w obu grupach traktowani byli jednakowo?

Jakie wyniki uzyskano?

- Jak silny był efekt terapeutyczny? Jak precyzyjna była ocena efektu leczenia?

Czy wyniki badania będą pomocne w mojej praktyce lekarskiej?

- Czy mogę zastosować wyniki badania w opiece nad moimi pacjentami?
- Czy wzięto pod uwagę wszystkie istotne kliniczne skutki terapii?
- Czy spodziewane korzyści z leczenia przewyższają potencjalne niebezpieczeństwa i koszty?

przenosić wniosków z badania na populację pacjentów różniącą się od tej, jaka brała udział w badaniu (*external validity*).

Inne ważne kryterium to ukończenie obserwacji u wszystkich pacjentów zgodnie z założonym czasem trwania badania. Obserwacja jest ukończona u tych pacjentów, którzy doświadczyli założonego punktu końcowego (*complete*) lub nie zaznali takiego zdarzenia na koniec zaplanowanego okresu obserwacji (*censored*). Dane takie są podstawą przeprowadzenia analizy statystycznej o typie analizy przeżycia. Brak weryfikacji losów

wszystkich pacjentów na koniec badania może wpłynąć na wiarygodne oszacowanie wyniku. Może się przecież zdarzyć, że właśnie ta część pacjentów, o których nie wiemy czy na koniec obserwacji zachorowała, w grupie interwencji zaznała niekorzystnych zdarzeń, a w grupie kontrolnej nie doszło do zachorowań (*worst-case scenario*). Wtedy może dojść do przeszacowania korzystnego efektu leczenia. O wiarygodności badania w tym zakresie świadczy zawarcie w artykule diagramu prezentującego przepływ pacjentów od momentu fazy przesiewowej badania do ostatniej wizyty/oceny w czasie trwania obserwacji. Zaplanowany czas trwania pozostaje w związku z wielkością próby, która jest wymagana, aby wykazać skuteczność leczenia na wskazanym w hipotezie badania poziomie. Dane będące podstawą wyliczenia wielkości próby są niezbędnym składnikiem opisu metod użytych w badaniu i jako takie powinny być zawarte w części dotyczącej opisu metod statystycznych.

Kolejnym kryterium wiarygodności jest trzymanie się zasady *intention-to-treat (ITT) analysis* – analiza zgodnie z zamiarem leczenia. Zakłada ona przeprowadzenie analizy pacjentów i ich danych o zdarzeniach klinicznych w przebiegu badania w tej grupie, do której zostali losowo przydzieleni bez względu na to czy pacjent w grupie aktywnego leczenia przestał przyjmować lek (np. z powodu działań niepożądanych) lub w grupie placebo zaczął przyjmować lek (bo np. pojawiły się wskazania do leczenia). Przestrzeganie tej zasady pozwala ocenić praktyczną skuteczność leczenia. Teoretyczna skuteczność ma znacznie mniejsze znaczenie dla praktyki klinicznej i oceniana jest w tzw. analizie *as treated*, tzn. w grupach pacjentów, którzy faktycznie otrzymywali leczenie, nie doświadczając z tego powodu żadnych niedogodności, które spowodowałyby jego przerwanie.

Zasada ITT umożliwia uniknięcie błędów związanych z przechodzeniem pacjentów do innego ramienia interwencji (*crossover*) a także błędów wynikających z wypadaniem pacjentów z badania (*dropout*). Pozwala ona przez to w większym stopniu uzyskać dane o bezpieczeństwie leczenia obejmując analizą wszystkich uczestników. Pełne dane o tolerancji i działaniach niepożądanych leku rzadko udaje się uzyskać w trybie badań eksperymentalnych. Niektóre z powikłań pojawiają się na tyle rzadko lub po dłuższym czasie stosowania, że daje się je wykryć dopiero w oparciu o badania obserwacyjne z dużą liczebnością uczestników przeprowadzone po rejestracji leku czy nowej technologii medycznej (*signal detection* w badaniach farmakoepidemiologicznych). Stąd niezmiernie ważne jest, aby w badaniach klinicznych rejestrować również wszystkie niezamie-

rzone działania leku, o których nie musimy przed zakończeniem badania wiedzieć, że mają związek z interwencją, tzw. zdarzenia niepożądane (*adverse events*).

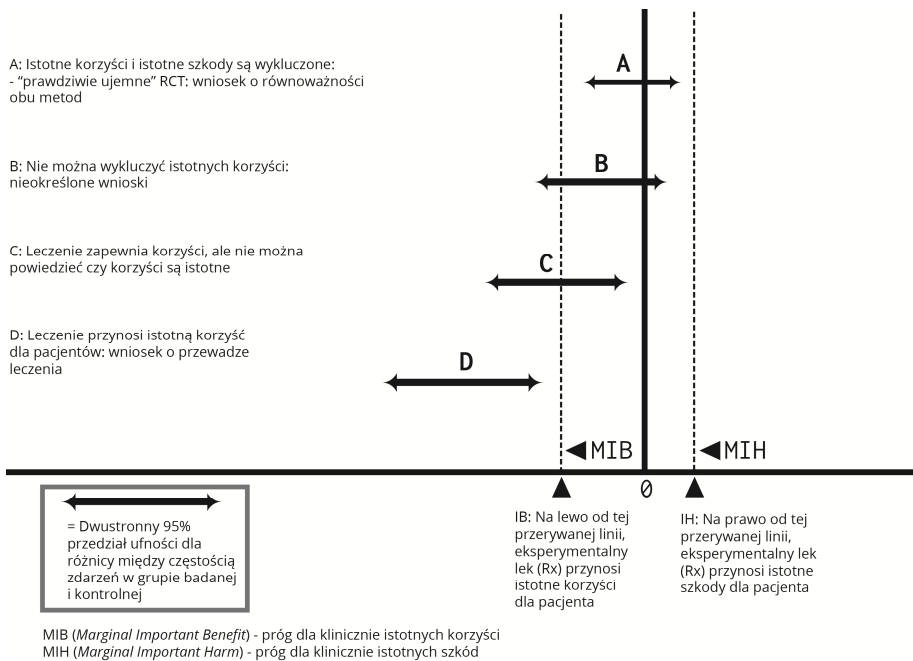
Kryterium drugorzędym, ale istotnie wpływającym na możliwość popełnienia błędu jest zastosowanie maskowania, a więc procedur uniemożliwiających uzyskanie przez pacjentów i leczących ich lekarzy wiedzy o tym czy otrzymują aktywny lek, czy nie. Jeśli nie było maskowania, istotne staje się ustalenie czy sposób leczenia i weryfikacji obecności/diagnozowania chorób jako punktów końcowych był identyczny w grupie interwencji i grupie kontrolnej.

Artykuły o szkodliwości i rokowaniu z reguły mają charakter badania obserwacyjnego. Wówczas, przy braku losowego rozkładu zmiennych zakłócających, kluczowymi kryteriami dla wiarygodności badania obserwacyjnego staje się to czy w momencie rozpoczęcia badania grupa z obecnością czynnika szkodliwego była porównywalna pod względem ważnych cech rokowniczych z grupą kontrolną oraz czy sposób weryfikacji obecności choroby i/lub czynnika szkodliwego był taki sam w porównywanych grupach pacjentów.

Miarą siły efektu terapeutycznego lub szkodliwości jest wartość ryzyka względnego (*Relative Risk* – RR), wskaźnika hazardu (*Hazard Ratio*) w badaniach oceniających czas do zdarzenia (analiza przeżycia) lub ilorazu szans (*Odds Ratio*) w badaniach kliniczno-kontrolnych. Ryzyko względne to stosunek ryzyka w grupie aktywnego leczenia (R_a) do ryzyka w grupie kontrolnej (R_c): $RR = R_a / R_c$. Jeśli wartość wskaźnika jest większa od jedności, wskazuje on na zwiększenie ryzyka, jeśli mniejsza od jedności, wskazuje on na zmniejszenie ryzyka. Im większe zmniejszenie lub zwiększenie, tym silniejszy efekt. Efekt terapeutyczny na poziomie 20% zmniejszenia ryzyka względnego ($RR=0,8$) można uznać za klinicznie istotny. Szkodliwość, która oszacowana jest na poziomie 50% i więcej zwiększenia ryzyka względnego ($RR \geq 1,5$) uznawana jest w zdrowiu publicznym za istotną i każde poszukiwać sposobów na jej zapobieganie.

Aby ocenić jaki jest efekt leczenia u pojedynczego uczestnika należy ocenić wartość ryzyka bezwzględnego ($ARR = R_a - R_c$). To czy efekt ten jest statystycznie istotny oceniamy na podstawie 95% przedziału ufności dla wartości wskaźnika ryzyka (RR, HR, OR) – jeśli nie zawiera on jedności, wówczas mamy do czynienia z efektem istotnym statystycznie (dla wartości ARR przedział ufności, który nie zawiera „0” wskazuje na efekt statystycznie istotny). Granice przedziału ufności są miarą precyzji w badaniu.

Rycina 6 pokazuje jak interpretować wyniki badania klinicznego z grupą kontrolną z placebo (adaptowano z Hayes R. B., Sackett D. L., Guyatt G. H., Tugwell P. *Clinical Epidemiology: How to Do Clinical Practice Research*, 2004). Ocena tego, czy efekt leczenia jest klinicznie istotny czy nie, wymaga kompetencji i wiedzy klinicznej w danej dziedzinie medycyny. Z pewnością jednak nie będziemy zainteresowani 1% zmniejszeniem ryzyka zachorowania nawet w przypadku, gdyby efekt był istotny statystycznie.



Ryc. 6. Możliwe wyniki badania klinicznego (RCT – *Randomized Controlled Trial*)

Diagnostyka

Artykuły dotyczące diagnostyki powinny zawierać w sobie odpowiedzi na poniższe pytania:

- Czy wyniki proponowanego testu diagnostycznego porównano z niezależnie wykonanym i interpretowanym testem referencyjnym?
- Czy pacjenci uwzględnieni w analizie w pełni reprezentowali populację z badanym stanem zdrowia?

- Czy sposób przeprowadzania testu i metody interpretacji opisano w sposób umożliwiający ich odtworzenie?
- Jaka jest czułość i swoistość testu?
- Czy podano dane umożliwiające wyliczenie wartości wskaźników wiarygodności dla wyniku dodatniego/ujemnego (ang. *Likelihood Ratio* – LR)?
- Czy dzięki zastosowaniu testu poprawię rokowanie pacjentów?

Poniżej tabela czteropolowa pomocna w definiowaniu sprawności testu diagnostycznego.

Tab. 3. Parametry trafności testu diagnostycznego

		Wynik testu referencyjnego	
		+	-
Wynik testu ocenianego	+	a	b
	-	c	d

czułość = $a/(a+c)$

swoistość = $d/(b+d)$

wartość predykcji wyników dodatnich (PPV) = $a/(a+b)$

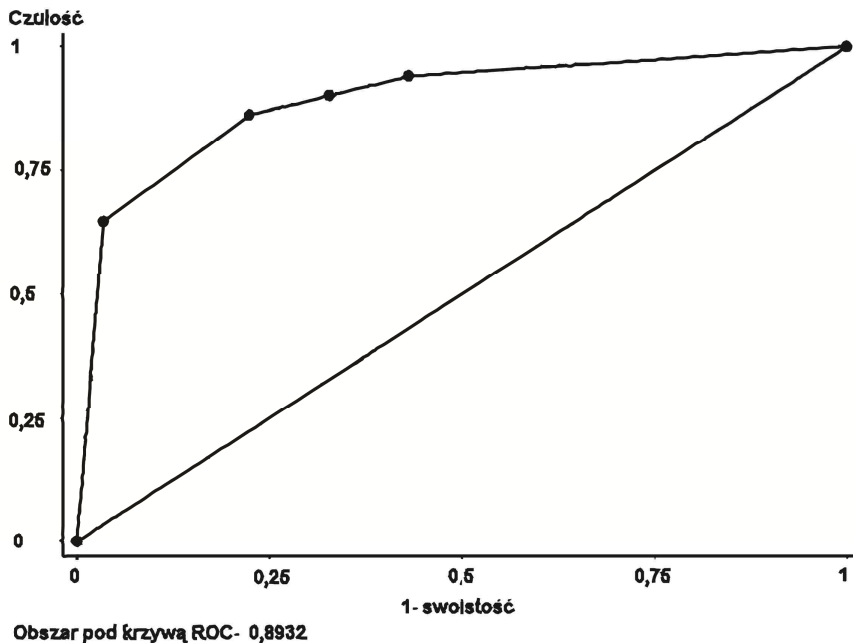
wartość predykcji wyników ujemnych (NPV) = $d/(c+d)$

wskaźnik wiarygodności dla wyniku dodatniego (LR+) = $[a/(a+c)]/[b/(b+d)]$

wskaźnik wiarygodności dla wyniku ujemnego (LR-) = $[c/(a+c)]/[d/(b+d)]$

Wyliczenie wartości wskaźników wiarygodności pozwala ocenić jak bardzo dokładny jest test diagnostyczny w odróżnianiu osób chorych od zdrowych, a także może być pomocne w wyznaczaniu punktów odcięcia decydujących o potwierdzeniu choroby lub jej wykluczeniu. Im większy wskaźnik wiarygodności dla wyniku dodatniego (w zakresie od 1 do ∞), tym dokładniejszy jest test – wartości > 10 potwierdzają obecność choroby. Im mniejszy wskaźnik wiarygodności dla wyniku ujemnego (w zakresie od 0 do 1), tym dokładniejszy jest test – wartości $< 0,1$ wykluczają obecność choroby.

Artykuły o metodzie diagnostycznej często prezentują tzw. krzywe ROC (*Receiver Operating Characteristic Curve*) – krzywe charakterystyki skuteczności percepcji. Powstaje ona poprzez wykreślenie odsetka wyników prawdziwie dodatnich (wartości czułości) względem odsetka wyników fałszywie dodatnich (1 – swoistość) dla różnych wartości punktów odcięcia. Rycina 7 przedstawia krzywą ROC dla przewidywania zaburzeń neurologicznych przy pomocy pięciostopniowej skali oceny obrazu tomografii komputerowej mózgu.



Ryc. 7. Krzywa ROC

Ten punkt odcięcia, który znajduje się najbliżej lewego górnego rogu wykresu ROC, charakteryzuje się największą czułością i najmniejszym odsetkiem wyników fałszywie dodatnich. Krzywa, pod którą pole powierzchni jest bliskie jedności, charakteryzuje idealny test diagnostyczny. Pole pod krzywą bliskie 0,5 wskazuje na test zupełnie nieprzydatny w diagnozowaniu zaburzenia – zdolność rozróżniania chorych od zdrowych jest czysto przypadkowa. Pole powierzchni w zakresie wartości 0,6-0,8 wskazuje na umiarkowaną sprawność testu, 0,8-0,9 dobrą, 0,9-1 doskonałą. Istotność statystyczną oszacowania pola powierzchni pod krzywą ocenia się na podstawie 95% przedziału ufności dla tej wartości. Jeśli ar-

tykuł zajmuje się analizą testu diagnostycznego w badaniach przesiewowych, należy doniesienie oceniać jak każde badanie oceniające skuteczność interwencji. Celem badania skringowego jest wcześniejsze wykrycie choroby w porównaniu do jej naturalnego przebiegu i przez to uzyskanie zmniejszenia ryzyka zgonu z tej przyczyny. Wiarygodne badanie tego rodzaju powinno mieć charakter randomizowanego badania eksperymentalnego. Badanie obserwacyjne nie dostarcza wystarczających dowodów na możliwość zmniejszenia umieralności z powodu choroby, której poszukuje się testem przesiewowym z powodów znanych błędów przeszacowujących efekt interwencji w takim kontekście (*lead time bias, lenght biased sampling*).

Ekonomika

Analiza ekonomiczna jest zagadnieniem wtórnym do oceny skuteczności leczenia. Analiza farmakoekonomiczna sprowadza do *wspólnego mianownika koszty i konsekwencje danej metody leczenia i umożliwia w ten sposób dokonanie wyboru pomiędzy alternatywnymi metodami terapeutycznymi* (za Drummond, 1987). Analiza ekonomiczna zajmuje się sposobem dokonywania wyborów między różnymi metodami leczenia w określonych warunkach systemu opieki zdrowotnej. Próbuje dostarczyć odpowiedzi na pytanie w jaki sposób można rozpoznać i wyjaśnić kryteria podejmowania decyzji o finansowaniu terapii w sytuacji dysponowania ograniczoną ilością zasobów.

Analiza farmakoekonomiczna powinna zawierać następujące elementy:

- rodzaje kosztów uwzględnione w analizie;
- w jakich jednostkach wyrażono wynik analizy opłacalności (kosztów i efektywności);
- jakie rodzaje badań były podstawą oceny skuteczności porównywanych procedur;
- która z porównywanych procedur jest tańsza;
- która z porównywanych procedur jest bardziej opłacalna (efektywniejsza kosztowo);
- czy w moim systemie opieki zdrowotnej koszty są podobne jak w analizie.

W tabeli 4 zawarto zestawienie różnych rodzajów kosztów.

Tab. 4. Rodzaje kosztów uwzględnianych w analizie ekonomicznej

Bezpośrednie (<i>direct</i>)	Pośrednie (<i>indirect</i>)	Niemierzalne (<i>intangible</i>)
- medyczne: leki, sprzęt jednorazowy, amortyzacja aparatury, opieka lekarza, pielęgniarki, pobyt w szpitalu, testy diagnostyczne, monitoring terapii; - niemedyczne: transport, hotel, dieta, opieka nad dzieckiem.	- absencja w pracy; - spadek wydajności w pracy; - wczesna umieralność; - obniżenie lub utrata dochodów.	- ból, cierpienie rozłąka z rodziną; - frustracja, mniejsza zdolność do nauki, życia towarzyskiego, wypełniania rutynowych obowiązków domowych; - poczucie gorszej jakości życia.

Wyróżnia się następujące rodzaje analizy farmakoekonomicznej:

- analiza kosztów i korzyści (Cost-Benefit Analysis, w skrócie CBA);
- analiza kosztów i efektywności (Cost-Effectiveness Analysis, w skrócie CEA);
- analiza kosztów użyteczności (Cost-Utility Analysis, CUA);
- analiza minimalizacji kosztów (Cost-Minimisation Analysis).

W tabeli 5 zawarto zestawienie rodzajów kosztów i korzyści w zależności od typu analizy farmakoekonomicznej.

Tab. 5. Jednostki stosowane w analizie ekonomicznej

Monetarne	Naturalne	Użyteczności
- USD; - PLN.	- obniżenie ciśnienia krwi o mmHg; - obniżenie poziomu glukozy o mmol/ l; - podwyższenie przeżywalności; - skrócenie pobytu w szpitalu (w dniach).	- QALY – lata życia ze skorygowaną jakością; - DALY – lata życia ze skorygowaną niesprawnością; - QoL – jakość życia; - HRQoL – jakość życia zależna od stanu zdrowia; - LYG – zyskane lata życia.

Tab. 6. Rodzaje kosztów i korzyści w zależności od typu analizy farmakoekonomicznej

Rodzaj analizy	Jednostki wartościowania kosztów	Jednostki wartościowania korzyści
Analiza kosztów i korzyści (CBA)	Monetarne	Monetarne
Analiza kosztów i efektywności (CEA)	Monetarne	Naturalne: MmHg, mmol/l, dni, lata
Analiza kosztów i użyteczności (CUA)	Monetarne	Dodatkowe: QALY, DALY, QoL, LYG
Analiza minimalizacji kosztów	Monetarne	Monetarne

Analiza kosztów i korzyści porównuje koszty i korzyści alternatywnych procedur medycznych wyrażane w jednostkach monetarnych. Wynikiem jest wartość wskaźnika stosunku kosztów do korzyści (*costs/benefits ratio*, C/B) lub korzyści netto (*net benefit*), czyli różnica pomiędzy korzyścią a kosztem. Wybiera się procedurę medyczną przynoszącą największą korzyść netto lub najniższy wskaźnik kosztów do korzyści.

Analiza kosztów i efektywności porównuje koszty procedur medycznych wyrażane w jednostkach monetarnych z korzyściami wyrażanymi w jednostkach naturalnych (mmHg, mmol/l, lata, dni itp.). Stosowana jest gdy dwie lub więcej różne procedury medyczne osiągają korzyści o podobnym charakterze, np. zyskane lata życia. Wynikiem tej analizy jest koszt na jednostkę efektu terapeutycznego, a więc ilość pieniędzy potrzebna na rok uratowanego życia, obniżenie ciśnienia krwi o x mmHg, obniżenie poziomu glukozy we krwi o x mmol/l, skrócenie pobytu w szpitalu o x dni, zyskanie jednego roku życia. W wyniku tej analizy procedurę medyczną można uważać za efektywną, gdy jest mniej kosztowna niż inne, a w równym stopniu skuteczna.

Analiza kosztów użyteczności jest rozszerzeniem analizy kosztów i efektywności przez uzupełnienie efektywności oceną stanu zdrowia z punktu odczucia pacjenta. Ocena ta wyrażana jest w jednostkach dodatkowych: QALY (*Quality-Adjusted Life-Year*, lata życia ze skorygowaną jakością), PYLL (*Potential-Year of Life Lost*, utracone lata potencjonalnego życia),

QoL (*Quality of Life*, jakość życia) itp. Jednostki te nazywane są wskaźnikami użyteczności.

Analiza minimalizacji kosztów jest zawężoną i praktycznie nieużywaną odmianą analizy kosztów i efektywności. Porównuje ona koszty i efekty wyrażane w jednostkach pieniężnych. Są one związane z procedurami leczniczymi o analogicznej skuteczności, profilu klinicznym oraz działaniach ubocznych. Zdarza się to w medycynie bardzo rzadko. Wynik jest różnicą kosztów obu metod, z których wybierana jest najtańsza.

Wytyczne praktyki klinicznej

Powszechnie funkcjonujące klasy rekomendacji w opracowywaniu standardów postępowania opisano poniżej.

Klasa I. Sytuacje, w odniesieniu do których istnieją dowody naukowe lub powszechne przekonanie, że rozpatrywana procedura jest korzystna, przydatna i skuteczna.

Klasa II. Sytuacje, w odniesieniu do których dane z badań naukowych są niejednoznaczne i/lub istnieją rozbieżne opinie dotyczące przydatności/skuteczności danej metody leczenia.

Klasa II a. Sytuacje, w odniesieniu do których przeważają dowody/opinie potwierdzające przydatność/skuteczność metody.

Klasa II b. Sytuacje, w odniesieniu do których dowody/opinie nie potwierdzają wystarczająco przydatności/skuteczności metody.

Klasa III. Sytuacje, w odniesieniu do których istnieją dowody naukowe lub powszechne przekonanie, że procedura/sposób leczenia jest nieprzydatna/nieskuteczna, a w niektórych przypadkach może być szkodliwa.

Wytyczne praktyki klinicznej powinny być oparte na systematycznym przeglądzie literatury. Najlepiej jeśli standardy postępowania mają swoje uzasadnienie w wynikach metaanaliz lub wynikach dużych i dobrze zaprojektowanych badań klinicznych. Jeśli takich nie ma wiarygodność dowodu naukowego, na którym oparto wytyczne, spada o poziom niżej.

Poziomy wiarygodności dowodów naukowych

Poziom A (najwyższy): dobrze przeprowadzone, duże badania kliniczne z randomizacją, metaanalizy badań z jednorodnymi wynikami.

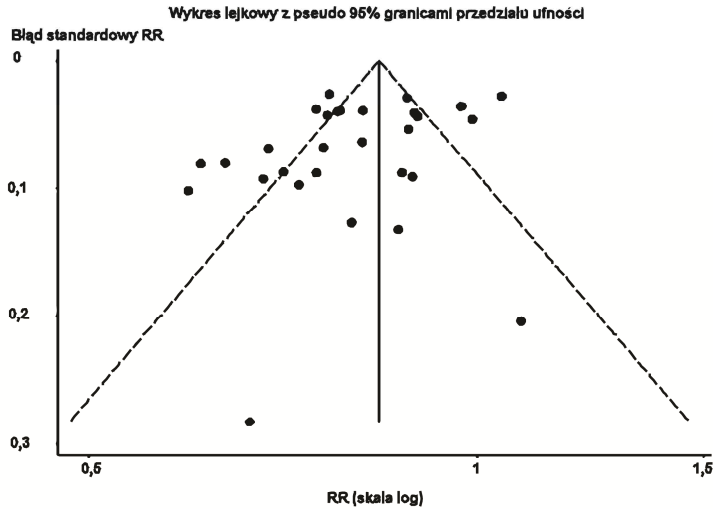
Poziom B (pośredni): nieliczne badania kliniczne z randomizacją z ograniczeniami metodologicznymi, badania bez randomizacji.

Poziom C (najniższy): dobrej jakości badania kohortowe, badania kliniczno-kontrolne, przegląd przypadków, rejestry, zgodna opinia ekspertów.

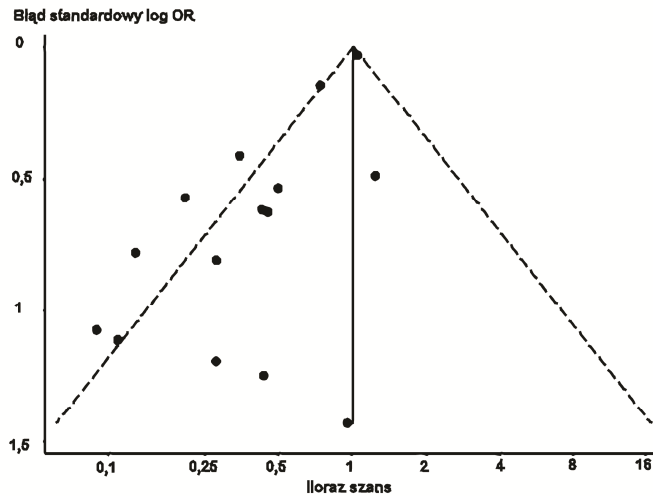
Należy wskazać na istotne ograniczenie współczesnej wiedzy medycznej związane z dostępnością publikowanych dowodów naukowych w medycynie. Jak myślisz, jakie badanie łatwiej opublikować: takie, które pokazuje korzystne efekty interwencji czy też pokazujące brak korzyści z leczenia? Odpowiedź jest oczywista. Fenomen związany z tym zjawiskiem jest znany jako błąd publikacji (*publication bias*). Skutkuje on tym, że więcej jest publikowanych badań z pozytywnymi wynikami niż z negatywnymi. Wiarygodne metaanalizy powinny uwzględniać w swoich wynikach prezentację tzw. wykresów lejkowych (*funnel plot*), które pokazują relację między precyzją (wielkością badania) a uzyskanym wynikiem. Jest to idealne narzędzie do wykrycia błędu publikacji w danej dziedzinie wiedzy. Wykres jest symetryczny przy braku błędu jak na rycinie 8 pokazującej badania oceniające związek między aktywnością fizyczną a umieralnością (Woodcock J et al. *Int. J. Epidemiol.* 2011;40:121-138). Wykres asymetryczny wskazuje na możliwość błędu – brak małych badań o negatywnym wyniku jak na rycinie 9. Mimo początkowych doniesień o korzystnych efektach stosowania magnezu w zawale serca, duże badania (ISIS-4) i ostatnie dobrej jakości badania (MAGIC) pokazały wyniki negatywne. Duże i poprawne metodologicznie badania uzyskują wyniki zbliżone do oszacowanego efektu w metaanalizie i znajdują się blisko osi symetrii i szczytu lejka. Pamiętaj, że firmy farmaceutyczne są zainteresowane publikacją tylko takich badań, które pokazują ich produkty w korzystnym świetle.

Czasopismo *British Medical Journal* w ostatnim czasie bardzo intensywnie opowiada się za pełną przejrzystością w zakresie informacji o wynikach wszystkich przeprowadzanych badań klinicznych, w szczególności tych sponsorowanych przez przemysł. Interesujące były starania autorów związanych z BMJ o uzyskanie pełnego wglądu w dane o wynikach badań nad lekiem Tamiflu (oseltamiwir) zalecanym w leczeniu grypy.

Czytając rekomendacje i stanowiska, warto też wiedzieć czy ich autorzy nie pozostawali w konflikcie interesów, otrzymując wynagrodzenia od firm produkujących leki będące przedmiotem oceny.



Ryc. 8. Wykres lejkowy dla badań oceniających wpływ umiarkowanej aktywności fizycznej na umieralność ogólną – symetryczny układ wyników badań względem wyniku metaanalizy



Ryc. 9. Wykres lejkowy dla 15 badań oceniających efekty stosowania magnezu po zawale serca – przykład asymetrii świadczącej o *publication bias*

Podsumowanie

1. Badania oceniające skuteczność interwencji wiarygodne są wtedy, kiedy mają charakter zaplanowanego eksperymentu, a przydział interwencji był losowy (zastosowano randomizację).
2. W badaniach nieeksperymentalnych podstawowym kryterium wiarygodności jest identyczny/podobny sposób przeprowadzania procedur badawczych w porównywanych grupach pacjentów.
3. W badaniach oceniających test diagnostyczny podstawowym kryterium użyteczności testu jest wykazanie możliwości poprawy rokowania u pacjentów, u których go zastosujemy.
4. Czytając badania oceniające ekonomiczne aspekty leczenia, sprawdź czy uwzględnione koszty są takie same, jakie są obecne w Twoim systemie opieki zdrowotnej.
5. Rekomendacje dotyczące leczenia powinny być oparte na systematycznym przeglądzie (metaanalizach) badań klinicznych i sformułowane przez niezależnych ekspertów.

Polecana literatura

1. *Epidemiologia kliniczna*. Brzeziński Z. J., Szamotulska K. Wydawnictwo Lekarskie PZWL, 1997.
2. *Clinical Epidemiology: How to Do Clinical Practice Research*. Haynes R. B., Sackett D. L., Guyatt G. H., Tugwell P. Lippincott Williams & Wilkins, 2005.
3. *How to Read a Paper: The Basics of Evidence-Based Medicine*. Greenhalgh T. BMJ Books, 2010.