

# Ocena jakości danych i klasyfikacja siły zaleceń

Grading quality of evidence and strength of recommendations

Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) Working Group

British Medical Journal, 2004; 328: 1490–1494

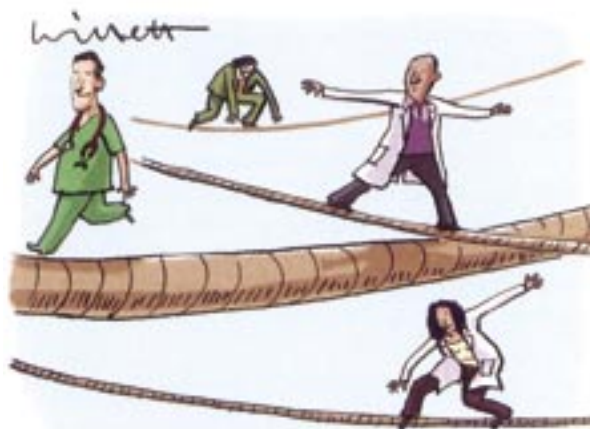
(Od Redakcji: W opracowaniu systemu przedstawionego w tym artykule brali czynny udział członkowie zespołu redakcyjnego „Medycyny Praktycznej” – prof. Roman Jaeschke, dr Jacek Mrukowicz i dr Jan Brożek, a jedno ze spotkań Grupy Roboczej GRADE odbyło się w Krakowie i zostało zorganizowane przez Polski Instytut Evidence Based Medicine i Medycynę Praktyczną.)

Jakość wytycznych praktyki klinicznej zależy od wiarygodności danych, na których je oparto, oraz od trafności decyzji podejmowanych w procesie ich tworzenia. Celem systemu GRADE jest ułatwienie użytkownikom wytycznych oceny procesu formułowania poszczególnych zaleceń.

Stosując się do wytycznych praktyki klinicznej i innych zaleceń, pracownicy służby zdrowia powinni wiedzieć, na ile mogą im ufać. Systematyczna i jasno sprecyzowana metodologia tworzenia wytycznych może zmniejszyć ryzyko sformułowania błędnych zaleceń oraz poprawić ich czytelność. Wypracowaliśmy system oceny jakości danych i klasyfikacji siły zaleceń, który można zastosować do rozmaitych interwencji i w różnych sytuacjach klinicznych. W niniejszym artykule przedstawiamy podsumowanie naszego podejścia z punktu widzenia użytkowników wytycznych.

## Czym się charakteryzują dobre wytyczne?

Ocena danych i zaleceń jest złożona. Weźmy na przykład wybór między trójpierścieniowymi lekami przeciwdepresyjnymi a wybiórczymi inhibito-



rami wychwyty zwrotnego serotoniny w leczeniu umiarkowanej depresji. Dokonując wyboru, trzeba zdecydować, które efekty działania leków (czyli tzw. punkty końcowe) uwzględnić, które dane dotyczące każdego z nich wziąć pod uwagę, jak ocenić wiarygodność tych danych i jak rozstrzygnąć, czy stosowanie wybiórczych inhibitorów wychwyty zwrotnego serotoniny przyniesie więcej korzyści niż szkód w porównaniu z lekami trójpierścieniowymi. Ponadto konieczne może być również określenie, czy jakiegokolwiek dodatkowe korzyści są warte dodatkowo poniesionych kosztów, ponieważ zasoby są zawsze ograniczone i pieniądze wydanych na sfinansowanie leczenia wybiórczymi inhibitorami wychwyty zwrotnego serotoniny nie będzie można wykorzystać na inne cele.

Dla poszczególnych klinicystów (czyli wszystkich osób z wykształceniem medycznym opiekujących się chorymi, a więc zarówno lekarzy, jak i pielęgniarek, rehabilitantów itd. – przyp. tłum.) i pacjentów rozważanie wszystkich tych aspektów przy podejmowaniu każdej decyzji jest niepraktyczne. W tej sytuacji klinicyści i pacjenci często posługują się wytycznymi. Użytkownicy wytycznych powinni jednak wiedzieć, na ile mogą zaufać

**Tabela 1. Porównanie systemu GRADE z innymi systemami oceny jakości danych i klasyfikacji siły zaleceń**

Aspekt	Inne systemy	GRADE	Zalety systemu GRADE*
pojęcia	nie ma precyzyjnych definicji stopni jakości danych i siły zaleceń	precyzyjnie zdefiniowane pojęcia	jasno określa, co oznaczają poszczególne stopnie i jakie czynniki należy uwzględnić przy ich ustalaniu
zasady oceny	nie ma jasno sprecyzowanych zasad wyboru ważnych punktów końcowych oraz oceny jakości danych dla każdego z nich, ogólnej jakości danych, bilansu korzyści i szkodliwości, wagi dodatkowych korzyści zdrowotnych	jasno sprecyzowane zasady sekwencyjnego dokonywania ocen	objaśnia zasady oceny każdego z aspektów i zmniejsza ryzyko popełnienia błędu lub spaczoności oceny
zasadnicze kryteria jakości danych	nie są rozważane dla każdego ważnego punktu końcowego; ocena jakości danych często oparta wyłącznie na rodzaju badania	systematyczne i jasno sprecyzowane uwzględnienie rodzaju i jakości badań oraz zgodności i odniesienia ich wyników	zapewnia właściwe uwzględnienie tych czynników
inne czynniki mogące wpłynąć na jakość danych	nie są uwzględniane w sposób jasno sprecyzowany	jasno określone uwzględnienie nieprecyzyjności oszacowania efektów, zbyt małej liczby danych, prawdopodobieństwa nieopublikowania części badań na dany temat, siły związku między interwencją a punktem końcowym, zależności efektu od dawki oraz czynników zakłócających	zapewnia właściwe uwzględnienie tych czynników
ogólna jakość danych	oceniana w sposób niesprecyzowany, na podstawie jakości danych dotyczących tylko korzystnych efektów interwencji	oceniana na podstawie najniższej jakości danych dla któregośkolwiek z punktów końcowych krytycznych dla podjęcia decyzji	zmniejsza ryzyko błędnej oceny ogólnej jakości danych w przypadku, gdy nie ma danych dotyczących krytycznego punktu końcowego
względna waga poszczególnych punktów końcowych	uwzględniana w sposób nieokreślony	jasno sprecyzowana ocena, które z punktów końcowych są krytyczne dla podjęcia decyzji, ważne, ale nie krytyczne, i które można pominąć	zapewnia właściwe uwzględnienie każdego z punktów końcowych w ocenie ogólnej jakości danych i określaniu siły zalecenia
bilans korzyści zdrowotnych i szkodliwości	niewuwzględniany w sposób jasno sprecyzowany	jasno sprecyzowana ocena bilansu istotnych korzyści i szkodliwości, jakości danych na temat korzyści i szkodliwości, możliwości odniesienia danych do określonej sytuacji klinicznej oraz stopnia pewności co do ryzyka podstawowego	objaśnia i zwiększa przejrzystość oceny bilansu korzyści i szkodliwości

bilans dodatkowych korzyści zdrowotnych i kosztów	nieuwzględniany w sposób jasno sprecyzowany	jasno sprecyzowana ocena bilansu korzyści i kosztów po uprzednim uznaniu jednoznacznej przewagi korzyści z danej interwencji	zapewnia przejrzystość decyzji o wadze korzyści zdrowotnej z określonej interwencji
zestawianie danych i wyników	niejednolite sposoby przedstawiania	powtarzalne zestawienia danych, zawierające ocenę jakości danych i podsumowanie wyników	zapewnia członkom zespołu dokonywanie ocen na podstawie tych samych danych i dostępność tych danych dla innych zainteresowanych osób
zakres zastosowania	rzadko wykorzystywane przez więcej niż jedną instytucję i rzadko lub nigdy niepoddane praktycznej weryfikacji	międzynarodowa współpraca wielu różnych organizacji w tworzeniu i ocenie systemu	wykorzystuje wcześniejsze doświadczenia do stworzenia bardziej racjonalnego i wiarygodnego systemu nadającego się do szerokiego stosowania

\* Większość innych systemów nie ma żadnej z wymienionych zalet, natomiast część może mieć niektóre z nich.

danym, które wykorzystano przy ich tworzeniu, oraz sformułowanym zaleceniom. W niniejszym artykule opisujemy czynniki, na których powinniśmy opierać nasze zaufanie do wytycznych, oraz systematyczne podejście do złożonych ocen stanowiących jasno sprecyzowaną (albo tylko domyślną) część procesu formułowania zaleceń. Starając się uprościć nasz wywód, nie poruszamy tutaj wszystkich niuansów zagadnienia; niektóre z nich są omówione bardziej szczegółowo w dłuższej wersji tego artykułu, dostępnej na stronach internetowych [bmj.com](http://bmj.com).

Grupa Robocza GRADE powstała w wyniku nieformalnej współpracy osób zainteresowanych rozwiązaniem problemu niedociągnięć aktualnych systemów klasyfikacji zaleceń. Niedociągnięcia te, wraz z propozycjami ich usunięcia, przedstawiliśmy w tabeli 1. System GRADE umożliwia dokonywanie oceny jakości danych oraz bilansu korzyści i szkodliwości w sposób bardziej powtarzalny, a przejrzyste opisanie tego procesu i jego wyników może sprzyjać podejmowaniu lepiej uzasadnionych decyzji w opiece zdrowotnej. Poszczególne etapy opracowywania i wdrażania wytycznych – od ustalenia hierarchii problemów aż do oceny przestrzegania zaleceń, przedstawiliśmy w ramce 1. Poniżej skupimy się na ocenie jakości danych i klasyfikacji siły zaleceń.

## Objaśnienie pojęć

W naszych rozważaniach używamy następujących pojęć: jakość danych oznacza przeświadczenia, że wielkość efektu danej interwencji została prawidłowo oszacowana; siła zalecenia wskazuje stopień przeświadczenia, że postępowanie zgodnie z danym zaleceniem przyniesie więcej korzyści niż szkód.

Na poszczególnych etapach naszego systemu kolejno rozstrzygamy:

- jaka jest jakość danych uzyskanych ze wszystkich badań klinicznych, w których oceniano poszczególne ważne punkty końcowe (*czyli zdarzenia mające istotne znaczenie dla pacjentów, np. wystąpienie udaru mózgu – przyp. tłum.*);
- które z tych punktów końcowych mają krytyczne znaczenie dla podjęcia decyzji;
- jaka jest ogólna jakość danych dla wszystkich krytycznych punktów końcowych;
- jaki jest bilans korzyści i szkodliwości związanych z daną interwencją;
- jaka jest siła zalecenia.

### Ramka 1. Kolejne etapy procesu tworzenia wytycznych

#### Etap pierwszy

1. **Rozpoczęcie procesu** – np. ustalenie hierarchii problemów, wybór członków zespołu, ujawnienie sprzeczności interesów, uzgodnienie zasad pracy zespołu.

#### Etapy wstępne

2. **Przegląd systematyczny** – w pierwszej kolejności należy znaleźć już przeprowadzone przeglądy systematyczne i ocenić ich wiarygodność albo sporządzić nowy przegląd systematyczny najlepszych dostępnych danych na temat wszystkich ważnych punktów końcowych.

3. **Przygotowanie zestawienia danych dotyczących ważnych punktów końcowych** – potrzebne są oddzielne zestawienia dla każdej części populacji lub grupy ryzyka uwzględnionej w przeglądzie systematycznym: powinny zawierać ocenę jakości danych i podsumowanie wyników.

#### Ocena jakości danych i klasyfikacja siły zaleceń

4. **Ocena jakości danych dla każdego ważnego punktu końcowego** – oceny dokonuje się na podstawie informacji zebranej w zestawieniu danych, stosując kryteria przedstawione w ramce 2.

5. **Określenie względnej wagi punktów końcowych** – w zestawieniach danych powinny zostać ujęte tylko ważne punkty końcowe, które następnie należy określić jako krytyczne albo ważne (ale nie krytyczne) dla podjęcia decyzji.

6. **Ocena ogólnej jakości danych** – za ogólną jakość danych przyjmuje się najniższą jakość danych dla któregokolwiek z krytycznych punktów końcowych.

7. **Ocena bilansu korzyści i szkodliwości** – bilans korzyści i szkodliwości powinien być ustalony jako: jednoznaczna przewaga korzyści, przewaga korzyści dyskusyjna, przewaga korzyści niewiadoma, nie ma przewagi korzyści.

8. **Ocena bilansu korzyści i kosztów** – Czy dodatkowe korzyści zdrowotne są warte poniesionych kosztów? Ponieważ dostępne środki są zawsze ograniczone, formułując zalecenie należy wziąć pod uwagę koszty (zużycie zasobów).

9. **Siła zalecenia** – zalecenia powinny być tak formułowane, aby odzwierciedlały swoją siłę, to znaczy stopień, w jakim można mieć pewność, że postępowanie zgodnie z zaleceniem przyniesie więcej korzyści niż szkód.

#### Etapy kolejne

10. **Wdrażanie i weryfikacja** – np. stosowanie skutecznych strategii wdrażania, które uwzględniają przeszkody w zmianie zachowań, weryfikacja wdrożenia, zapewnienie aktualizacji wytycznych

Wszystkie te rozstrzygnięcia zależą od jasnego sformułowania pytania i uwzględnienia wszystkich punktów końcowych przypuszczalnie ważnych dla osób, których dotyczą. Pytanie powinno precyzować, jakie opcje postępowania się porównuje (np. wybiórcze inhibitory wychwytu zwrotnego serotoniny i trójpierścieniowe leki przeciwdepresyjne), u kogo będzie się je stosować (u chorych na umiarkowaną depresję) i w jakich okolicznościach (w podstawowej opiece zdrowotnej w Anglii).

### Jakość danych

Ocena jakości danych powinna polegać na przeglądzie systematycznym wszystkich dostępnych danych. Osoby oceniające jakość danych powinny wziąć pod uwagę 4 kluczowe zagadnienia (ramka 2): rodzaj badań, ich jakość metodologiczną, zgodność wyników poszczególnych badań i ich odniesienie. Rodzaje badań klinicznych podzieliłmy bardzo ogólnie na badania obserwacyjne i badania z randomizacją. Jakość badania odnosi się do jego szczegółowej metodologii i przeprowadzenia. Zgodność wyników oznacza podobieństwo oszacowanego efektu interwencji w poszczególnych badaniach. Odniesienie wyników dotyczy stopnia, w jakim populacja, interwencja i jej efekty w uwzględnionych badaniach odpowiadają sytuacji, dla której chcemy sformułować zalecenie. Jeśli nie porównano określonych interwencji bezpośrednio w jednym badaniu klinicznym i trzeba to zrobić na podstawie danych z oddzielnych badań, to mamy do czynienia z danymi pośrednimi.

Uwzględniając każdy z wymienionych 4 elementów, można określić jakość danych dla każdego ważnego punktu końcowego. W pierwszej kolejności klasyfikujemy dostępne dane na podstawie rodzaju badania – na pochodzące z badań z randomizacją albo z badań obserwacyjnych (ramka 2). Następnie proponujemy rozważyć, czy badania te miały poważne uchybienia, czy nie ma istotnych niezgodności między ich wynikami i czy istnieją uzasadnione wątpliwości co do odniesienia tych wyników.

Dodatkowe czynniki mogące obniżyć jakość danych to: nieprecyzyjne oszacowanie efektu lub zbyt mała liczba danych oraz duże prawdopodobieństwo, że część badań na dany temat nie została opublikowana. Dodatkowe czynniki mogące podwyższyć jakość danych to: bardzo silny zwią-

**Ramka 2. Kryteria oceny jakości danych****rodzaj badania**

badanie z randomizacją = jakość wysoka  
 badanie obserwacyjne = jakość niska  
 jakiegokolwiek inne dane = jakość bardzo niska

**czynniki obniżające jakość danych**

- poważne (–1) albo bardzo poważne (–2) ograniczenie jakości badania
- ważna niezgodność wyników (–1)
- umiarkowana (–1) lub duża (–2) niepewność co do możliwości odniesienia danych
- nieprecyzyjne oszacowanie efektów lub zbyt mała liczba danych (–1)
- duże prawdopodobieństwo, że część badań na dany temat nie została opublikowana (–1)

**czynniki zwiększające jakość danych**

- silny związek pomiędzy interwencją a punktem końcowym – statystycznie istotne ryzyko względne  $>2$  (albo  $<0,5$ ) oszacowane na podstawie zgodnych wyników  $\geq 2$  badań obserwacyjnych, bez prawdopodobnych czynników zakłócających mogących zmniejszyć zaobserwowany efekt (+1)
- bardzo silny związek pomiędzy interwencją a punktem końcowym – statystycznie istotne ryzyko względne  $>5$  (albo  $<0,2$ ) na podstawie danych bezpośrednich, bez poważnych zastrzeżeń co do ich wiarygodności (+2)
- wykazanie zależności efektu od dawki (+1)
- wszystkie prawdopodobne, ale niewzięte pod uwagę czynniki zakłócające najprawdopodobniej zmniejszyły zaobserwowany efekt (+1)

zek pomiędzy interwencją a punktem końcowym (np. 50-krotne zwiększenie ryzyka zgonu z powodu zatrucia w przypadku trójpierścieniowych leków przeciwdepresyjnych [tab. 2]) albo silny związek (np. 3-krotne zwiększenie ryzyka urazu głowy u rowerzystów nienoszących kasków ochronnych, w porównaniu z jeżdżącymi w kaskach<sup>1</sup>), a także wykazana zależność efektu od dawki. Proponowane definicje stopni jakości danych przedstawiono w ramce 3.

Powyższe zasady należy stosować, oceniając jakość danych zarówno pod względem korzyści, jak i szkodliwości określonej interwencji. Ważne prawdopodobne szkodliwości danej interwencji mogą i powinny być ujęte w zestawieniach danych, nawet jeżeli istnieją tylko pośrednie przesłanki ich występowania. Na przykład, jeżeli obawiamy się, że badania przesiewowe w kierunku

**Ramka 3. Stopniowanie jakości danych**

**jakość wysoka** = dalsze badania prawdopodobnie nie zmienią naszego przekonania o trafności oszacowania efektu interwencji

**jakość średnia** = dalsze badania prawdopodobnie będą miały istotny wpływ na nasze przekonanie o trafności oszacowania efektu i mogą zmienić to oszacowanie

**jakość niska** = dalsze badania najpewniej będą miały istotny wpływ na nasze przekonanie o trafności oszacowania efektu i prawdopodobnie zmienią to oszacowanie

**jakość bardzo niska** = jakiegokolwiek oszacowanie efektu jest bardzo niepewne

czerniaka złośliwego mogą wywołać niepokój u badanych osób, a nie dysponujemy żadnymi bezpośrednimi danymi o tym świadczącymi, to właściwe może być wykorzystanie danych dotyczących innych badań przesiewowych.

Jakość danych dla ważnych punktów końcowych ze wszystkich badań klinicznych powinno się oceniać na podstawie przeglądów systematycznych, takich jak przeglądy dokonywane przez Cochrane Collaboration. Ocena ogólnej jakości danych, bilansu korzyści i szkodliwości oraz klasyfikacja siły zaleceń zwykle wymaga dodatkowych informacji, których nie ma w wynikach przeglądu systematycznego.

W innych systemach ocenę ogólnej jakości danych opierano najczęściej tylko na jakości danych dotyczących korzyści z określonej interwencji. Kiedy ryzyko wystąpienia skutku niepożądanego ma zasadnicze znaczenie dla podjęcia decyzji o zaleceniu określonej interwencji, a dane na temat tego ryzyka są niższej jakości niż na temat korzyści, to zignorowanie niepewności co do ryzyka stwarza problem. Sugerujemy zatem, aby o ogólnej jakości danych orzekać na podstawie najniższej jakości danych dla któregośkolwiek z krytycznych punktów końcowych.

**Zalecenia****Czy korzyści ze stosowania interwencji przeważają nad jej szkodliwością?**

Formułowanie zaleceń wiąże się z oceną bilansu korzyści i szkodliwości. Dokonując tej oceny, każdemu punktowi końcowemu przypisujemy, w spo-

**Tabela 2. Ocena jakości danych z badań, w których porównywano wybiórcze inhibitory wychwyty zwrotnego serotoniny (SSRI) z trójpierścieniowymi lekami przeciwdepresyjnymi (TAD) u chorych na umiarkowaną depresję w podstawowej opiece zdrowotnej?**

liczba badań	Ocena jakości danych				Podsumowanie wyników				
	rodzaj badania	jakość badania	zgodność wyników	odniesienie	inne czynniki*	liczba chorych	efekt	jakość	waga
						SSRI	względny (95% CI)	bez-względny	
<b>punkt końcowy: nasilenie depresji (oceniane za pomocą skali depresji Hamiltona po upływie 4–12 tygodni)</b>									
citalopram (8)	badania	bez	bez	niewielkie	nie	5044	WMD 0,034	nie ma	średnia krytyczna
fluoksetyna (38)	z rando-	poważnych	poważnej	(punkt		4510	(od -0,007	różnicy	
fluwoksamina (25)	mizacją	ograniczeń	niezgod-	końcowy)**			do 0,075)		
nefazodon (2)			ności						
paroksetyna (18)									
sertralina (4)									
wenlafaksyna (4)									
<b>punkt końcowy: wycofanie z badania z powodu przemieszających skutków ubocznych</b>									
citalopram (8)	badania	bez	bez	tak	nie	1948/7032	RRR 13%	5/100	wysoka krytyczna
fluoksetyna (50)	z rando-	poważnych	poważnej			2072/6334	(5–20%)		
fluwoksamina (27)	mizacją	ograniczeń	niezgod-			(33%)			
nefazodon (4)			ności						
paroksetyna (23)									
sertralina (6)									
wenlafaksyna (5)									
<b>punkt końcowy: zgon z powodu zatrucia</b>									
UK Office for National Statistics (1)	badania obserwacyjne	poważne ograniczenie#	tylko jedno badanie	tak	bardzo silny związek między interwencją a punktem końcowym	1/100 000/ rok leczenia	RRR 98% (97–99%)#	6/10 000	średnia krytyczna

\* Nieprecyzyjne oszacowanie efektu lub zbyt mała liczba danych, silny lub bardzo silny związek pomiędzy interwencją a punktem końcowym, duże prawdopodobieństwo, że część badań nie została opublikowana, wykazanie zależności efektu od dawki, wszystkie prawdopodobne a nieuwzględnione czynniki zakłócające zmniejszyłyby zaobserwowany efekt.

\*\* niepewność co do odniesienia wyników ze względu na krótki okres obserwacji

# Możliwe, że osoby obciążone mniejszym ryzykiem miały większą szansę otrzymania SSRI i nie wiadomo, czy zmiana leku przeciwdepresyjnego zmniejszyłaby częstość prób samobójczych.

## niepewne podstawowe ryzyko zgonu z powodu zatrucia lekiem

RRR (*relative risk reduction*) – względne zmniejszenie ryzyka, WMD (*weighted mean difference*) – średnia ważona różnic

sób jasno sprecyzowany albo nie, jakąś względną wagę. Proponujemy, aby zawsze najpierw dokonywać jasno sprecyzowanej oceny bilansu najważniejszych korzyści i szkodliwości, a potem dopiero uwzględniać koszty. Czy korzyści ze stosowania interwencji przeważają nad jej szkodliwością?

Gdy korzyści i szkodliwości są zróżnicowane w zależności od populacji lub sytuacji klinicznej, to zalecenia dotyczące danej interwencji muszą się odnosić do określonych sytuacji i populacji. Na przykład: czy powinniśmy zalecać, by chorzy z migotaniem przedsionków otrzymywali doustny antykoagulant w celu zmniejszenia ryzyka udaru mózgu, pomimo jednoczesnego zwiększenia ryzyka krwawienia? Zalecenia lub ich siła prawdopodobnie będą różne, gdy możliwe jest monitorowanie intensywności antykoagulacji i gdy takiej możliwości nie ma. Ponadto zalecenia (lub ich siła) prawdopodobnie będą różne dla chorych obciążonych małym ryzykiem udaru (w wieku <65 lat i bez chorób współistniejących) i dla chorych bardziej zagrożonych (np. starszych z niewydolnością serca), ze względu na różnicę w bezwzględny zmniejszeniu ryzyka. Zalecenia muszą się więc odnosić do określonej grupy chorych i sytuacji klinicznej.

Osoby tworzące zalecenie powinny wziąć pod uwagę 4 główne czynniki:

- bilans korzyści i szkodliwości danych interwencji, uwzględniając oszacowaną wielkość efektu, precyzję tego oszacowania i względną wagę przypisaną każdemu z głównych punktów końcowych;
- jakość danych;
- możliwość odniesienia danych pochodzących z badań naukowych do praktyki klinicznej w określonych warunkach, uwzględniając ważne czynniki mogące wpłynąć na wielkość oczekiwanych efektów interwencji, takie jak odległość do najbliższego szpitala lub dostępność wykwalifikowanego personelu medycznego;
- niepewność co do wielkości ryzyka podstawowego w populacji, do której zalecenie będzie się odnosiło.

Niepewność co do możliwości odniesienia danych do praktyki w określonych warunkach lub co do ryzyka podstawowego może zmniejszyć nasze przekonanie o słuszności zalecenia. Weźmy na przykład interwencję przynoszącą ważne korzyści zdrowotne, ale jednocześnie powodującą poważne skutki niepożądane – w przypadku gdy ryzyko podstawowe w danej populacji nie jest do-

kładnie znane, to zalecenie prawdopodobnie będzie znacznie słabsze, niż wówczas gdy ryzyko to jest znane.

Proponujemy następujące kategorie zaleceń:

- **„rób to”** albo **„nie rób tego”** – wskazuje, że zdecydowana większość osób dysponujących odpowiednimi danymi zgodziłaby się na taką ocenę bilansu korzyści i szkodliwości;
- **„raczej rób to”** albo **„raczej nie rób tego”** – wskazuje, że większość osób dysponujących odpowiednimi danymi zgodziłaby się na taką ocenę bilansu korzyści i szkodliwości, ale znacząca część miałaby odmienne zdanie.

Zalecenie stosowania albo zaniechania danej interwencji nie oznacza, że wszystkich chorych powinno się leczyć jednakowo. Nie oznacza to również, że lekarz nie powinien uwzględniać zdania pacjenta przy podejmowaniu decyzji lub przedstawiać mu zalet alternatywnego postępowania. Niemniej jednak, ponieważ większość pacjentów dysponujących odpowiednimi informacjami dokonałaby takiego samego wyboru, względne korzyści z alternatywnych sposobów postępowania można im przedstawić stosunkowo krótko. Każde zalecenie jest tworzone z zamiarem ułatwienia podjęcia słusznej decyzji dotyczącej pojedynczego pacjenta lub populacji. Powinno ono zatem odzwierciedlać opcję, którą pacjenci najpewniej by wybrali na podstawie dostępnych danych i własnego systemu wartości lub preferencji oraz oczekiwanych efektów. Zalecenie, aby coś „raczej robić”, wskazuje na potrzebę bardziej wnikliwego rozważenia przez klinicystę systemu wartości i preferencji pacjenta, gdy proponuje się mu określony sposób postępowania.

W pewnych okolicznościach sformułowanie zalecenia byłoby niewłaściwe, z powodu niejasnego bilansu korzyści i szkodliwości lub niezgodności w jego ocenie. Jeśli tego przyczyną jest brak danych dobrej jakości, to powinno się postulować przeprowadzenie odpowiednich badań klinicznych, które dostarczyłyby danych niezbędnych do sformułowania zalecenia.

## Czy dodatkowe korzyści zdrowotne są warte poniesionych kosztów?

Ponieważ wydanie pieniędzy na jedną interwencję oznacza mniej pieniędzy do wydania na inną, to tworząc zalecenia rozpatrujemy (w sposób jasno sprecyzowany albo nie) wartość dodatkowych korzyści zdrowotnych w odniesieniu do dodatko-



wych kosztów. Koszty – wartość pieniężna zużytych zasobów – są ważnym czynnikiem przy formułowaniu zaleceń, ale zależą od okoliczności i zmieniają się z czasem, a ich wielkość bywa trudna do oszacowania. Zdając sobie sprawę z trudności w szacowaniu kosztów, proponujemy przy rozważaniu spodziewanych korzyści zdrowotnych i szkodliwości różnych alternatywnych interwencji uwzględniać w sposób przejrzysty i możliwie jasno sprecyzowany dodatkowe koszty z nimi związane. Tam gdzie jest to celowe i możliwe, w zestawieniu danych powinno się – wraz z ważnymi punktami końcowymi – podawać „rozłączne” koszty interwencji (różnice zużycia zasobów). Jakość danych dotyczących zużycia zasobów należy ocenić w taki sam sposób, jak opisano powyżej dla innych ważnych punktów końcowych.

### Jak to wygląda w praktyce?

Tabela 2 zawiera przykład zastosowania systemu GRADE do oceny jakości danych z opublikowanego w 1997 roku przeglądu systematycznego<sup>2</sup>, w którym porównano wybiórcze inhibitory wychwyty zwrotnego serotoniny z trójpierścienowymi lekami przeciwdepresyjnymi. Po dyskusji uzgodniliśmy, że jakość danych dotyczących względnego wpływu obu grup leków na nasilenie depresji i związanego z ich stosowaniem ryzyka zgonu z powodu zatrucia jest średnia, a danych na temat przemijających skutków ubocznych – wysoka. Następnie się zgodziliśmy, że ogólna jakość danych jest średnia i że korzystniejsze jest stosowanie inhibitorów wychwyty zwrotnego serotoniny (bez różnicy w nasileniu depresji, mniej przemijających skutków ubocznych i mniej zgonów z powodu zatrucia). Mimo to ostatecznie sformułowaliśmy zalecenie, aby „raczej” stosować wybiórcze inhibitory wychwyty zwrotnego serotoniny, ponieważ nie byliśmy pewni jakości danych. Nie dysponowaliśmy danymi na temat kosztów. Gdybyśmy je mieli, nasze zalecenie mogłoby się zmienić.

### Podsumowanie

W naszym systemie oceny jakości danych i klasyfikacji siły zaleceń staraliśmy się zrównoważyć prostotę i jednoznaczność. Niezależnie od tego jak prosty albo skomplikowany jest system, zawsze wymaga oceny bilansu korzyści i szkodliwości oraz innych aspektów. Nasz system stanowi szkie-

let procesu podejmowania decyzji i może pomóc zagwarantować, by oceny w ramach tego procesu były właściwe, natomiast ich nie eliminuje.

#### Główne punkty

- Różne instytucje używają różnych systemów do oceny jakości danych i klasyfikacji siły zaleceń.
- Różnice i niedoskonałości tych systemów mogą wprowadzać w błąd i utrudniać skuteczne porozumiewanie się.
- W niniejszym artykule przedstawiono systematyczne i jasno sprecyzowane podejście do oceny jakości danych i klasyfikacji siły zaleceń.
- Podejście to uwzględnia metodologię i jakość przeprowadzenia badań oraz zgodność i odniesienie ich wyników w ocenie jakości danych dotyczących każdego z ważnych punktów końcowych.
- W ocenie siły zaleceń uwzględnia się bilans korzyści i szkodliwości, jakość danych, możliwość ich odniesienia do konkretnej sytuacji klinicznej oraz stopień pewności co do ryzyka podstawowego.

Osoby uczestniczące w przygotowaniu tego artykułu oraz źródki: [bmj.com](http://bmj.com)

**Konflikt interesów:** Większość członków Grupy Roboczej GRADE jest żywotnie zainteresowana innym systemem oceny jakości danych i klasyfikacji siły zaleceń.

Tłumaczył lek. Jan Brożek

Reprinted with permission of BMJ Publishing Group

### Piśmiennictwo

1. Thompson D.C., Rivara F.P., Thompson R.: Helmets for preventing head and facial injuries in bicyclists. *Cochrane Database Syst. Rev.*, 2000; (2): CD001855
2. North of England Evidence Based Guideline Development Project. Evidence based clinical practice guideline: the choice of antidepressants for depression in primary care. Newcastle upon Tyne, Centre for Health Services Research, 1997

## Komentarz

Zamieszczony w tym numerze „Medycyny Praktycznej” artykuł grupy GRADE jest wynikiem kilkuletniej pracy wielu autorów z Europy, Ameryki Północnej, Azji, Australii i Nowej Zelandii. Na co dzień pracują oni w różnych międzynarodowych lub krajowych instytucjach (p. ramka), których zadaniem – głównym lub jednym z ważniejszych – jest opracowywanie mniej lub bardziej kategorycznych zaleceń dla służby zdrowia (personelu medycznego, administratorów, ubezpieczycieli), dotyczących stosowania lub finansowania różnych procedur medycznych. Ich praca jest wieloetapowa i w odniesieniu do pojedynczego zalecenia polega w skrócie na:

- sprecyzowaniu problemu, którego zalecenie ma dotyczyć,
- zebraniu wszystkich dostępnych danych,
- ocenie jakości zgromadzonych danych (czyli określeniu stopnia ich wiarygodności),
- ilościowym zestawieniu danych,
- sformułowaniu na ich podstawie zalecenia postępowania, odzwierciedlającego jednocześnie stopień pewności, że efektem przestrzegania tego zalecenia będzie lepsza jakość opieki zdrowotnej.

Te różne osoby nawiązały współpracę w ramach grupy GRADE, gdy sobie uświadomiły, że ich zadania i napotykanne trudności w dużej mierze są podobne. Jasne się też stało, że koncepcje i techniczne aspekty rozwiązań tych problemów również są wspólne. W komentarzu do oryginalnej publikacji GRADE nie będziemy powtarzać rozwiązań zaproponowanych przez autorów. Jako uczestnicy prac nad systemem GRADE, reprezentujący Polski Instytut Evidence Based Medicine, chcielibyśmy się podzielić z Czytelnikami przemyśleniami ze spotkań grupy GRADE.

Po pierwsze: upewniliśmy się wzajemnie, że systemy opieki zdrowotnej i instytucje za nią odpowiedzialne w poszczególnych krajach borykają się w podobnym stopniu z problemami związanymi z koniecznością dokonywania wyborów i decydowania o zasadności finansowania różnorodnych procedur medycznych. Dotyczy to tak Polski, jak i bogatych Stanów Zjednoczonych Ameryki. Przyczyną takiej sytuacji są z jednej strony ograniczone środki przeznaczane na ochronę zdrowia (stanowiące określoną część budżetu każdego państwa), z drugiej zaś – stale rosnąca liczba nowych procedur medycznych (badań diagnostycznych i metod leczenia). Prowadzenie badań przesiewowych w kierunku różnych nowotworów, leczenie nadciśnienia tętniczego skurczowego u osób w podeszłym wieku czy wszczepianie au-

tomatycznych kardiowerterów-defibrylatorów – to tylko przykłady interwencji, które się pojawiły w ciągu kilku ostatnich lat. Dzisiaj pacjent po zawale serca wychodzi ze szpitala z zaleceniem przyjmowania leku przeciwplatekowego,  $\beta$ -bloкера, inhibitora konwertazy angiotensyny i leku hipolipemizującego, a przecież korzyści ze stosowania tych leków udokumentowano dopiero niedawno. Ponadto decyzja często dotyczy nie tylko tego „czy coś zrobić”, ale też „jak to zrobić” (np. wybór określonej techniki operacyjnej czy rodzaju leku hipotensyjnego). Zalew informacji dotyczących każdej szczegółowej sytuacji klinicznej trudno jest opanować.

Po drugie: okazało się, że w większości, jeśli nie we wszystkich krajach powstały już lub powstają zespoły, których zadaniem jest koordynacja działań zmierzających do obiektywnej selekcji i syntezy dostępnych informacji na potrzeby osób podejmujących decyzje w opiece zdrowotnej (personelu medycznego, administratorów, ubezpieczycieli i in.). Zespoły te pracują pod różnymi szyldami: epidemiologii klinicznej, wytycznych praktyki klinicznej, *evidence based medicine* (EBM), oceny technologii medycznych (*health technology assessment* – HTA), farmakoekonomiki itd., i w ramach różnych instytucji, zazwyczaj niezależnych. Przede wszystkim mają one wykształcić i zorganizować odpowiednią kadrę specjalistów potrafiących obiektywnie radzić sobie z ilością i jakością informacji, jak i z często sprzecznymi interesami grupowymi (np. grup zawodowych, płatników za świadczenia zdrowotne, producentów leków i sprzętu medycznego) – jest to jedna z najlepszych inwestycji dostępnych dla zarządzających służbą zdrowia. Wobec implikacji zdrowotnych i finansowych decyzji podejmowanych w opiece zdrowotnej każdy krok w kierunku racjonalizacji tych decyzji jest wręcz bezcenny.

Po trzecie: dyskusje podobne do tych w ramach spotkań grupy GRADE toczyły się lub toczą w wielu krajach i organizacjach międzynarodowych. Owocem tych dyskusji bywają rozwiązania uważane przez ich autorów za jedynie słuszne. Dyskusje w ramach GRADE unaocznily stopień przywiązania każdego z uczestników do własnych wcześniejszych rozwiązań. Zaakceptowanie cudzych pomysłów wymagało czasu, chęci porozumienia i kompromisów. Bez dobrej woli wszystkich zainteresowanych wspólne wypracowanie zasad podejmowania decyzji nie byłoby możliwe. Odnosi się to także do prac podobnych zespołów w poszczególnych krajach.

Naszym zdaniem najważniejszym wkładem grupy GRADE do metodologii tworzenia zaleceń jest jasne



rozgraniczenie dwóch etapów tego procesu. Pierwszy obejmuje sprecyzowanie problemu, którego zalecenie ma dotyczyć, zebranie dostępnych danych, określenie ich jakości, a następnie ich ilościowe zestawienie. Drugi etap polega na dokonaniu szeregu subiektywnych ocen (np. bilansu korzyści i szkodliwości związanych ze stosowaniem określonych metod leczenia) i ich syntezy w formie zalecenia, odzwierciedlającego także stopień przekonania autorów, że postępowanie zgodnie z tym zaleceniem jest wskazane i korzystne (siła zalecenia). Na tym etapie nie należy się spodziewać pełnej jednomyślności, ale raczej trzeba oczekiwać różnych stanowisk i propozycji. Wielokrotnie podczas spotkań grupy GRADE poszczególni uczestnicy przedstawiali problem i dane naukowe opisujące konsekwencje alternatywnych rozwiązań, spodziewając się, że wybór postępowania będzie jasny dla wszystkich; wielokrotnie się przekonaliśmy, że różne osoby mające do dyspozycji te same dane niejednokrotnie podejmują odmienne decyzje i uważają je za słuszne. Rozbieżności te wynikają nie tylko z uwarunkowań kulturowych, prawnych, religijnych czy finansowych, ale często też z osobistych przekonań, doświadczeń i preferencji. Dostrzegając, że dane wymagają subiektywnej interpretacji, grupa GRADE wzbogaciła nasze zrozumienie procesu tworzenia zaleceń i zmusiła niejako ich autorów do wzięcia odpowiedzialności za pozanaukowe przesłanki prowadzące do różnych decyzji i ujawnienia ich.

Powyżej opisaliśmy pokrótce to, co grupa GRADE już zrobiła. Obecnie zajmujemy się następnymi obszarami decyzyjnymi. Pierwszy z nich to zalecenia i wytyczne dotyczące zastosowania procedur diagnostycznych w sytuacjach, w których decyzje o ich wykonaniu opiera się na danych dotyczących zdolności do ustalenia określonego rozpoznania (a nie dalszych tego konsekwencji). Śledząc dyskusje na temat wykonywania pozytronowej tomografii emisyjnej, tomografii komputerowej całego ciała (reklamowanej na stacjach metra i przy autostradach w USA) czy mammografii u kobiet w różnych grupach wiekowych, uświadamiamy sobie jasno, że ustalenie diagnozy to tylko początek (i to niepewny) drogi do poprawy stanu zdrowia lub przedłużenia życia.

Drugi obszar zainteresowań grupy GRADE to etap procesu tworzenia zaleceń, na którym powinno się brać pod uwagę koszty związane z określonymi interwencjami oraz stopień, w jakim należy je uwzględnić. W niektórych krajach instytucjom tworzącym zalecenia ustawowo zabrania się uwzględniania kosztów, a w innych nakazuje się je brać pod uwagę.

Istnieją argumenty za jednym i drugim podejściem, a uważni Czytelnicy artykułu GRADE dostrzegą, że aspekty ekonomiczne zostały w nim potraktowane powierzchownie.

Wszystkie problemy wspomniane w tym komentarzu dotyczą również sytuacji w Polsce. Takie zadania jak decydowanie o refundacji leków czy o koszcie świadczeń gwarantowanych to nic innego, jak tworzenie zaleceń dotyczących poszczególnych procedur medycznych. Mamy nadzieję, że rozwiązania wypracowane przez grupę GRADE będą przydatne przy rozwiązywaniu problemów polskiego systemu opieki zdrowotnej.

**Roman Jaeschke MD MSc**

lek. Jan Brożek

dr med. Jacek Mrukowicz

Polski Instytut Evidence Based Medicine, Kraków

#### **Niektóre instytucje, w których pracują autorzy systemu GRADE**

- Australasian Cochrane Centre, Park Holm, Australia
- Centers for Disease Control and Prevention, Atlanta, USA
- Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, Rockville, USA
- Centre for Evidence-based Medicine, Oxford, Wielka Brytania
- Centre for Health Services Research, University of Newcastle upon Tyne, Newcastle, Wielka Brytania
- Centro per la Valutazione della Efficacia della Assistenza Sanitaria (CeVEAS), Modena, Włochy
- Department of Health Services Research, Norwegian Centre for Health Services, Oslo, Norwegia
- Departments of Clinical Epidemiology and Biostatistics and Medicine, McMaster University, Hamilton, Kanada
- Fédération Nationale des Centres de Lutte Contre le Cancer, Lyon, Francja
- Finnish Medical Society Duodecim, Helsinki, Finlandia
- German Cochrane Centre, Freiburg, Niemcy
- Global Programme on Evidence for Health Policy, World Health Organisation, Genewa, Szwajcaria
- National Institute for Clinical Excellence, Londyn, Wielka Brytania
- Polski Instytut Evidence Based Medicine, Kraków, Polska
- Scottish Intercollegiate Guidelines Network, Edynburg, Wielka Brytania