

Od danych naukowych do praktycznych zaleceń – tworzenie wytycznych według metodologii GRADE

Wiktoria Leśniak¹, Małgorzata M. Bała^{2,3}, Roman Jaeschke⁴, Jan L. Brożek^{5,6}

1 Pracownia Podejmowania Decyzji Klinicznych, II Katedra Chorób Wewnętrznych, Uniwersytet Jagielloński, Collegium Medicum, Kraków

2 Ośrodek Przeglądów Systematycznych – Polska Filia Cochrane, Uniwersytet Jagielloński, Collegium Medicum, Kraków

3 Zakład Higieny i Dietetyki, Uniwersytet Jagielloński, Collegium Medicum, Kraków

4 Department of Medicine and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Kanada

5 Department of Clinical Epidemiology & Biostatistics and Division of Immunology & Allergy, Department of Medicine, McMaster University, Hamilton, Ontario, Kanada

6 University of Freiburg, Medical Informatics and Clinical Epidemiology, Institute for Medical Biometry and Statistics, Freiburg im Breisgau, Niemcy

SŁOWA KLUCZOWE

punkty końcowe, siła zaleceń, wiarygodność danych, wytyczne praktyki klinicznej

STRESZCZENIE

Wytyczne praktyki klinicznej to zbiór zaleceń, sformułowanych w sposób systematyczny, które pomagają lekarzom i innym pracownikom instytucji opieki zdrowotnej w podejmowaniu optymalnych decyzji w opiece nad pacjentami w określonych okolicznościach. Odgrywają również ważną rolę w polityce zdrowotnej. Wiele towarzystw naukowych i innych organizacji publikuje co roku dokumenty i nazywa je wytycznymi. Nie wszystkie tworzone są rzetelnie, w systematyczny sposób. W artykule przedstawiono proces, według którego powinny powstawać wytyczne, aby zwiększyć prawdopodobieństwo ich wiarygodności i przydatności. Metodologię tego procesu opracowała i ciągle doskonali międzynarodowa grupa metodologów zwana GRADE. Rola GRADE polega głównie na zdefiniowaniu pytań klinicznych, wybraniu punktów końcowych i przypisaniu im wagi, ocenie jakości danych i zinterpretowaniu tych danych (głównie w kontekście oszacowanych efektów względnych i bezwzględnych) oraz na uzasadnieniu siły zalecenia (czyli stopnia pewności, że dane zalecenie należy stosować).

Co to są wytyczne praktyki klinicznej Wytyczne praktyki klinicznej (*clinical practice guidelines*) definiuje się jako zbiór stwierdzeń (zaleceń), sformułowanych w sposób systematyczny, które mają pomagać lekarzom (a także innym pracownikom instytucji opieki zdrowotnej) w podejmowaniu optymalnych decyzji w opiece nad pacjentami w określonych okolicznościach.¹ Wytyczne można porównać do rady udzielanej przez bardziej doświadczoną osobę, osobie mniej doświadczonej, np. przez bardziej doświadczonego lekarza młodszemu koledze. Różnica polega na tym, że w wytycznych grupa bardziej doświadczonych ekspertów udziela porady dużej grupie mniej doświadczonych lekarzy, pielęgniarek i innych pracowników opieki zdrowotnej.

Wytyczne mają na celu:

- skoncentrowanie ogromnej wiedzy medycznej w poręcznej, łatwej do stosowania formie
- pomóc pracownikom opieki zdrowotnej i pacjentom w podjęciu najważniejszej w danej

sytuacji decyzji odnośnie do wyboru testów diagnostycznych, profilaktyki lub leczenia

- poprawę jakości opieki nad pacjentami i zwiększenie efektywności wykorzystania ograniczonych zasobów.

Kilka słów o historii, czyli od GOBSAT do GRADE

W 1990 roku prof. Gordon Guyatt z uniwersytetu McMaster w Kanadzie użył po raz pierwszy terminu „*evidence-based medicine*” do określenia zasad praktykowania medycyny opartego na podejmowaniu decyzji klinicznych na podstawie znajomości i zrozumienia piśmiennictwa medycznego (albo wiedzy, że nie ma opublikowanych danych naukowych na określony temat). Celem takiego podejścia było świadome podejmowanie decyzji, wiedząc na jakich danych naukowych opiera się decyzja, na ile dane te są wiarygodne i jakie wnioski można z nich wyciągnąć. Stosowanie tej strategii wymagało jasnego sformułowania odpowiednich pytań, dokładnego przeszukania

Adres do korespondencji:
dr med. Wiktoria Leśniak, Pracownia
Podejmowania Decyzji Klinicznych,
II Katedra Chorób Wewnętrznych,
Uniwersytet Jagielloński, Collegium
Medicum, ul. Skawińska 8,
31-066, Kraków, tel.: 12 430 52 66,
e-mail: lesniak@mp.pl
Praca wpłynęła: 30.10.2015.
Przyjęta do druku: 30.10.2015.
Pol Arch Med Wewn. 2015;
125 (Special Issue): 26-41
Copyright by Medycyna Praktyczna,
Kraków 2015

piśmiennictwa odnoszącego się do tych pytań, krytycznej oceny danych i możliwości ich odniesienia do określonej sytuacji klinicznej oraz wyważonego zastosowania wyciągniętych wniosków do rozwiązania problemu klinicznego.² Początkowo zwolennicy tego poglądu uważali, że każdy lekarz powinien być biegły w wyszukiwaniu, krytycznej analizie i praktycznym wykorzystaniu danych z piśmiennictwa naukowego. Z czasem przekonali się, że pracownicy opieki zdrowotnej mieli zbyt mało czasu, aby opanować odpowiednie umiejętności, a nawet gdy je opanowali – aby je wykorzystywać w codziennej praktyce. Okazało się także, że do zapewnienia opieki klinicznej o wysokiej jakości wystarczy umiejętność znalezienia i zrozumienia odpowiednich podsumowań danych naukowych, a nawet samych zaleceń na nich opartych.³ Zrodziła się więc m.in. potrzeba tworzenia wytycznych postępowania, będących źródłem wiedzy dla praktykujących lekarzy, innych pracowników opieki zdrowotnej i pacjentów. Takie wytyczne miały się różnić od pojedynczych badań i podsumowań wielu badań naukowych (np. przeglądów systematycznych) tym, że oprócz syntezy dostępnej wiedzy miały zawierać wskazówki (porady), jak postępować w określonej sytuacji klinicznej. Już wtedy zdawano sobie sprawę z tego, że aby uniknąć udzielania błędnych porad (zaleceń) proces tworzenia wytycznych powinien być usystematyzowany.⁴

Niestety większość wytycznych stworzonych w latach 90. ubiegłego wieku przez towarzystwa naukowe nie spełniała podstawowych kryteriów wiarygodności.⁵ W jednym badaniu, na 431 ocenionych wytycznych, w 67% nie wymieniono osób uczestniczących w ich tworzeniu, w 88% nie podano informacji na temat wyszukiwania badań, a w 82% nie stopniowano siły zaleceń. Wszystkie 3 kryteria spełniało tylko 5% dokumentów.⁶ W innym badaniu na 279 wytycznych niektóre ze standardów metodologicznych spełniało około 51%.⁷

W związku z tym, że wytyczne nie spełniały kryteriów jakości, co jest niezbędne, aby mogły osiągnąć swoje cele, potrzebne było opracowanie zasad ich tworzenia. Dlatego też, m.in., międzynarodowy zespół badaczy i autorów wytycznych (Appraisal of Guidelines Research and Evaluation [AGREE] Collaboration) opracował zestaw kryteriów jakości wytycznych – tzw. instrument AGREE, który pośrednio był źródłem wskazówek, jak powinny wyglądać dobre wytyczne (p. niżej).^{8,9} Nawet jednak, jeśli w wytycznych oceniano jakość danych i podawano siłę zaleceń, różne zespoły ekspertów używały różnych skal i symboli, co utrudniało użytkownikom korzystanie z wytycznych. Aby to zmienić, Gordon Guyatt i Andy Oxman zainicjowali dyskusję, która w 2000 roku doprowadziła do utworzenia Grupy Roboczej GRADE (Grading of Recommendations Assessment, Development and Evaluation), mającej opracować jednolity schemat tworzenia wiarygodnych wytycznych.^{1,10} Schemat taki powstał i jest ciągle udoskonalany (p. niżej). W 2008 roku grupa ta liczyła 30 członków, obecnie około 500.

Przyzwyczajenie jest drugą naturą człowieka, dlatego początkowo wiele grup ekspertów nie mogło zrozumieć, że powstawanie wytycznych opartych jedynie na ich opiniach (zgodnie z zasadą GOBSAT – Good Old Boys Sat Around a Table) nie jest już zalecane, ponieważ stwierdzenia bazujące jedynie na opinii klinicznej, doświadczeniu i wyselekcjonowanych doniesieniach naukowych, a nieuwzględniające wszystkich wiarygodnych i aktualnych danych naukowych, mogą być wypaczone, np. przez konflikt interesów.⁵

Z czasem jednak coraz liczniejsze organizacje przyjmowały system GRADE, obecnie jest ich ponad 90, a wśród nich:

- 1 towarzystwa naukowe (m.in. American College of Chest Physicians, American Thoracic Society, European Respiratory Society, Endocrine Society, Society of Critical Care Medicine, American College of Physicians, Canadian Cardiovascular Society)
- 2 organizacje międzynarodowe (World Health Organization [WHO], Cochrane Collaboration)
- 3 międzynarodowe grupy ekspertów opracowujące wytyczne (np. Kidney Disease: Improving Global Outcome, Surviving Sepsis Campaign)
- 4 agencje HTA i narodowe ośrodki tworzące wytyczne (np. National Institute for Health and Clinical Excellence, Scottish Intercollegiate Guidelines Network)
- 5 redakcje podręczników (np. UpToDate)
- 6 wydawcy (np. BMJ).

W tym roku obchodzimy 15. rocznicę nieprzerwanej, aktywnej działalności grupy GRADE. Warto wspomnieć, że jedno ze spotkań roboczych tej grupy odbyło się w Krakowie w 2003 roku i było współorganizowane przez Medycynę Praktyczną.

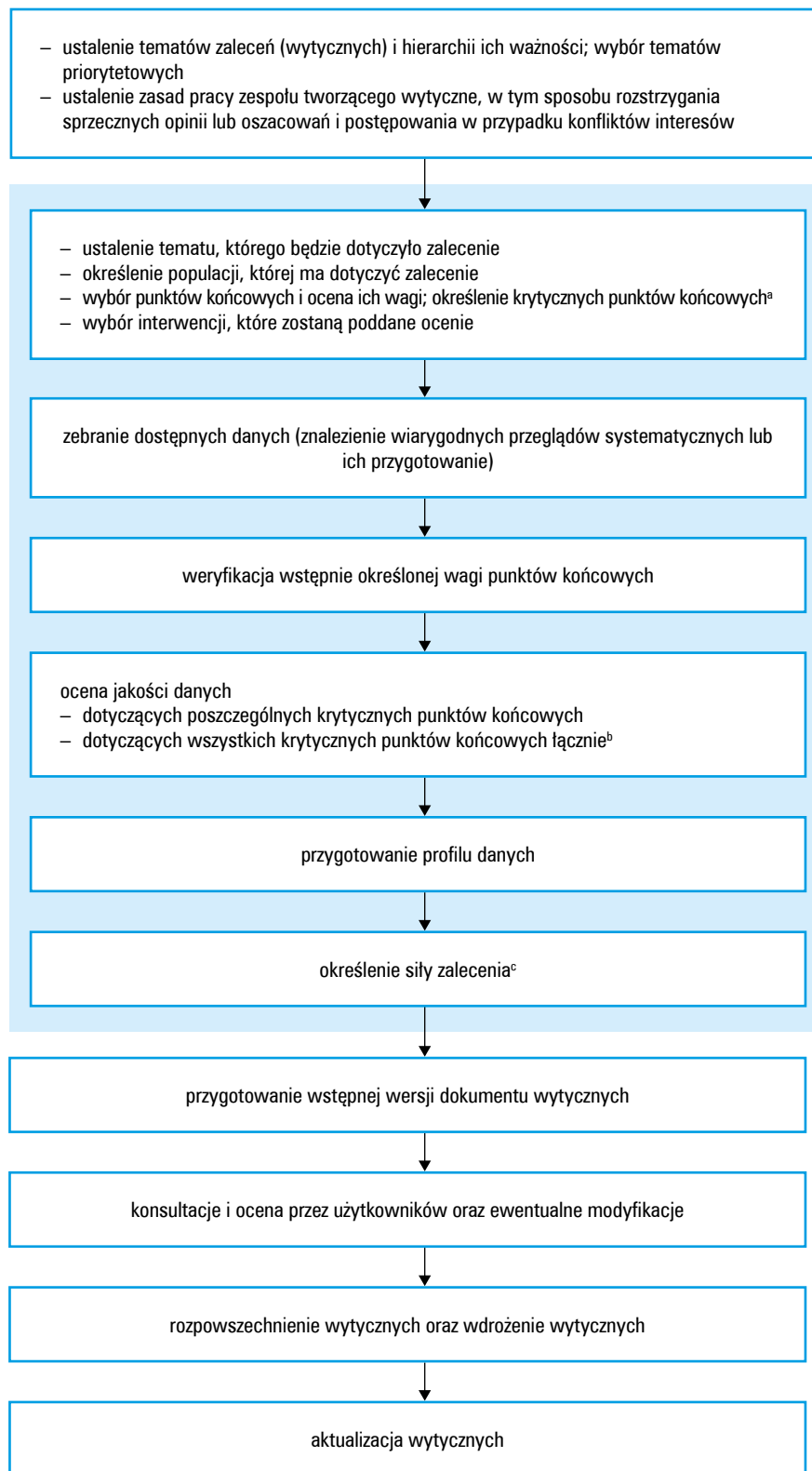
Metodologia GRADE, czyli jak powinny powstawać wytyczne

Poniżej opisano kolejne etapy tworzenia wytycznych zgodnie z GRADE. Schematycznie przedstawiono je na **RYCINIE 1**.¹¹ Autorzy przeglądów systematycznych i wytycznych praktyki klinicznej muszą sobie zdawać sprawę z tego, że nie do wszystkich etapów tego procesu odnosi się bezpośrednio system GRADE. Rola GRADE polega głównie na zdefiniowaniu pytań, wybraniu punktów końcowych i przypisaniu im wagi, ocenie jakości danych i zinterpretowaniu tych danych, głównie w kontekście oszacowanych efektów względnych i bezwzględnych (p. niżej).

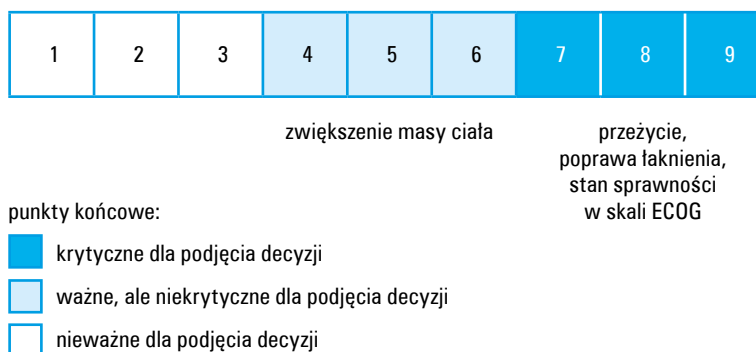
Ustalenie zakresu wytycznych i wybór zespołu, który je będzie opracowywał

Proces tworzenia wytycznych rozpoczyna się od ustalenia zakresu wytycznych (np. rozpoznawanie, leczenie i zapobieganie infekcyjnemu zapaleniu wsierdza, postępowanie w ostrych zespołach wieńcowych bez uniesienia odcinka ST, rozpoznawanie i leczenie cukrzycy), hierarchii ważności poszczególnych szczegółowych tematów i wyboru tematów priorytetowych. Następnie należy ustalić, kto będzie wchodził w skład zespołu (panelu) przygotowującego wytyczne. Ważne, aby ten zespół był zróżnicowany pod względem doświadczenia

RYCINA 1 Etapy tworzenia wytycznych (na podstawie Guyatt i wsp.,¹¹ zmodyfikowane)



- a** punkty końcowe mające krytyczne znaczenie dla podjęcia decyzji o zalecaniu albo niezalecaniu określonego postępowania
- b** zwykle determinowana przez najniższą jakość danych dotyczących któregośkolwiek z krytycznych punktów końcowych
- c** odzwierciedla stopień pewności autorów zalecenia, że jego przestrzeganie przyniesie więcej skutków korzystnych niż niekorzystnych; uwzględnia jakość danych, bilans skutków korzystnych i niekorzystnych oraz wartości i preferencje
- Uwaga: na niebieskim tle umieszczono etapy, które opisuje metodologia GRADE



i aby poszczególni członkowie uzupełniali się pod względem potrzebnych kompetencji, gdyż tylko to gwarantuje, a przynajmniej zwiększa prawdopodobieństwo stworzenia optymalnych zaleceń. Dlatego obecnie takie zespoły są coraz liczniejsze i obejmują metodologów, lekarzy różnych specjalności, innych pracowników opieki zdrowotnej, przedstawicieli pacjentów i przedstawicieli płatnika. Przy dużych zróżnicowanych zespołach istnieje ryzyko problemów związanych ze sprzecznymi interesami i opiniami. Ważne jest więc, aby przed rozpoczęciem prac ustalić zasady działania, sposoby rozwiązywania sprzecznych opinii i dochodzenia do konsensusu oraz zasady postępowania w przypadku konfliktów interesów.¹²

Ustalenie tematów zaleceń i pytań klinicznych Następnie należy sformułować pytania kliniczne, na które zalecenia w wytycznych miałyby odpowiedzieć, ze zdefiniowanymi:

- 1 populacją – chorymi, których będzie dotyczyć zalecenie
- 2 interwencjami, czyli dostępnymi metodami postępowania oraz
- 3 skutkami zdrowotnymi (w badaniach klinicznych nazywane punktami końcowymi), wiążącymi się z tymi interwencjami.¹³

Ważne jest, by sprecyzować, z czym porównywana jest dana interwencja. Szczególnie jeśli ocenia się skuteczność kilku interwencji (np. kilku leków), autorzy wytycznych powinni jednoznacznie określić, czy wszystkie te interwencje są zalecane w takim samym stopniu, czy któreś z nich są korzystniejsze od pozostałych.¹⁴

W pytaniach klinicznych zadawanych w wytycznych należy również uściślić, gdzie (w jakim kraju) wytyczne będą stosowane. Zalecenia, których można przestrzegać w krajach rozwiniętych, mogą być niewykonalne w krajach biednych, rozwijających się.

Ocena wagi punktów końcowych Na wstępie należy też ustalić, które skutki zdrowotne trzeba uwzględnić przy podejmowaniu decyzji o zalecaniu albo niezalecaniu danego postępowania.

Tę ocenę wagi punktów końcowych należy przeprowadzić przed przystąpieniem do przeglądu systematycznego dostępnych danych naukowych, a następnie powtórzyć, gdy przegląd jest już gotowy.

Według GRADE klasyfikuje się punkty końcowe w skali 1–9 (RYCINA 2), w której:

- 1 7–9 oznacza punkty końcowe krytyczne dla podjęcia decyzji
- 2 4–6 ważne dla podjęcia decyzji
- 3 1–3 punkty końcowe o małej wadze dla podjęcia decyzji.¹⁴

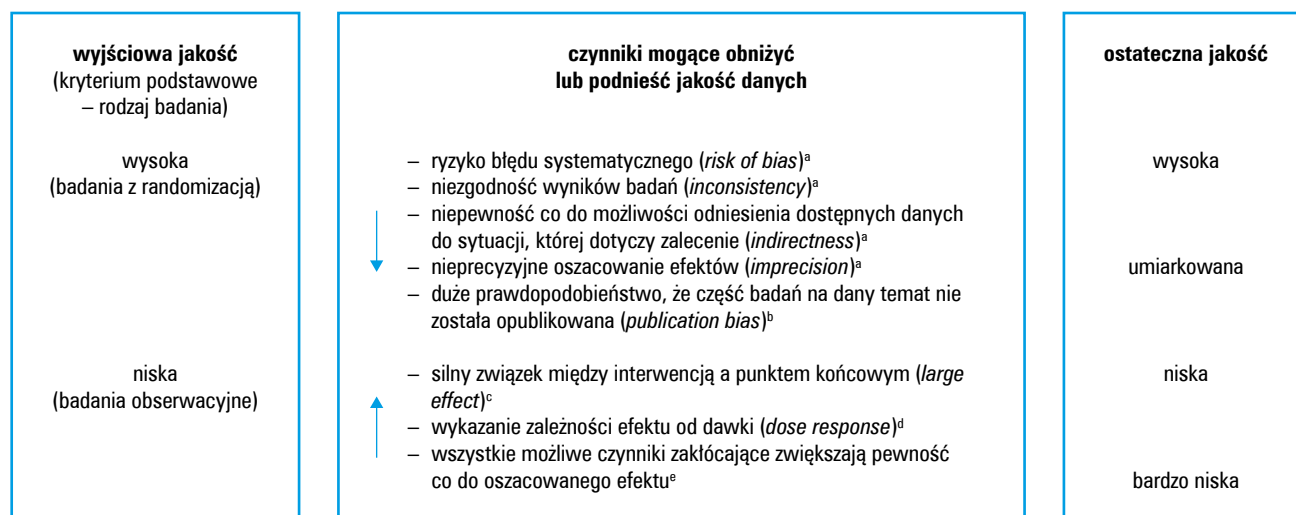
Uszeregowanie punktów końcowych według ich znaczenia pozwala autorom wytycznych na skupienie się na tych najważniejszych.

Ta ocena wagi punktów końcowych może zależeć od uwarunkowań kulturowych, jak również od pełnionej roli (pacjent, lekarz czy przedstawiciel płatnika). Panel ekspertów opracowujący wytyczne musi podjąć decyzję, czyja perspektywa zostanie uwzględniona. Zwykle powinna to być perspektywa grupy, do której są kierowane konkretne wytyczne.

Ocena wagi poszczególnych punktów końcowych może ulec zmianie po wykonaniu przeglądu systematycznego dostępnych danych.¹⁴ Może dojść do następujących sytuacji, gdy:

- 1 punkt końcowy oceniony wstępnie jako krytyczny (np. zgon z jakiegokolwiek przyczyny) może stać się mniej ważny, jeśli nie ma na jego temat danych, a korzystny wpływ na inne krytyczne punkty końcowe (np. zgon z przyczyn sercowych) jest wystarczający do podjęcia decyzji o stosowaniu danej interwencji. Ważne jest jednak, aby autorzy wytycznych byli pewni, że w opisywanej przykładowej sytuacji dana interwencja nie zwiększa ryzyka zgonu z jakiegokolwiek przyczyny, a zmniejszenie częstości zgonu z przyczyn sercowych było na tyle duże, że uzasadnia stosowanie interwencji mimo nieznanego wpływu na częstość zgonu z jakiegokolwiek przyczyny
- 2 w przypadku nowej, niestosowanej wcześniej interwencji, początkowo jej działania niepożądane mogą nie być znane i dopiero z czasem na podstawie pojawiających się danych staje się jasne, że określone szkodliwe działanie jakiegoś leku jest krytyczne, aby podjąć decyzję o jego niestosowaniu.¹⁴

Zebranie dostępnych danych Przy tworzeniu wytycznych postępowania konieczne jest znalezienie już istniejącego aktualnego i wiarygodnego przeglądu systematycznego dostępnych danych albo przeprowadzenie takiego przeglądu (odrębnie



- a** w przypadku poważnego (dużego) wpływu tego czynnika obniżenie jakości o 1 poziom; w przypadku bardzo poważnego (bardzo dużego) wpływu – o 2 poziomy
- b** w przypadku dużego prawdopodobieństwa tego błędu obniżenie jakości o 1 poziom; w przypadku bardzo dużego prawdopodobieństwa – o 2 poziomy (rzadko)
- c** w przypadku silnego związku podniesienie jakości o 1 poziom; w przypadku bardzo silnego związku – o 2 poziomy
- d** podniesienie jakości o 1 poziom
- e** tzn. wszystkie możliwe czynniki zakłócające zmniejszyłyby wielkość oszacowanego korzystnego efektu lub wskazywałyby na niekorzystny efekt w sytuacji, gdy nie zaobserwowano różnicy; podniesienie jakości o 1 poziom

RYCINA 3 Ocena jakości danych według systemu GRADE (na podstawie Guyatt i wsp.,¹¹ zmodyfikowane)

dla każdego pytania klinicznego; uwzględniającego wszystkie ważne punkty końcowe).¹³ Przegląd systematyczny z metaanalizą badań pierwotnych stanowi podstawę oszacowania wielkości wpływu danego postępowania na skutki zdrowotne ważne dla pacjentów i określenia jakości danych wspierających określone zalecenie. Niekiedy, aby sformułować jedno zalecenie potrzebny jest więcej niż 1 przegląd systematyczny. Nie we wszystkich badaniach są oceniane wszystkie ważne punkty końcowe. Często zdarza się, że skuteczność interwencji ocenia się w badaniach z randomizacją (*randomized controlled trials* – RCT), a bezpieczeństwo (częstość działań niepożądanych) w badaniach obserwacyjnych. Tymczasem publikowane przeglądy systematyczne ograniczają się często do RCT. Dlatego, aby podsumować wszystkie dostępne dane dla wszystkich ważnych punktów końcowych, konieczne jest sięgnięcie do kilku przeglądów systematycznych lub przeprowadzenie własnego lub własnych przeglądów systematycznych.¹¹

Zdarza się, że nie ma badań, w których oceniano punkty końcowe uznane przez zespół opracowujący wytyczne za ważne. W takiej sytuacji można się posłużyć badaniami oceniającymi zastępcze punkty końcowe (zastępujące te, których brakuje), ale jakość takich danych jest mniejsza (ze względu na niepewność co do możliwości uogólnienia tych danych – p. niżej).

Ocena jakości danych Podczas tworzenia wytycznych kolejnym etapem jest ocena jakości danych dotyczących każdego ważnego punktu końcowego. Jakość danych świadczy o tym, jaka jest nasza pewność, że oszacowanie efektu jest prawdziwe

(właściwe) i może być wsparciem dla konkretnego zalecenia. Przy ocenie jakości danych bierze się pod uwagę ogólną metodykę każdego badania (np. badanie z randomizacją vs badanie obserwacyjne), która decyduje o wyjściowej jakości danych i dodatkowo uwzględnia się inne czynniki mogące obniżyć lub podnieść jakość danych. Z **RYCINY 3** wynika, iż na jakość tę wpływa nie tylko ryzyko błędu systematycznego, ale i inne czynniki.¹⁵ Omówiono je szczegółowo w podrozdziałach poniżej.

W systemie GRADE podkreśla się rolę przejrzystości w procesie oceny jakości danych. Opinie formułujących zalecenia muszą się w czytelny sposób opierać na dostępnych danych naukowych. Zarówno podnoszenie, jak i obniżanie wyjściowej jakości danych należy uzasadnić i udokumentować.¹⁵ Nie oznacza to, że GRADE gwarantuje jednakową ocenę jakości danych przez różnych oceniających. Interpretacja jakości danych może być różna. Przywiązywanie niewystarczającej uwagi do jakości danych zwiększa ryzyko sformułowania niewłaściwych zaleceń, co odbije się niekorzystnie na zdrowiu pacjentów.

Arbitralne klasyfikowanie jakości danych do 4 kategorii jest pewnym ograniczeniem i uproszczeniem. Ale korzyści w postaci prostoty i przejrzystości takiego rozwiązania przewyższają jego ograniczenia.

W ramach GRADE oddzielnie traktuje się ocenę jakości danych oraz formułowanie zaleceń. Siła zalecenia zależy bowiem nie tylko od jakości danych (p. niżej). Dane o niskiej lub bardzo niskiej jakości mogą prowadzić do silnego zalecenia, a dane o wysokiej jakości mogą stanowić podstawę do sformułowania słabego zalecenia.

Czynniki obniżające jakość danych *Ograniczenia badań – ryzyko błędu systematycznego* Główne ograniczenia badań z randomizacją (*study limitations [risk of bias]*), które zwiększają ryzyko błędu systematycznego obejmują:¹⁶

- 1 brak utajnienia randomizacji
 - 2 brak zaślepienia – jego znaczenie jest większe w przypadku subiektywnych punktów końcowych (takich jak np. jakość życia, natężenie bólu) niż przy obiektywnych (np. zgon, zawał serca, udar mózgu, krwawienie)
 - 3 duży odsetek chorych, których utracono z obserwacji (*lost to follow-up*)
 - 4 niezastosowanie analizy zgodnie z zaplanowanym leczeniem (*intention-to-treat analysis*)
 - 5 wcześniejsze niż zaplanowano przerwanie badania z powodu zaobserwowanej korzyści – z dostępnych danych wynika, że w takich badaniach efekt leczenia jest zawyżony, szczególnie jeśli przed zakończeniem badania zarejestrowano mniej niż 500 zdarzeń.¹⁷ W przypadku przeglądów systematycznych, do których włączono dużo badań zakończonych przedwcześnie, powinno się przeprowadzić analizę wrażliwości polegającą na objęciu analizą wszystkich badań, w tym tych zakończonych wcześniej oraz tylko tych badań, których nie przerwano wcześniej. Jeśli oszacowane wyniki różnią się między tymi 2 sytuacjami, powinno się uwzględnić tylko badania, które się nie zakończyły wcześniej, jako bardziej wiarygodne
 - 6 selektywne raportowanie wyników dla poszczególnych zaplanowanych punktów końcowych. Aby ocenić występowanie tego ograniczenia, konieczne jest sprawdzenie protokołu badania.
- Główne ograniczenia badań obserwacyjnych obejmują:¹⁶

- 1 niewłaściwy dobór grupy kontrolnej
- 2 nieprzyjęcie odpowiedniej poprawki na nierównowagę czynników prognostycznych.

Ryzyko błędu systematycznego może być różne w przypadku oceny różnych punktów końcowych (np. może być wiadomo, ilu chorych przeżyło 1 rok – ocena ryzyka zgonu, a liczba chorych, u których po roku od zatrzymania krążenia oceniono stan neurologiczny może być mniejsza).

Przy podejmowaniu decyzji, czy obniżyć jakość danych ze względu na ograniczenia metodologii badań autorzy wytycznych nie powinni podawać średniej jakości (wypadkowej) dla wszystkich badań, lecz raczej włączyć do przeglądu systematycznego jedynie badania, w których ryzyko błędu systematycznego jest mniejsze.¹⁶

Niezgodność wyników badań Należy pamiętać, że różnica ryzyka między grupą z interwencją i grupą kontrolną (bez badanej interwencji) – *risk difference*, zwana też *absolute risk reduction* może się różnić między poszczególnymi badaniami (w których ryzyko podstawowe, czyli to w grupie kontrolnej często znacznie się różni). Wtedy nie mówimy o niezgodności wyników, lecz oczekujemy, że autorzy wytycznych wyodrębnią podgrupy i sformułują różne zalecenia dla każdej.

Parametry opisujące względny efekt interwencji (takie jak ryzyko względne [*relative risk* – RR], hazard względny [*hazard ratio* – HR], iloraz szans [*odds ratio* – OR]) zwykle są podobne, nawet w populacjach różniących się pod względem ryzyka podstawowego. Jeśli więc różnice między wynikami wyrażonymi przez parametry względne znacznie się różnią i nie są znane przyczyny tych różnic, autorzy wytycznych powinni obniżyć jakość danych, zwłaszcza gdy wyniki niektórych badań sugerują korzyść danej interwencji, a innych – nie występowanie efektu lub efekt szkodliwy. Mniejsze przesłanki do obniżenia jakości danych występują, gdy wyniki poszczególnych badań różnią się pod względem wielkości zaobserwowanego efektu (np. różnego stopnia korzyści).¹⁸

Różnice te są widoczne na etapie tworzenia przeglądu systematycznego i mogą wynikać z tego, że w badaniach zadano inne pytanie kliniczne, czyli nieco inne są populacje (np. inne zaawansowanie choroby, choroby współistniejące), inne interwencje (np. inne leki z tej samej grupy, inna dawka leku, inna droga podania, interwencje dodatkowe), inne oceniane punkty końcowe (inne kryteria ich rozpoznania, inny czas obserwacji) lub inna była metodologia badania (np. badania z randomizacją mniej [większe ryzyko błędu systematycznego] lub bardziej wiarygodne). W przypadku, gdy różnice w pytaniu klinicznym stanowią przyczynę niezgodności wyników, autorzy wytycznych powinni sformułować inne zalecenia dla różnych podgrup chorych lub podgrup interwencji. Jeśli natomiast różnice wynikają z innej metodologii badań, autorzy wytycznych powinni podejmować decyzję na podstawie oszacowania efektu interwencji z badań bardziej wiarygodnych.¹⁸

Kryteria uwzględniane przy podejmowaniu decyzji o zmniejszeniu jakości danych z powodu niezgodności wyników (*inconsistency*) to:¹⁸

- 1 oszacowane w badaniach wyniki znacznie różnią się od siebie
- 2 przedziały ufności (*confidence intervals [CI]*) dla oszacowanych efektów z poszczególnych badań nie zachodzą na siebie wcale lub w niewielkim stopniu
- 3 wartość *p* dla testów statystycznych oceniających niejednorodność wyników (*tests for heterogeneity*) badań włączonych do przeglądu systematycznego jest mała; wynik tych testów może zależeć od liczby włączonych badań
- 4 wartość I^2 (parametr oceniający, jaka część różnic między badaniami włączonymi do przeglądu systematycznego)¹⁹ jest duża; wielkość I^2 zależy w pewnym stopniu od wielkości próby we włączonych badaniach (jeśli badania są duże, to nawet niewielka różnica ich wyniku może powodować dużą wartość I^2).

Niepewność co do możliwości odniesienia wyników do sytuacji, której dotyczy zalecenie (indirectness) Jakość danych jest mniejsza, gdy autorzy wytycznych nie mają pewności, czy można odnieść

wyniki dostępnych badań do sytuacji, której dotyczy zalecenie (*indirectness*). Można wyróżnić 4 takie sytuacje, gdy:

- 1 istnieją zasadnicze różnice między populacjami włączonymi do badań a tymi, których dotyczą wytyczne
- 2 istnieją zasadnicze różnice między interwencją testowaną w badaniu a tą, której dotyczą wytyczne
- 3 istnieją zasadnicze różnice między punktami końcowymi, które oceniano w badaniach a tymi, które interesują autorów wytycznych (ważnymi dla pacjentów – *patient-important outcomes*)
- 4 nie ma badań bezpośrednio porównujących 2 interwencje (*head-to-head comparisons*). Wtedy korzysta się z pośrednich porównań (*indirect comparisons*), np. oceniając różnicę efektu interwencji A w porównaniu z B, wykorzystuje się dostępne dane dotyczące porównania interwencji A z C i interwencji B z C.²⁰ Te sytuacje występują często, ale nie zawsze się je uwzględnia, oceniając jakość danych.

Uważa się, że nie należy obniżać jakości danych w przypadku różnic w populacjach, jeśli nie ma wystarczających przesłanek patofizjologicznych, pozwalających sądzić, że wielkość efektu interwencji w badanej populacji będzie się znacząco różnić w porównaniu z populacją, której dotyczy zalecenie.²⁰

W przypadku różnic w interwencjach obniża się jakość danych, jeśli różnice te mogły wpłynąć na częstość zaobserwowanych punktów końcowych (np. skomplikowany zabieg operacyjny przeprowadzany w ramach badania w wyspecjalizowanym ośrodku może skutkować większym prawdopodobieństwem przeżycia przy małej częstości powikłań, a przeprowadzenie go przez chirurga bez odpowiedniego doświadczenia może mieć niekorzystne następstwa). Trudno też uogólniać wyniki badania, w którym np. zastosowano złożone interwencje psychospołeczne, trudne do dokładnego powtórzenia w innym miejscu.

W przypadku punktów końcowych największy problem stanowi ocenianie w badaniach tylko zastępczych punktów końcowych (np. glikemii, gęstości mineralnej kości, wysokości ciśnienia tętniczego, stężenia cholesterolu LDL, stężenia kreatyniny, saturacji krwi) zamiast punktów końcowych ważnych dla pacjenta (takich jak np. odpowiednio powikłania cukrzycowe, złamania, udar mózgu, zawał serca, konieczność dializoterapii, zgon). Wymaga to zawsze obniżenia jakości danych przez autorów wytycznych. Ewentualnie można nie obniżać jakości danych jedynie wtedy, gdy w wielu badaniach z randomizacją wykazano konsekwentnie silny związek między zmianami danego zastępczego punktu końcowego a częstością punktu końcowego ważnego dla pacjenta dla leków z tej samej grupy, co oceniano (np. można nie obniżać jakości danych, jeśli skuteczność nowej statyny oceniano na podstawie zmian stężenia cholesterolu LDL, zamiast częstości incydentów wieńcowych, ponieważ przeprowadzono wcześniej wiele badań z użyciem różnych statyn,

dla których wykazano zmniejszenie częstości incydentów wieńcowych; nawet tutaj jednak istnieje pewne ryzyko nieprawidłowego rozumowania).²⁰

Trudno też uogólniać wyniki badania, gdy częstość punktu końcowego oceniano w krótkim okresie (np. zgon w okresie okołozabiegowym), a autorów wytycznych interesuje długotrwały wpływ interwencji na rokowanie.

W przypadku porównań pośrednich również zaleca się obniżenie jakości danych.²⁰

Szczególną sytuacją jest ocena jakości danych z metaanalizy sieciowej (*network meta-analysis* – NMA), nazywanej również metaanalizą wielu metod leczenia (*multiple-treatment meta-analysis*) lub porównaniem różnych metod leczenia (*mixed-treatment comparison*). Ta nowa metoda analizy statystycznej umożliwia jednoczesne porównanie wielu interwencji w jednym przeglądzie systematycznym i łączy bezpośrednio i pośrednie porównania różnych metod leczenia ze wszystkich badań.²¹ Ocena jakości danych z metaanalizy sieciowej jest trudna i wieloetapowa, wymaga osobnej oceny jakości danych dla porównań bezpośrednich i pośrednich. Grupa GRADE sugeruje, żeby jako ostateczną jakość wszystkich danych z metaanalizy sieciowej przyjąć wyższą wartość (czyli np. umiarkowaną jakość danych, jeśli jakość danych z porównań bezpośrednich jest umiarkowana, a z porównań pośrednich niska).²²

Nieprecyzyjne oszacowanie efektów Do oceny precyzji oszacowania efektu wykorzystuje się 95% przedział ufności dla różnicy efektu między grupą poddaną interwencji badanej a grupą kontrolną dla każdego punktu końcowego.

Autorzy wytycznych powinni obniżyć jakość danych ze względu na nieprecyzyjność oszacowania wyniku (*imprecision*), jeśli decyzja kliniczna byłaby różna w zależności od tego, czy rzeczywisty wynik znajdowałby się na górnej czy dolnej granicy przedziału ufności.²³ Jeżeli przedział ufności nie jest wystarczająco wąski, obniża się zwykle jakość danych o 1 poziom, a jeśli jest bardzo szeroki o 2 poziomy. Przy formułowaniu zaleceń, np. co do stosowania danego leku wymaga się zwykle większego efektu i większej precyzji oszacowania wyniku, jeśli ta interwencja:

- 1 wiąże się z poważnymi działaniami niepożądanymi i potencjalna korzyść musi je przeważać (np. długotrwałe stosowanie podwójnej terapii przeciwpłytkowej po zawale serca w celu uniknięcia ponownego zawału serca, ale kosztem zwiększonego ryzyka krwawienia)
- 2 jest niedogodna dla pacjenta
- 3 wiąże się z dużymi kosztami.²³

W takich sytuacjach autorzy wytycznych są bardziej skłonni do oceny jakości danych jako niższej przy węższym przedziale ufności niż przy oszacowaniu efektu interwencji, która jest w pełni bezpieczna (np. stosowanie ekstraktu z żurawin w celu zmniejszenia ryzyka nawrotu zakażeń układu moczowego).

Dodatkowo powinno się rozważyć obniżenie jakości danych, jeśli oszacowany efekt jest duży,

a przedział ufności wąski, ale całkowita wielkość próby i liczba zaobserwowanych punktów końcowych jest mała.²³ Powinno się w takiej sytuacji policzyć liczbę pacjentów, która byłaby niezbędna, aby pojedyncze badanie miało odpowiednią moc (jest to tzw. optymalna wielkość próby – *optimal information size*).²⁴

Błąd systematyczny związany z wybiórczym publikowaniem badań Badania, w których uzyskano istotne statystycznie wyniki mają większą szansę na publikację niż badania o podobnej wiarygodności, w których nie wykazano znamiennej zmiany. Skutkiem może być błąd systematyczny związany z wybiórczym publikowaniem badań (*publication bias*) (patrz także str. 22).²⁵ W wyniku tego błędu może dochodzić do znacznego zawyżenia efektu interwencji. Istnieją metody (graficzne [np. wykres lejkowy *funnel plot*] i statystyczne), które pozwalają podejrzewać istnienie tego błędu, ale za pomocą żadnej z nich nie można go potwierdzić. Dlatego ich wyniki należy interpretować ostrożnie.

Ten błąd systematyczny należy szczególnie podejrzewać, jeśli przegląd systematyczny obejmuje:

- 1 wczesne badania (czyli testujące nową interwencję), zwłaszcza obejmujące niewielką liczbę uczestników i w których zaobserwowano mało punktów końcowych
- 2 głównie małe badania sponsorowane przez producentów leków.²⁶

W obu tych sytuacjach autorzy wytycznych powinni rozważyć ocenę jakości danych jako niższą niż oszacowana wyjściowo.

Jednakże w związku z tym, że nie da się potwierdzić obecności tego błędu, grupa GRADE sugeruje, aby przy podejrzeniu tego błędu obniżyć jakość danych tylko o 1 poziom.²⁶

Czynniki mogące podnieść jakość danych Rzadko podnosi się wyjściową jakość danych. Najczęstszą przyczyną oceny jakości danych jako wyższej niż podstawowa oceniona na podstawie metodologii badań jest zaobserwowanie dużego efektu, czyli silnego związku między interwencją a punktem końcowym.

Grupa GRADE sugeruje, by podnieść jakość danych o 1 poziom lub 2 poziomy, jeśli w metodologicznie poprawnych badaniach obserwacyjnych wykazano zwiększenie lub zmniejszenie ryzyka wystąpienia punktu końcowego, odpowiednio co najmniej 2-krotne ($RR \geq 2$ lub $0,2-0,5$) lub co najmniej 5-krotne ($RR \geq 5$ lub $<0,2$).²⁷ W przypadku ilorazu szans przyjęcie wyższych progów wydaje się właściwsze. Jako duży efekt należy również traktować szybką odpowiedź na leczenie.

Inne sytuacje, w których możliwa jest ocena jakości danych z badań obserwacyjnych jako wyższej to:

- 1 wykazanie zależności efektu od dawki
- 2 sytuacja, gdy wszystkie możliwe czynniki zakłócające zmniejszyłyby wielkość zaobserwowanego korzystnego efektu lub wskazywałyby

na niekorzystne następstwa przy niezaobserwowaniu żadnego istotnego efektu.²⁷

Warunkiem podniesienia jakości danych z badań obserwacyjnych jest ich poprawność metodologiczna obejmująca:

- 1 dokładny pomiar czynników rokowniczych związanych z punktem końcowym, będącym przedmiotem zainteresowania autorów wytycznych
- 2 małą liczbę uczestników, których utracono z obserwacji (*lost to follow-up*)
- 3 właściwy pomiar punktów końcowych
- 4 właściwą analizę statystyczną uwzględniającą różnice w rozkładzie czynników rokowniczych w grupie z interwencją badaną i w grupie kontrolnej.²⁷

Nie należy oceniać jakości danych jako wyższej niż podstawowa, jeśli podejrzewa się występowanie błędów systematycznych, w tym związanego z selektywnością publikowania lub nieprecyzyjność wyniku.

Przygotowanie profilu danych Następnym etapem tworzenia wytycznych jest zebranie danych (informacji na temat jakości danych odnośnie do wszystkich ważnych punktów końcowych i wielkości oszacowanych efektów) w tabelarycznej formie profilu danych (*evidence profile* – EP).¹¹ W EP (RYCINA 4) znajduje się dokładna ocena wszystkich czynników wpływających na jakość danych (wymienionych na RYCINIE 3 i opisanych szczegółowo wyżej) w przypadku każdego ważnego punktu końcowego oraz zestawienia danych liczbowych świadczących o wielkości efektu danej interwencji (*summary of findings table* – tabela SoF). Niekiedy jedyną formą zebrania danych jest tabela SoF, w której podana jest tylko ogólna jakość danych, bez szczegółowego uzasadnienia (RYCINA 5). EP i tabelę SoF przygotowuje się w innym celu i dla innych odbiorców. EP jest dokładniejszym dokumentem, gdyż na jego podstawie można dokładnie prześledzić, skąd wynika ostateczna ocena jakości danych. Dzięki temu ta ocena jest przejrzysta. EP są używane przez autorów wytycznych i autorów przeglądów systematycznych. Tabele SoF są przeznaczone dla szerszej grupy odbiorców, w tym osób, które wykorzystują wyniki przeglądów systematycznych w codziennej praktyce. Tabela SoF zawiera podsumowanie informacji najważniejszych dla podjęcia decyzji i pokazuje, co jest podstawą konkretnego zalecenia wytycznych.¹¹ Tabela SoF jest obecnie stałym elementem przeglądu Cochrane. Zaleca się, by uwzględniono w tabeli SoF nie więcej niż 7 najważniejszych punktów końcowych.²⁸ W 2 małych badaniach z randomizacją wykazano, że umieszczenie tabeli SoF ułatwia korzystanie z przeglądów Cochrane, szukanie przydatnych informacji i zrozumienie wyników.²⁹

W wydaniu 1 z 2009 roku w Bibliotece Cochrane znajdowały się 3 przeglądy Cochrane zawierające tabelę SoF, w marcu 2012 roku – 502 takie przeglądy, a we wrześniu 2014 roku blisko 1000.²⁸

Specjalne oprogramowanie GRADEpro³⁰ pozwala przygotować zarówno EP, jak i tabelę SoF.

Author(s):
Date:
Question: Should megestrol be used for cachexia-anorexia syndrome?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Megestrol	Control	Relative (95% CI)	Absolute		
Survival (follow-up median 1 years)												
2	randomised trials	no serious risk of bias	no serious inconsistency ¹	no serious indirectness	no serious imprecision	publication bias	55/250 (22%)	53/248 (21.4%)	RR 1.02 (0.73 to 1.42)	4 more per 1000 (from 58 fewer to 90 more)	⊕⊕⊕○ MODERATE	CRITICAL
Appetite improvement (follow-up 8-12 weeks)												
5	randomised trials	no serious risk of bias	serious ¹	no serious indirectness	no serious imprecision	dose response gradient	170/301 (56.5%)	47/262 (17.9%)	RR 3.0 (1.86 to 4.84)	359 more per 1000 (from 154 more to 689 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Physical status worsening												
3	randomised trials	no serious risk of bias	serious ¹	no serious indirectness	no serious imprecision	publication bias	103/225 (45.8%)	107/175 (61.1%)	RR 0.65 (0.39 to 1.08)	190 fewer per 1000 (from 0 more to 380 more) ²	⊕⊕○○ LOW	CRITICAL
								0%		-		

¹ Heterogeneity of the results

RYCINA 4 Tabela profilu danych (*evidence profile* – EP) – widok z GRADEpro (na podstawie Leśniak i wsp.⁴⁴)

Megestrol for cachexia-anorexia syndrome						
Patient or population: patients with cachexia-anorexia syndrome						
Settings:						
Intervention: megestrol						
Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk Control	Corresponding risk Megestrol				
Survival Follow-up: median 1 years	214 per 1000	218 per 1000 (156 to 303)	RR 1.02 (0.73 to 1.42)	498 (2 studies)	⊕⊕⊕⊕ moderate ^{1,2}	
Appetite improvement Follow-up: 8-12 weeks	179 per 1000	538 per 1000 (334 to 808)	RR 3.0 (1.86 to 4.84)	563 (5 studies)	⊕⊕⊕⊕ high ^{1,3}	
Physical status worsening	611 per 1000	397 per 1000 (238 to 660)	RR 0.65 (0.39 to 1.08)	400 (3 studies)	⊕⊕⊕⊕ low ^{1,2}	
		2				

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio;

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

¹ Heterogeneity of the results
² High probability of publication bias
³ Strong dose response gradient

RYCINA 5 Tabela zestawienia danych liczbowych świadczących o wielkości efektu danej interwencji (*summary of findings table* – tabela SoF) – widok z GRADEpro (na podstawie Leśniak i wsp.⁴⁴)

W tabeli SoF (**RYCINA 5**) znajdują się następujące informacje:³¹

- 1 lista wszystkich ważnych punktów końcowych (korzystnych i niekorzystnych)
- 2 typowa częstość występowania danego punktu końcowego (ryzyko jego wystąpienia w grupie kontrolnej – *assumed risk*). Uważa się, że jego wyliczanie z badań z randomizacją, w których uczestnicy są zwykle ściśle wyselekcjonowani jest mylące. Dlatego uważa się, że to ryzyko powinno być oszacowane na podstawie poprawnych metodologicznie badań obserwacyjnych. Jeśli takich badań nie ma, to powinno się je uzyskać jako medianę (nie średnią ważoną)

- wartości tego ryzyka ze wszystkich badań włączonych do przeglądu lub wybrać ryzyko w grupie kontrolnej w jednym badaniu, jeśli jest ono zdecydowanie większe niż pozostałe. Dla tego parametru nie podaje się przedziału ufności.
- 3 ryzyko w grupie z interwencją (*corresponding risk*) albo różnica ryzyka w obu grupach
- 4 względny efekt interwencji (wyrażony przez RR, HR lub OR, przy czym najczęściej stosuje się RR jako najprostszy i najbardziej intuicyjny)
- 5 liczba uczestników i badań, w których oceniano dany punkt końcowy

6 ocena jakości danych dla każdego punktu końcowego

7 komentarze.

Przy tworzeniu wytycznych przeprowadza się następnie ocenę jakości danych dotyczącą wszystkich krytycznych punktów końcowych łącznie. Determinowana jest ona przez najniższą jakość danych dotyczących któregośkolwiek z krytycznych punktów końcowych.³² Ta łączna jakość danych jest potrzebna, aby określić na jakich danych opiera się konkretne zalecenie.

Określenie siły zalecenia Następnym etapem tworzenia wytycznych jest formułowanie zaleceń wraz z podaniem ich siły.

Siła zalecenia jest wynikiem oceny bilansu korzystnych i niekorzystnych skutków porównywanych opcji postępowania, z uwzględnieniem skutków zdrowotnych, uciążliwości, możliwości implementacji w danym systemie opieki zdrowotnej, akceptowalności wśród pacjentów, lekarzy i innych pracowników opieki zdrowotnej, sprawiedliwości społecznej w dostępie do świadczeń zdrowotnych, a także środków (kosztu) związanych z wprowadzeniem i przestrzeganiem zalecenia. Te korzystne i niekorzystne skutki to punkty końcowe uznane za krytyczne lub ważne (wybrane na początku procesu tworzenia wytycznych, potwierdzone po uzyskaniu danych i uwzględnione w zestawieniu danych).

Ujmując rzecz bardziej szczegółowo, na kierunek i siłę zalecenia wpływają:

- 1** wielkość oszacowanego wpływu danej interwencji na ważne punkty końcowe i pewność co do trafności tego oszacowania (im mniejsza ta pewność, tym mniejsze prawdopodobieństwo silnego zalecenia)
- 2** oszacowanie typowych wartości i preferencji i pewność co do trafności tego oszacowania (im mniejsza ta pewność, tym mniejsze prawdopodobieństwo silnego zalecenia)
- 3** różnorodność wartości i preferencji (im większa, tym mniejsze prawdopodobieństwo silnego zalecenia)
- 4** konieczne zasoby i koszty (im większe, tym mniejsze prawdopodobieństwo silnego zalecenia).³³

O ile ocenę jakości danych na podstawie ściśle określonych kryteriów można uznać za proces stosunkowo obiektywny i powtarzalny, o tyle ocena bilansu korzystnych i niekorzystnych skutków rozpatrywanego postępowania w dużej mierze zależy od systemu wartości i preferencji osób decydujących o tym bilansie. Subiektywnych sądów nie można wyeliminować, ale powinno się je przejrzysto udokumentować.

Siłę zalecenia można wyobrazić sobie jako zmienną ciągłą, ale dla ułatwienia interpretacji i korzystania z wytycznych w metodologii GRADE wyróżnia się 4 kategorie zaleceń:

- 1** silne za stosowaniem danej interwencji
- 2** silne przeciwko jej stosowaniu
- 3** słabe (zwane inaczej warunkowymi) za
- 4** słabe przeciwko.

Wydaje się, że korzyści z takiego uproszczenia przeważają nad potencjalnymi ograniczeniami.

Silne zalecenie należy interpretować w następujący sposób:

- 1** w odniesieniu do lekarzy – u zdecydowanej większości pacjentów, do których się ono odnosi, powinno się określić interwencję zastosować (lub przynajmniej taką interwencję im zaproponować)
- 2** w odniesieniu do pacjentów – większość z nich, po uzyskaniu wyczerpujących informacji na temat danej interwencji i alternatywnych sposobów postępowania, wybrałaby tę interwencję
- 3** w odniesieniu do decydentów – stosowanie (lub proponowanie pacjentom) danej interwencji może być wskaźnikiem jakości opieki zdrowotnej.

Niepewność co do bilansu korzystnych i niekorzystnych skutków (np. z powodu niskiej jakości danych, małych efektów interwencji, znacznych kosztów, wątpliwości co do możliwości implementacji itd.) mogą zmniejszyć siłę zalecenia. W sytuacjach, kiedy przekonanie autorów o słuszności zalecenia jest mniejsze (choć uważają, że prawdopodobnie korzyści z postępowania zgodnie z tym zaleceniem przeważają nad skutkami niekorzystnymi) formułuje się zalecenie słabe (warunkowe), które oznacza, że znaczna część pacjentów może wybrać odmienny sposób postępowania, a podjęcie decyzji wymaga od lekarza większej uwagi i dłuższego czasu, aby bardziej szczegółowo przedstawić pacjentowi możliwe opcje postępowania i ich efekty. Wspólna decyzja odzwierciedla wartości i preferencje pacjenta.

W wytycznych formułuje się często przypuszczenia co do wartości i preferencji populacji docelowej dla zaleceń wytycznych. W przypadku zaleceń, których siła w istotny sposób zależy od wartości i preferencji, zaleca się, by w treści zalecenia wprost napisać, którym wartościom przy formułowaniu zalecenia przypisano dużą wagę, a którym mniejszą.³⁴

W podejściu GRADE najczęściej przypisuje się zaleceniom silnym „1” oraz określenia „zaleca się”, „należy”, „powinno się”, a zaleceniom słabym – „2” oraz określenia „sugeruje się”, „można rozważyć”, „warunkowo zaleca się”. Niektóre organizacje oznaczają siłę zaleceń cyframi (np. rzymskimi – **TABELA 1**), literami albo symbolami (gwiazdki, strzałki itd.).

Sporadycznie autorzy wytycznych mogą się zdecydować na sformułowanie zaleceń odnoszących się tylko do badań, a nie codziennej praktyki („*only-in-research*” recommendation) lub na nieformułowanie zaleceń.³⁴ Zalecenia odnośnie do stosowania danej interwencji tylko w badaniach klinicznych dotyczą sytuacji, gdy ta interwencja ma potencjalny „obietujący” korzystny efekt, na który nie ma na razie wystarczających danych, ale przyszłe badania mogą zmniejszyć niepewność co do tej skuteczności. Natomiast niezalecenie jej grozi wstrzymaniem dalszych badań.

Autorzy wytycznych mogą się zdecydować na nieformułowanie zaleceń, jeśli:

- 1** nie mają pewności co do oszacowania efektu

TABELA 1 Klasyfikacja zaleceń i danych naukowych wg European Society of Cardiology (ESC) oraz American College of Cardiology i American Heart Association (ACC/AHA); przedruk z Jaeschke i wsp.¹³

	ESC	ACC/AHA
klasa zalecenia		
I (oznacza: zdecydowanie stosować)	dane naukowe lub powszechnie akceptowana opinia wskazują, że określone postępowanie diagnostyczne lub lecznicze jest korzystne, użyteczne i skuteczne	stany, w których określone postępowanie (diagnostyczne lub lecznicze) jest korzystne, przydatne i skuteczne, na co wskazują wyniki badań lub powszechnie akceptowana opinia
II	dane naukowe lub opinie dotyczące przydatności lub skuteczności określonego postępowania diagnostycznego lub leczniczego nie są zgodne	stany, w których wyniki badań lub opinie co do przydatności lub skuteczności określonego postępowania nie są zgodne
Ila (oznacza: raczej stosować)	dane naukowe lub opinie przemawiają za użytecznością lub skutecznością	wyniki badań lub panujące opinie przemawiają za przydatnością lub skutecznością określonego postępowania
IIb (oznacza: można rozważyć stosowanie)	użyteczność lub skuteczność jest słabiej potwierdzona przez dane naukowe lub opinie	przydatność lub skuteczność określonego postępowania jest słabiej potwierdzona przez wyniki badań lub panujące opinie
III (oznacza: zdecydowanie nie stosować)	dane naukowe lub powszechnie akceptowana opinia wskazują, że określone postępowanie diagnostyczne lub lecznicze nie jest użyteczne ani skuteczne, a w niektórych przypadkach może być szkodliwe	stany, w których określone postępowanie jest nieprzydatne lub nieskuteczne, a w niektórych przypadkach może być szkodliwe – na co wskazują wyniki badań lub powszechnie akceptowana opinia
stopień wiarygodności danych		
A	dane pochodzące z licznych badań z randomizacją lub z metaanaliz	dane pochodzące z licznych badań klinicznych z randomizacją lub z metaanaliz
B	dane pochodzące z jednego badania z randomizacją lub z dużych badań bez randomizacji	dane pochodzące z jednego badania z randomizacją lub z badań bez randomizacji
C	uzgodniona opinia ekspertów lub dane pochodzące z małych badań, lub z badań retrospektywnych bądź rejestrów	uzgodniona opinia ekspertów, opisy przypadków, standardowe postępowanie

2 bilans korzystnych i niekorzystnych następstw jest wyrównany

3 są przekonani, że 2 opcje postępowania wiążą się z różnymi niekorzystnymi następstwami i w związku z tym decyzje poszczególnych pacjentów mogą znacznie się różnić.³⁴

Wydaje się jednak, że mimo wszystko autorzy ci powinni sformułować zalecenia (a na pewno przynajmniej przedstawić przyczyny ich niepodania, gdyż i tak dysponują większą wiedzą niż pojedynczy lekarz, który musi podejmować decyzje w takich problematycznych sytuacjach).

Sytuacje szczególne – dane na temat rokowania i testów diagnostycznych Opracowano już system oceny jakości danych (EP i tabela SoF) dla pytań klinicznych dotyczących testów diagnostycznych.³⁵⁻³⁷

Przygotowywane są systemy zarówno dla pytań dotyczących rokowania,³⁸ jak i dla danych jakościowych.

Słabe strony i wypaczenia systemu GRADE Wiele organizacji tworzących wytyczne przyjęło system GRADE. Nie wszystkie jednak stosują go konsekwentnie i we właściwy sposób. W 2007 roku

WHO zaadaptowała system GRADE do tworzenia wytycznych z zakresu zdrowia publicznego. Okazało się, że wiele silnych zaleceń (>50%) opiera się na danych niskiej lub bardzo niskiej jakości. GRADE dopuszcza taką możliwość w 5 sytuacjach:

- 1 mały koszt w stosunku do prawdopodobnej korzyści (silne zalecenie, by stosować)
- 2 duży koszt może uzasadniać zalecenie niestosowania interwencji (silne zalecenie, by nie stosować)
- 3 duży koszt w stosunku do korzyści może wspierać zalecenie mniej szkodliwej alternatywnej interwencji (silne zalecenie, by stosować mniej szkodliwą lub tańszą interwencję)
- 4 duży koszt jednej z alternatywnych interwencji (silne zalecenie, by nie stosować interwencji bardziej szkodliwej lub droższej)
- 5 duży koszt potencjalnie szkodliwej interwencji może uzasadnić zalecenie stosowania mniej szkodliwej (silne zalecenie, by nie stosować danej interwencji albo by stosować interwencję mniej szkodliwą lub tańszą, jeśli porównuje się 2 interwencje).³⁹

W przypadku WHO tylko w około 15% silne zalecenia oparte na danych niskiej jakości można było uzasadnić tymi 5 sytuacjami.³⁹ Wynikało

TABELA 2 Instrument AGREE II – kryteria oceny jakości wytycznych praktyki klinicznej (na podstawie: *The AGREE II Instrument. The AGREE Research Trust, 2009* (<http://www.agreertrust.org/>); przedruk z Jaeschke i wsp.¹³)

zakres i cel	1. Ogólny cel (lub cele) wytycznych jest (są) precyzyjnie opisany (opisane). 2. Pytanie (pytania) kliniczne objęte wytycznymi jest (są) precyzyjnie opisane. 3. Populacja (pacjenci, społeczeństwo itd.), do której wytyczne się odnoszą, jest precyzyjnie opisana.
udział użytkowników	4. W skład zespołu opracowującego wytyczne wchodzi przedstawiciele wszystkich grup zawodowych, których one dotyczą. 5. Uwzględniono punkt widzenia i preferencje populacji (pacjentów, społeczeństwa itd.). 6. Użytkownicy wytycznych są jasno określani.
poprawność metodyki	7. Wyszukiwanie danych przeprowadzono w sposób systematyczny. 8. Kryteria doboru danych są jasno opisane. 9. Zalety i ograniczenia wszystkich danych zostały przejrzysto opisane. 10. Metody formułowania zaleceń są jasno opisane. 11. Przy formułowaniu zaleceń uwzględniono korzyści zdrowotne, skutki uboczne i zagrożenia. 12. Zalecenia zostały wyraźnie powiązane z danymi, na podstawie których je sformulowano. 13. Przed publikacją wytyczne zostały zrecenzowane przez zewnętrznych ekspertów. 14. Podano procedurę aktualizacji wytycznych.
przejrzystość i format	15. Zalecenia są konkretne i jednoznaczne. 16. Różne możliwości postępowania w danej sytuacji są jasno przedstawione. 17. Główne zalecenia są łatwe do zidentyfikowania.
użyteczność	18. Omówiono potencjalne bariery i ułatwienia w stosowaniu zaleceń. 19. Wytyczne są uzupełnione poradami i/lub narzędziami ułatwiającymi ich stosowanie. 20. Uwzględniono potencjalne skutki ekonomiczne postępowania zgodnego z zaleceniami. 21. Wytyczne zawierają kryteria służące do monitorowania lub kontroli jakości.
niezależność redakcyjna	22. Poglądy instytucji finansujących proces powstawania wytycznych nie miały wpływu na ich treść. 23. Potencjalne konflikty interesów autorów wytycznych zostały ujawnione i podjęto w związku z nimi odpowiednie działania.

to z nieznanymi zasadami GRADE przez autorów tych wytycznych.⁴⁰

Obok czynników determinujących siłę zalecenia wymienionych wyżej, w przypadku decyzji z zakresu zdrowia publicznego należy uwzględnić częstość występowania choroby, dostępność danej interwencji, wykonalność i akceptowalność zaleceń, czynniki utrudniające i czynniki ułatwiających implementację, zakres obowiązujących suboptimalnych praktyk i wpływ na nierówności w dostępie do opieki zdrowotnej.

Członkowie grupy GRADE uważają, że istnieje pilna potrzeba opracowania zasad korzystania z systemu GRADE i kształcenia osób tworzących wytyczne, aby uniknąć niewłaściwego używania tego systemu i tym samym dezorientacji użytkowników wytycznych.

Co nowego? W ramach 7 Ramowego Programu Unii Europejskiej finansowany jest projekt DECIDE (2011–2015).⁴¹

Projekt ten ma dostarczyć nowe narzędzia przydatne w procesie tworzenia i upowszechniania wytycznych (wśród lekarzy, decydentów i pacjentów). Wśród tych nowych narzędzi znajdują się między innymi:

1 iSoF – interaktywna wersja tabeli SoF (*interactive summary of findings table*) – kompatybilna z programem GradePro

2 GRADEproGDT (GRADEpro Guideline Development Tool) – narzędzie do tworzenia zaleceń, zawierające tabelę SoF dla testów diagnostycznych

3 GET IT – słownik terminów medycznych związanych z metodologią badań naukowych, przeznaczony dla osób niezwiązanych profesjonalnie z opieką zdrowotną.

Jak rozpoznać dobre wytyczne – instrument AGREE II

Niestety nie wszystkie publikowane wytyczne praktyki klinicznej powstają zgodnie z zasadami opisanymi powyżej. Aby ułatwić ocenę jakości wytycznych (w szczególności wykrycie ich ograniczeń), międzynarodowy zespół badaczy i autorów wytycznych (AGREE Collaboration) opracował zestaw kryteriów jakości wytycznych – tzw. instrument AGREE, który został następnie zaktualizowany i opublikowany jako instrument AGREE II.⁴²

Kryteria instrumentu AGREE II (opisujące czynniki, które warunkują wiarygodność wytycznych) są zgrupowane w 6 domenach (TABELA 2).^{9,13}

1 Pierwsza domena dotyczy określenia celu wytycznych, precyzji pytań klinicznych, na które odpowiadają zalecenia zawarte w wytycznych oraz sprecyzowania, do jakiej populacji pacjentów wytyczne się odnoszą i dla jakich odbiorców są przeznaczone.

2 Druga domena dotyczy tego, w jakim stopniu wytyczne odzwierciedlają poglądy osób,

których dotyczą (zarówno ich użytkowników, jak i pacjentów). To w jaki sposób interpretuje się obiektywną wiedzę i jakie metody postępowania się wybiera zależy od uwarunkowań kulturowych, religijnych, prawnych, społecznych itd. Dlatego w zespołach opracowujących wytyczne powinni się znaleźć przedstawiciele wszystkich zainteresowanych grup, także pacjentów.

- 3 Trzecia domena dotyczy wyszukiwania i syntezy danych (etapu stosunkowo obiektywnego, sprawdzalnego i w dużym stopniu powtarzalnego) oraz procesu formułowania i aktualizowania zaleceń (etapu ze swej natury subiektywnego). Jest to krytyczny element oceny wiarygodności wytycznych. Systemem zapewniającym wysoką jakość zaleceń jest opisany system GRADE. Ważna jest również aktualność wytycznych, ponieważ postęp wiedzy w medycynie nie jest szybki. W praktyce większość wytycznych wymaga aktualizacji średnio po upływie 3 lat, ale ten czas zależy od dynamiki postępu w danej dziedzinie medycyny.
- 4 Czwarta domena dotyczy formy przedstawienia wytycznych – ich przejrzystości, jednoznaczności i przystępności wytycznych dla użytkowników.
- 5 Piąta domena zawiera kryteria dotyczące aspektów związanych z praktycznym wykorzystaniem wytycznych (załączenie poradników i narzędzi ułatwiających ich stosowanie, rozważenie skutków ekonomicznych i organizacyjnych ich zastosowania, określenie wskaźników jakości do monitorowania stosowania zaleceń i oceny efektów określonego postępowania).
- 6 Szósta domena zawiera kryteria pozwalające na ocenę prawdopodobieństwa, że zalecenia mogły zostać wypaczone przez ingerencję organizacji lub firm finansujących tworzenie wytycznych bądź wskutek finansowych powiązań autorów wytycznych z przemysłem medycznym. Oczekuje się, że autorzy wytycznych złożą deklaracje co do potencjalnych sprzeczności interesów.

Ograniczeniami instrumentu AGREE są:

- 1 ocena tylko formy przedstawienia (opisania) wytycznych oraz metodyki ich tworzenia, a nie zawartości merytorycznej (np. jakości przeglądu systematycznego)
- 2 brak formalnych zasad, pozwalających na interpretację sumarycznego wyniku oceny jakości wytycznych.

Jak korzystać z wytycznych Należy pamiętać, że wytyczne nie zastępują myślenia klinicznego w każdym indywidualnym przypadku. Stanowią one tylko jeden z czynników (choć ważny i przydatny) uwzględnianych w procesie podejmowania decyzji klinicznych. Dlatego nie należy ich traktować jak sztywnych standardów.

Lekarz, który chce się posilkować wytycznymi w codziennej praktyce, musi umieć nie tylko ocenić ich wiarygodność, ale też właściwie interpretować i stosować zalecenia w nich zawarte.

Musi ustalić, czy wytyczne są aktualne i czy odnoszą się do jego pacjenta (lub grupy pacjentów), gdyż w niektórych sytuacjach zaleceń nie można ekstrapolować.¹³

Istnieje też ryzyko bezkrytycznego stosowania wszystkich zalecanych interwencji, co może stwarzać poważne problemy w przypadku współistnienia chorób, zwłaszcza u osób w starszym wieku.

Wytyczne praktyki klinicznej bywają także wykorzystywane, nieświadomie lub świadomie, w celach, dla których nie zostały stworzone. Wśród przedstawionych wyżej kryteriów wiarygodności wytycznych znalazły się wskaźniki służące do monitorowania stosowania zaleceń i oceny efektów określonego postępowania. Głównie dotyczą one stosowania się lekarzy (lub innych pracowników ochrony zdrowia) do zaleceń, gdyż ocena wpływu wprowadzenia wytycznych na losy chorych jest znacznie trudniejsza. Można więc śledzić efektywność procesu wdrażania wytycznych i zmiany postępowania. Taką kontrolę jakości postępowania próbuje się jednak wykorzystywać do oceny pracy indywidualnych lekarzy i uzależniać od jej wyniku ich wynagrodzenie, a nawet zatrudnienie. Skutkiem tego może być ignorowanie preferencji pacjenta przy podejmowaniu decyzji o zastosowaniu określonej, zalecanej w wytycznych, metody diagnostycznej lub leczniczej.¹³

Wytyczne można też niekiedy wykorzystywać w celach marketingowych. Trzeba podkreślić, że ważne jest nie tylko to, czy opracowujący wytyczne rzeczywiście kierowali się przesłankami pozamerytorycznymi, ale również to, czy rozsądnie myśląc osoba może odnieść wrażenie, iż taka sytuacja zaistniała. Niezależnie od rzeczywistej rzetelności opracowania wytycznych, już samo wrażenie sprzeczności interesów autorów zaleceń może wpłynąć na losy całych wytycznych.¹³

Co robić w przypadku sprzecznych zaleceń? W swojej praktyce klinicznej lekarze spotykają się z sytuacją, gdy różne wytyczne podają sprzeczne zalecenia. Takie różnice mogą wynikać z wielu przyczyn, takich jak:⁴³

- 1 różna ocena przez autorów wytycznych dostępnych danych naukowych pod kątem ich przydatności
- 2 błąd systematyczny w badaniach, na podstawie których formułowano zalecenie
- 3 uwzględnienie w jednych z wytycznych wyników badań nieodnoszących się do konkretnej sytuacji klinicznej
- 4 różna ocena korzyści, działań niepożądanych i kosztów
- 5 konflikt interesów autorów jednych z wytycznych
- 6 nieznajomość niektórych danych, nieumiejętność krytycznej oceny wiarygodności danych i nieuwzględnienie ważnych punktów końcowych przez autorów jednych z wytycznych.

Lekarz może w takiej sytuacji:

- 1 zignorować wszystkie wytyczne, co nie jest jednak wskazane
- 2 potrafić rozróżnić te wytyczne, które powstały w systematyczny sposób, na podstawie przeglądu systematycznego (i w których opisano metody stosowane przy ich tworzeniu) i stosować się do nich, a odrzucać wytyczne, w których wystąpiła sprzeczność interesów autorów lub tworzyli je w przypadkowy sposób.⁴³

Dlatego lekarze nie powinni bezkrytycznie akceptować zaleceń.

Jak wdrażać wytyczne do praktyki Nawet najlepsze wytyczne nie przyniosą poprawy opieki nad pacjentami, jeśli nie zostaną skutecznie wdrożone. Strategia wdrażania wytycznych jest równie ważna jak metodologia ich tworzenia i obejmuje:

- 1 szerokie rozpowszechnienie wytycznych
- 2 aktywne wdrażanie – różne metody zachęcania do poznania wytycznych; egzekwowanie ich znajomości oraz postępowania zgodnie z zaleceniami
- 3 monitorowanie
 - a) postępowania zgodnie z wytycznymi
 - b) zmian efektów opieki zdrowotnej, czyli spodziewanych korzyści z wprowadzenia wytycznych
 - c) barier w realizacji zaleceń.¹³

Za liczne bariery w procesie od wydania wytycznych (zaleceń) do uzyskania korzyści zdrowotnych u pacjenta odpowiadają wszyscy uczestnicy procesu leczniczego (lekarze, system opieki zdrowotnej, pacjenci).

Problemy z wdrażaniem wytycznych na poziomie lekarza mogą być związane z:

- 1 utrudnionym dostępem do tekstu wytycznych, ich mało „przyjazną dla użytkownika” formą, wcześniej utrwalonymi przyzwyczajeniami
- 2 ograniczoną dostępnością leku (np. brak refundacji), konsultacji specjalistycznych, badań diagnostycznych, z organizacją systemu opieki zdrowotnej itp.
- 3 zapominaniem lub ignorowaniem zaleceń.¹³

Bariery na poziomie pacjenta mogą wynikać z niewyrażenia przez niego zgody oraz z nieprzestrzegania przez niego zaleceń lekarskich, zwłaszcza w przypadku interwencji długoterminowej.

Warunkiem przełamywania tych barier jest aktywne wdrażanie wytycznych.

Należy sobie zdawać sprawę z tego, że siła zalecenia nie jest bezpośrednio związana z tym, czy dane zalecenie należy traktować jako priorytetowe do wdrożenia.³⁴ Decydenci w opiece zdrowotnej, ustalając priorytety, uwzględniają:

- 1 częstość występowania danej choroby (zalecenia odnoszące się do częściej występującej choroby są priorytetowe),
- 2 łatwość wdrożenia zaleceń (priorytetowe te, które łatwiej w danej chwili można wdrożyć)
- 3 kwestię sprawiedliwości społecznej (priorytetowe te, które zmniejszają nierówności w dostępie do świadczeń zdrowotnych)

- 4 koszt dla społeczeństwa (w dalszej kolejności są wdrażane te, które wiążą się z dużymi kosztami)
- 5 potencjalną poprawę jakości opieki zdrowotnej (im większa, tym ważniejsze są zalecenia).³⁴

PIŚMIENNICTWO

- 1 Field M, Lohr K. Clinical practice guidelines: directions for a new program. Washington, DC: National Academic Press; 1990.
- 2 Guyatt G. Historia EBM. W: Podstawy EBM. Gajewski P, Jaeschke R, Brożek J, red. Kraków, Poland. Med Prakt. 2008; 17-23.
- 3 Guyatt GH, Meade MO, Jaeschke R i wsp. Practitioners of evidence based care. Not all clinicians need to appraise evidence from scratch but all need some skills. BMJ. 2000; 320: 954-955.
- 4 Feder G. Guidelines for clinical guidelines. BMJ. 1998; 317: 427-428.
- 5 Nathwani D. From evidence-based guideline methodology to quality of care standards. J Antimicrob Chemother. 2003; 5: 1103-1107.
- 6 Grilli R, Magrini N, Penna A i wsp. Practice guidelines developed by specialty societies: the need for critical appraisal. Lancet. 2000; 355: 103-106.
- 7 Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. JAMA. 1999; 281: 1900-1905.
- 8 Appraisal of Guidelines, Research, and Evaluation in Europe (AGREE) Collaborative Group. For the AGREE Collaborative Group. Guideline development in Europe: an international comparison. Int J Technol Assess Health Care. 2000; 16: 1036-1046.
- 9 Cluzeau F. Jakość wytycznych praktyki klinicznej – Projekt AGREE. Med Prakt. 2003; 3: 46-52.
- 10 The GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004; 328: 1490.
- 11 Guyatt G, Oxman AD, Akl EA i wsp. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011; 64: 383-394. doi:10.1016/j.jclinepi.2010.04.026.
- 12 Jaeschke R, Guyatt GH, Dellinger P i wsp. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ. 2008; 31; 337: a744. doi:10.1136/bmj.a744.
- 13 Jaeschke R, Gajewski P, Brożek J i wsp. Wytyczne praktyki klinicznej. W: Podstawy EBM. Gajewski P, Jaeschke R, Brożek J, red. Kraków, Poland: Medycyna Praktyczna; 2008: 143-158.
- 14 Guyatt GH, Oxman AD, Kunz R i wsp. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011; 64: 395-400.
- 15 Balshem H, Helfand M, Schünemann HJ i wsp. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011; 64: 401-406.
- 16 Guyatt GH, Oxman AD, Vist G i wsp. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol. 2011; 64: 407-415.
- 17 Bassler D, Briel M, Montori VM i wsp. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA. 2010; 303: 1180-1187.
- 18 Guyatt GH, Oxman AD, Kunz R i wsp. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. GRADE Working Group. J Clin Epidemiol. 2011; 64: 1294-1302. doi:10.1016/j.jclinepi.2011.03.017.
- 19 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003; 327: 557-560.
- 20 Guyatt GH, Oxman AD, Kunz R i wsp. GRADE guidelines: 8. Rating the quality of evidence—indirectness. GRADE Working Group. J Clin Epidemiol. 2011; 64: 1303-1310. doi:10.1016/j.jclinepi.2011.04.014.
- 21 Brignardello-Petersen R, Rochwerf B, Guyatt GH. Co to jest metaanaliza sieciowa i jak ją wykorzystać w praktyce klinicznej? Med Prakt. 2015; 3: 120-122.
- 22 Puhan MA, Schünemann HJ, Murad MH i wsp. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014; 349: g5630. doi:10.1136/bmj.g5630.
- 23 Guyatt GH, Oxman AD, Kunz R i wsp. GRADE guidelines 6. Rating the quality of evidence—imprecision. J Clin Epidemiol. 2011; 64: 1283-1293. doi:10.1016/j.jclinepi.2011.01.012.
- 24 Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials. 1997; 18: 580-593.
- 25 Leśniak W, Bała M, Mrukowicz J i wsp. Przegląd systematyczny i metaanaliza. W: Podstawy EBM. Gajewski P, Jaeschke R, Brożek J, red. Kraków, Poland: Medycyna Praktyczna; 2008: 111-126.
- 26 Guyatt GH, Oxman AD, Montori V i wsp. GRADE guidelines: 5. Rating the quality of evidence—publication bias. J Clin Epidemiol. 2011; 64: 1277-1282. doi:10.1016/j.jclinepi.2011.01.011.
- 27 Guyatt GH, Oxman AD, Sultan S i wsp. GRADE guidelines: 9. Rating up the quality of evidence. GRADE Working Group. J Clin Epidemiol. 2011; 64: 1311-1316. doi:10.1016/j.jclinepi.2011.06.004.

- 28 Langendam MW, Akl EA, Dahm P i wsp. Assessing and presenting summaries of evidence in Cochrane Reviews. *Syst Rev.* 2013; 2: 81. doi:10.1186/2046-4053-2-81.
- 29 Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol.* 2010; 63: 620-626. doi:10.1016/j.jclinepi.2009.12.014.
- 30 Brozek J, Oxman A, Schünemann H. GRADEpro. [Computer program]. Version 3.6 for Windows. McMaster University. 2008. <http://www.guidelinedevelopment.org/>.
- 31 Guyatt GH, Oxman AD, Santesso N i wsp. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol.* 2013; 66: 158-172. doi:10.1016/j.jclinepi.2012.01.012.
- 32 Guyatt G, Oxman AD, Sultan S i wsp. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol.* 2013; 66: 151-157. doi:10.1016/j.jclinepi.2012.01.006.
- 33 Andrews JC, Schünemann HJ, Oxman AD i wsp. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol.* 2013; 66: 726-735. doi:10.1016/j.jclinepi.2013.02.003.
- 34 Andrews J, Guyatt G, Oxman AD i wsp. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol.* 2013; 66: 719-725. doi:10.1016/j.jclinepi.2012.03.013.
- 35 Gopalakrishna G, Mustafa RA, Davenport C i wsp. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. *J Clin Epidemiol.* 2014; 67: 760-768. doi:10.1016/j.jclinepi.2014.01.006.
- 36 Schunemann HJ, Oxman AD, Brozek J i wsp. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ.* 2008; 336: 1106-1110.
- 37 Hsu J, Brozek JL, Terracciano L i wsp. Application of GRADE: making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implement Sci.* 2011; 6: 62. doi:10.1186/1748-5908-6-62.
- 38 Iorio A, Spencer FA, Falavigna M i wsp. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ.* 2015; 350: h870. doi:10.1136/bmj.h870.
- 39 Alexander PE, Brito JP, Neumann I i wsp. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. *J Clin Epidemiol.* 2014. doi:10.1016/j.jclinepi.2014.10.011.
- 40 Alexander PE, Gionfriddo MR, Li SA i wsp. A number of factors explain why WHO guideline developers make strong recommendations inconsistent with GRADE guidance. *J Clin Epidemiol.* 2015. doi:10.1016/j.jclinepi.2015.09.006.
- 41 DECIDE (2011 – 2015) Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence. <http://www.decide-collaboration.eu/>.
- 42 The AGREE Next Steps Consortium. APPRAISAL OF GUIDELINES FOR RESEARCH & EVALUATION II (AGREE II) Instrument. www.agreetrust.org.
- 43 Oxman AD, Glasziou P, Williams JW Jr. What should clinicians do when faced with conflicting recommendations? *BMJ.* 2008; 337: a2530.
- 44 Leśniak W, Bala M, Jaeschke R, Krzakowski M. Effects of megestrol acetate in patients with cancer anorexia-cachexia syndrome – a systematic review and meta-analysis. *Pol Arch Med Wewn.* 2008; 118: 636-644.

From evidence to recommendations: developing health care guidelines following the GRADE approach

Wiktoria Leśniak¹, Małgorzata M. Bała^{2,3}, Roman Jaeschke⁴, Jan L. Brożek^{5,6}

1 Clinical Decision Making Unit, 2nd Department of Internal Medicine, Jagiellonian University Medical College, Kraków, Poland

2 Systematic Reviews Unit – Polish Cochrane Branch, Jagiellonian University Medical College, Kraków, Poland

3 Department of Hygiene and Dietetics, Jagiellonian University Medical College, Kraków, Poland

4 Department of Medicine and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

5 Department of Clinical Epidemiology & Biostatistics and Division of Immunology & Allergy, Department of Medicine, McMaster University, Hamilton, Ontario, Canada

6 University of Freiburg, Medical Informatics and Clinical Epidemiology, Institute for Medical Biometry and Statistics, Freiburg im Breisgau, Germany

KEY WORDS

clinical practice
guidelines, endpoints,
evidence of validity,
strength of
recommendations

ABSTRACT

Clinical practice guidelines (CPGs) are systematically developed statements (recommendations) to assist practitioners and other health care providers in making optimal decision regarding patient care in specific clinical situations. They also play an important role in health policy. Every year numerous medical societies and other professional organizations publish documents which they call “guidelines”. Not all of them are developed correctly and in a systematic way. The article describes the methodology of CPG development which increases the probability that CPGs are valid and clinically useful. This methodology was developed and is still being improved by the international working group of methodologists called GRADE. The role of GRADE is to frame clinical questions, decide on important outcomes, rate the quality of evidence, interpret the results (mainly the estimated relative and absolute effects), and provide judgments on the strength of recommendation (namely, confidence that particular recommendations should be followed).

Correspondence to:

Wiktoria Leśniak, MD, PhD, Pracownia
Podejmowania Decyzji Klinicznych,
II Katedra Chorób Wewnętrznych,
Uniwersytet Jagielloński, Collegium
Medicum, ul. Skawieńska 8,
31-066, Kraków, Poland, phone:
+48 12 430 52 66,
e-mail: lesniak@mp.pl
Received: October 30, 2015.
Accepted: October 30, 2015.
Pol Arch Med Wewn. 2015;
125 (Special Issue): 26-41
Copyright by Medycyna Praktyczna,
Kraków 2015