# Jak oceniać skuteczność psychoterapii?

## Rekomendacje Koalicji dla Psychoterapii

2026

Uniwersytet SWPS
UNIWERSYTET JAGIELLOŃSKI W KRAKOWIE
UNIWERSYTET WARSZAWSKI
Instytut Psychologii Polska Akademia Nauk

# Stanowisko Koalicji dla Psychoterapii

## w sprawie kryteriów oceny skuteczności psychoterapii

Przyjęte dnia 20 maja 2026 roku

# 1. Wprowadzenie

Koalicję dla Psychoterapii (KdP) tworzą jednostki naukowe posiadające najwyższą ocenę (A+) jakości prowadzonych badań naukowych w dyscyplinie psychologii, tj. Instytut Psychologii Polskiej Akademii Nauk, Uniwersytet Jagielloński, Uniwersytet SWPS oraz Uniwersytet Warszawski.

Do najważniejszych celów KdP należy promowanie podejścia opartego na dowodach naukowych w psychoterapii oraz upowszechnianie wiedzy o skutecznych metodach psychoterapii. Jednym z podstawowych kroków w realizacji tych celów jest odpowiedź na pytanie o kryteria oceny skuteczności psychoterapii. KdP postuluje opracowanie kryteriów oceny skuteczności psychoterapii, które uwzględniałyby wyniki badań ilościowych i jakościowych oraz polski kontekst społeczny.

Zgodnie z definicją APA, praktyka psychologiczna oparta na dowodach polega na „integracji najlepszych dostępnych badań z wiedzą kliniczną w kontekście cech, środowiska kulturowego i preferencji pacjenta" (APA, 2006, s. 273).

Optymalny sposób oceny skuteczności psychoterapii powinien uwzględniać wyniki badań ilościowych i jakościowych, zgodnie z założeniem, że różne pytania badawcze wymagają różnych metodologii.

Ze względu na to, że jest to zadanie obszerne i złożone, KdP w niniejszym stanowisku skupia się na kryteriach ilościowych, jedynie wzmiankując o kryteriach jakościowych. KdP w niedalekiej przyszłości planuje stworzenie dokumentu, który obejmie te dwie grupy dowodów naukowych.

Zdaniem KdP w debacie nad naukowymi podstawami psychoterapii z wykorzystaniem metod ilościowych należy brać pod uwagę model oceny skuteczności psychoterapii i tworzenia rekomendacji zaproponowany przez Davida Tolina i współpracowników (2015), który jest stosowany przez Amerykańskie Towarzystwo Psychologiczne. Streszczenie tego modelu stanowi załącznik do niniejszego stanowiska.

# 2. Kluczowe rekomendacje

**Jednocześnie, w toku dyskusji wokół modelu Tolina i współpracowników (2015), KdP zaleca rozważenie następujących zagadnień:**

1. Warto pamiętać, że opublikowana rekomendacja może ulec zmianie w przypadku pojawienia się nowych dowodów za lub przeciwko stosowaniu danej metody psychoterapii.

2. Przy rozważaniu zastosowania danej metody należy brać pod uwagę nie tylko obowiązujące rekomendacje, lecz także wyniki nowych badań opublikowanych po ich wydaniu.

3. Zalecamy opisywanie metod, których koncepcje teoretyczne są zgodne z ustaleniami psychologii naukowej.

4. Do ograniczeń rekomendacji wynikających z modelu Tolina i wsp. (2015) należą: opóźnienie (nawet kilkuletnie) między czasem publikacji dowodów a przygotowaniem rekomendacji, brak wpływu na przedmiot przyszłych badań (analiza dotyczy dostępnych badań, a nie wskazuje, jakie badania są potrzebne). Ponadto, warto zwrócić uwagę, że rekomendacje bazują na badaniach, które również mają szereg ograniczeń, w tym organizacyjnych i finansowych, uniemożliwiających prowadzenie badań w perspektywie wieloletniej, np. niewielki udział RCT w odniesieniu do psychoterapii długoterminowych, niewielka liczba pomiarów długoterminowej trwałości efektów psychoterapii (np. powyżej roku), nadreprezentacja wykorzystania niezłożonych wskaźników zmiany (w przeciwieństwie do złożonej perspektywy pacjentów), w tym krótkich kwestionariuszy objawowych, a w mniejszym stopniu wykorzystywanie wywiadów diagnostycznych i informacji ze środowiska (rodzina, pracodawca, ZUS), wskaźników neurobiologicznych i ocen funkcjonowania w życiu codziennym.

5. Zalecamy, aby zespoły ekspertów opracowujące rekomendacje w oparciu o model Tolina składały się z osób dysponujących szerokim spektrum wysokich kompetencji i bogatym doświadczeniem zawodowym. Powinny one obejmować zarówno ekspertów o zaawansowanych kwalifikacjach naukowych w zakresie badań nad psychoterapią, jak i superwizorów reprezentujących różne nurty psychoterapeutyczne.

6. Warto rozważyć nadawanie większej rangi miarom niezwiązanym z subiektywną oceną (np. miary psychofizjologiczne, neuroobrazowanie) względem kwestionariuszy samoopisowych, co może być elementem modyfikującym finalną ocenę rekomendacji.

7. W zgodzie z duchem propozycji Tolina i wsp. (2015), równolegle do rekomendacji dla metod stworzonych do pomocy osobom z danym problemem/diagnozą, zalecamy tworzenie rekomendacji dla mechanizmów zmiany opartych na dowodach w odniesieniu do transdiagnostycznych procesów.

8. Niniejsze rekomendacje dotyczą nie tylko oceny skuteczności metod stosowanych w pracy bezpośrednio z osobami korzystającymi z psychoterapii (na żywo czy online), ale także szeroko rozumianych interwencji psychologicznych w formie cyfrowej (np. aplikacji).

9. W oryginalnych rekomendacjach Tolina zalecane jest korzystanie z AMSTAR do oceny przeglądów i metaanaliz. Obecnie powstał uaktualniony system - AMSTAR-2 (Shea i wsp., 2017) i zalecamy używanie tej wersji.

**10** Zalecamy wspieranie rozwoju dynamicznego agregowania danych („living databases"), które umożliwia bieżące aktualizowanie baz wiedzy o skuteczności psychoterapii oraz szeroki, dostępny dla praktyków, dostęp do wniosków z badań — także dla osób bez formalnego przygotowania metodologicznego (Sakaluk i wsp., 2023).

**11** Postulujemy rozwój badań nad psychoterapią w Polsce, ze szczególnym uwzględnieniem badań nad jej skutecznością oraz mechanizmami zmiany, uwzględniających specyfikę krajowego kontekstu klinicznego i systemowego. Wskazujemy, że warunkiem koniecznym tego rozwoju jest stworzenie stabilnych mechanizmów finansowania badań nad psychoterapią.

## Konkluzja

Postulujemy, aby w debacie nad psychoterapią uwzględniać – w odniesieniu do badań ilościowych – kryteria oceny skuteczności i bezpieczeństwa opisane przez D. Tolina, a w przyszłości opracować również kryteria dla badań jakościowych oraz kryteria uwzględniające polski kontekst kulturowy. Jakość tworzonych rekomendacji zależy również od dostępności badań prowadzonych w Polsce. W związku z tym postulujemy stworzenie klarownych mechanizmów finansowania takich badań.

### Literatura cytowana

Sakaluk, J. K., De Santis, C., Kilshaw, R., Pittelkow, M. M., Brandes, C. M., Boness, C. L., ... & van Ravenzwaaij, D. (2023). Reconsidering what makes syntheses of psychological intervention studies useful. Nature Reviews Psychology, 2(9), 569-583.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., ... & Henry, D. A. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ, 358.

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. Clinical Psychology: Science and Practice, 22(4), 317.

Tolin, D. F., Grasso, D., Boness, C. L., Beck, J. G., Keane, T. M., Leichsenring, F., ... & Weinand, J. (2024). A proposed definition of psychological treatment and its relation to empirically supported treatments. Clinical Psychology: Science and Practice, 32(3), 213.

ZAŁĄCZNIK

# Kryteria oceny skuteczności psychoterapii według modelu Tolina

## Opracowanie przygotowane przez członków Koalicji dla Psychoterapii

Przyjęte dnia 20 maja 2026 roku

Opracowanie stanowi streszczenie modelu oceny skuteczności psychoterapii i tworzenia rekomendacji zaproponowanego przez Davida Tolina i współpracowników (2015), który jest stosowany przez Amerykańskie Towarzystwo Psychologiczne.

Tekst przedstawia model Davida Tolina i współpracowników (2015),
który proponuje kryteria oceny skuteczności psychoterapii.
Opracowanie powstało przede wszystkim na podstawie artykułu Tolina i wsp. (2015).

# 1. Wprowadzenie

Model Tolina i wsp. (2015) opiera się na założeniu, że ocena skuteczności psychoterapii powinna uwzględniać wszystkie dostępne badania (także z wynikami negatywnymi i te prowadzone w warunkach codziennej praktyki terapeutycznej, a nie tylko rygorystyczne badania kliniczne) oraz brać pod uwagę zarówno efekt natychmiastowy psychoterapii, jak i ocenę utrzymania się efektów psychoterapii po jej zakończeniu. Wynikiem badań są m.in. wskaźniki wielkości efektu. Obecnie uznaje się, że nie można przyjąć konkretnego punktu odcięcia wielkości efektu, który oznaczałby, że dana metoda psychoterapii jest skuteczna; zależy on od kontekstu, w tym rodzaju zaburzenia i celów psychoterapii. Ważnym kryterium tworzenia rekomendacji powinna być ocena efektywności kosztowej.

Prowadzenie psychoterapii opartej na dowodach zakłada korzystanie z doświadczenia praktycznego oraz wiedzy płynącej z badań naukowych, a także osobistych preferencji i wartości osoby korzystającej z psychoterapii. Niniejsze rekomendacje stanowią zatem jedno ze źródeł informacji, jakie terapeuta powinien brać pod uwagę w toku swojej praktyki.

# 2. Kryteria oceny skuteczności psychoterapii

Model Tolina i wsp. (2015) zakłada, że ocena skuteczności
psychoterapii i tworzenia rekomendacji odbywa się w dwóch krokach.

## 2.1. Krok 1: Analiza systematycznych przeglądów badań

### 2.1.1. Nominacja metody i wskazanie bazy dowodów

Nominujący może dostarczyć istniejące przeglądy lub stworzyć nowy przegląd w celu przygotowania rekomendacji, jednak wszystkie przeglądy należy dokładnie ocenić pod kątem kompletności i ryzyka tendencyjności. Opublikowane lub nieopublikowane przeglądy systematyczne, które nie zostały uznane za spełniające rygorystyczne standardy jakości, nie powinny być brane pod uwagę.

### 2.1.2. Powołanie zespołu oceniającego dowody

Proces wyboru członków zespołu oceniającego dowody powinien być przejrzysty i obejmować otwarty proces nominacji, publiczne udostępnienie listy członków oraz zapewniać różnorodność w zakresie ich doświadczeń i kompetencji. Zespoły dokonujące przeglądów powinny:

a. składać się z osób o szerokim zakresie udokumentowanej wiedzy i doświadczenia,
b. ujawniać rzeczywiste i potencjalne konflikty interesów,
c. utrzymywać atmosferę otwartości i swobodnej wymiany poglądów,
d. posiadać jasno określone procesy i metody działania.

### 2.1.3. Ocena jakości przeglądu

Ocena metod psychoterapii powinna zostać przeprowadzona na podstawie istniejących (opublikowanych lub nieopublikowanych) ilościowych przeglądów badań przez powołany do tego zespół. Istnieje wiele sposobów oceny, czy systematyczny przegląd został przeprowadzony z wystarczającą przejrzystością i rygorem, aby zapewnić pewność, że jego wyniki są kompleksowe i odzwierciedlają najlepsze dostępne dowody, zgodnie z Cochrane Handbook (Higgins i Green, 2008) oraz Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Liberati i in., 2009).

Proponowane jest korzystanie z listy kontrolnej AMSTAR (Shea i in., 2007), której wynik nie jest sumowany w pojedynczy wskaźnik, co oznacza, że nie ma określonego progu, po przekroczeniu którego przegląd uznawany jest za wiarygodny. Zamiast tego elementy listy kontrolnej są wykorzystywane do subiektywnej decyzji zespołu, kiedy systematyczny przegląd jest wystarczająco wysokiej jakości i odpowiednio opisany, aby mógł zostać wykorzystany.

W niektórych przypadkach przegląd systematyczny może obejmować badania, w których psychoterapie różniły się między sobą pod względem sposobu stosowania interwencji, cech terapeuty, czasu trwania psychoterapii lub okresu obserwacji. W niektórych badaniach pewne elementy psychoterapii mogą być dodane lub usunięte, co powoduje dylemat dla członków zespołu – muszą oni ocenić, czy istnieje wystarczające podobieństwo między badaniami, aby uznać je za testujące tę samą zasadniczą metodę psychoterapii.

Pewien stopień heterogeniczności klinicznej (rozumianej jako zróżnicowanie sposobów stosowania metod, cech terapeutów, długości trwania psychoterapii czy follow up) musi być przewidziany i zaakceptowany; w przeciwnym razie niewiele grup badań mogłoby być włączone do przeglądu. Należy jednak dokładnie rozważyć, w jakim stopniu heterogeniczność kliniczna wpływa negatywnie na interpretację jednego zbiorczego wyniku ilościowego, zanim zespół rozważy przeprowadzenie metaanalizy. Standardowa procedura przeglądu powinna obejmować zgodę członków zespołu co do tego, że wszystkie wybrane badania są wystarczająco podobne, aby można było uznać je za odzwierciedlające jedną metodę psychoterapii.

Wykorzystanie systematycznych przeglądów nie wyklucza uwzględnienia badań pojedynczych przypadków, ponieważ tego typu badania, przy zastosowaniu odpowiedniej kontroli eksperymentalnej, mogą wykazać przyczynowość w sposób porównywalny z randomizowanymi kontrolowanymi badaniami (RCT). Mimo że mniejsza liczba uczestników może ograniczać uogólnienie wyników, opracowano metody obliczania wielkości efektu takich badań oraz modelowania bayesowskiego i wielopoziomowego.

Ocena jakości badań pojedynczych przypadków mogłaby wykorzystywać opublikowane wskaźniki jakości w sposób analogiczny do procedur oceny RCT. Zgodnie z aktualnymi podejściami do syntezy dowodów, nie jest jednak zalecane opieranie rekomendacji wyłącznie na dowodach pochodzących z badań pojedynczych przypadków.

Wyniki przeglądu powinny bazować głównie na syntezie danych z większych badań klinicznych.

Skuteczność danej metody psychoterapii może być oceniana na podstawie:

a. porównań z dobrze znanymi i ilościowo określonymi nieaktywnymi warunkami kontrolnymi, do których należą:
   i. lista oczekujących,
   ii. placebo farmakologiczne,
   iii. placebo psychologiczne

lub

b. porównań z alternatywnymi metodami psychologicznymi.

### 2.1.4. Ocena przydatności klinicznej

Kluczowe dla trafności zewnętrznej (możliwości uogólniania) wyników przeglądu systematycznego jest odpowiednie sformułowanie pytania badawczego. Tolin i wsp. (2015) rekomendują stosowanie akronimu **PICOTS**:

**P:** Populacja, z której pochodzi próba osób badanych,

**I:** Interwencja, która została zastosowana w badaniu,

**C:** Comparison bądź Comparator – warunek kontrolny, aktywny bądź nieaktywny, do którego została porównana interwencja,

**O:** Outcome, czyli mierzony wynik oddziaływania,

**T:** Timeline, czyli określona rama czasowa (12 tygodni, 6 miesięcy, długi termin, etc.),

**S:** Setting – środowisko, w którym prowadzona była interwencja (np. oddział psychiatryczny vs poradnia ambulatoryjna).

Dodatkowo, w przeglądzie należy uwzględnić ocenę badań pod kątem treatment fidelity (tj. zgodności stosowanych metod terapeutycznych z teoretycznymi założeniami czy protokołem). Tolin i wsp. (2015) nie przywołują konkretnego narzędzia do oceny treatment fidelity, ale wymieniają takie elementy jak ocena kompetencji i wyszkolenia terapeutów, użycie protokołu psychoterapii i ocena, na ile prowadzone interwencje były zgodne z protokołem.

### 2.1.5. Ocena ryzyka tendencyjności (risk of bias)

Ocena "metodologicznej adekwatności" badania jest ważna dla wewnętrznej i zewnętrznej trafności badania klinicznego. Tolin i wsp. (2015) rekomendują, żeby stosować Cochrane Risk of Bias Tool (Higgins et al., 2011). To narzędzie pozwala ocenić ryzyko tendencyjności związane z:

a. adekwatną randomizacją i przydziałem do warunków badania (allocation sequence generation),
b. ukryciem informacji o przydziale do grup (allocation concealment),
c. zaślepieniem uczestników badania, osób prowadzących interwencję oraz osób dokonujących pomiaru wyników psychoterapii (blinding),
d. brakami danych,
e. selektywnym raportowaniem wyników badania (selective outcome reporting),
f. innymi źródłami tendencyjności.

Użycie Cochrane Risk of Bias Tool umożliwia ocenę każdego z wyżej wymienionych elementów na skali: niskie ryzyko tendencyjności, niejasne ryzyko tendencyjności, wysokie ryzyko tendencyjności. Ocena tendencyjności jest oceną jakościową i przez to mogą pojawiać się niespójności i różne oceny w zakresie tego, jakie ryzyko tendencyjności niesie ze sobą zastosowanie określonego rozwiązania metodologicznego. Dlatego do oceny w tym zakresie potrzebna jest wiedza z zakresu metod prowadzenia badań klinicznych oraz określenie, czy dana metoda mogła doprowadzić do zwiększenia ryzyka tendencyjności.

## 2.2. Krok 2: Ocena dowodów przez ekspertów w oparciu o kryteria GRADE

Eksperci oceniają dowody zebrane w toku systematycznego przeglądu w celu dokonania rekomendacji. Oceny dokonują za pomocą zmodyfikowanego systemu GRADE (ang. Grading of Recommendations Assessment, Development, and Evaluation): **https://www.gradeworkinggroup.org**.

### 2.2.1. Ocena jakości poszczególnych badań

W pierwszej kolejności eksperci oceniają jakość badań uwzględnionych w systematycznym przeglądzie lub metaanalizie. Co do zasady, badania RCT są uznawane za wysokiej jakości, ale ocena ta może zostać obniżona.

Punktem wyjścia dla innych rodzajów badań jest ocena umiarkowanej lub niskiej jakości, ale ocena konkretnego badania może zostać podniesiona. Ocena jakości badania może zostać obniżona, kiedy ma następujące wady:

a. ograniczenia badania (np. brak standaryzacji procedury, nierzetelne i nietrafne narzędzia, p-hacking, niepoprawne analizy statystyczne, manipulacja danymi (ang. cherry-picking), zastosowanie celowo osłabionej interwencji w grupie kontrolnej (ang. straw therapy);
b. niespójność wyników - odnosi się do stopnia heterogeniczności wyników. W metaanalizie stosujemy miarę I2, która określa stopień, w jakim wyniki poszczególnych badań różnią się między sobą w sposób większy, niż można oczekiwać na podstawie samego błędu losowego. Zazwyczaj wartość 25% i mniej wskazuje na niską heterogeniczność, między 25% a 50% umiarkowaną, 50-75% wysoką i 75-100% bardzo wysoką. Należy również nadmienić, że wskaźnik I2 spełnia swoją funkcję, gdy jest prezentowany i interpretowany wraz z przedziałem ufności (Thorlund i in., 2012);
c. pośredniość dowodów dotyczy sytuacji, kiedy charakterystyka pacjentów, miar efektu lub zastosowanych interwencji odbiega od tych, które są przedmiotem zainteresowania oraz gdy porównania są niebezpośrednie, np. porównanie metody psychoterapii A i B z placebo, zamiast bezpośredniego porównania A z B;
d. nieprecyzyjność oznacza, że wielkość próby powinna być odpowiednio duża, aby wykryć efekt istotny klinicznie z mocą statystyczną na poziomie 0,80 oraz poziomem istotności α= 0,05; nieprecyzyjność jest związana również z szerokim przedziałem ufności wyników;
e. selektywne raportowanie wyników (ang. reporting bias).

### 2.2.2. Ocena całości zebranych dowodów

W drugiej kolejności eksperci analizują całość zebranych dowodów. System GRADE proponuje trzystopniową skalę oceny jakości zebranych dowodów:

a. przegląd badań wysokiej jakości – daje ekspertom dużą pewność, że prawdziwy efekt psychoterapii jest zbliżony do estymacji przedstawionych w badaniach. Na przykład, badania takie muszą spełnić następujące kryteria:
   i. istnieje szeroki zakres badań uwzględnionych w analizach, które nie mają poważnych ograniczeń,
   ii. istnieje niewielka wariancja wyników między badaniami,
   iii. estymator efektu psychoterapii ma wąski przedział ufności.
b. przegląd badań umiarkowanej jakości - prawdziwy efekt psychoterapii może być umiarkowanie różny

od estymacji przedstawionych w badaniach. Na przykład, badania możemy tak ocenić, gdy spełnione będzie przynajmniej jedno z kryteriów:

i. istnieje kilka badań i niektóre z nich mają wady, ale nie są one poważne,
ii. istnieje pewna wariancja wyników między badaniami lub estymator efektu psychoterapii ma szeroki przedział ufności.

c. przegląd badań niskiej jakości – prawdziwy efekt psychoterapii może być znacząco różny od estymacji przedstawionych w badaniach. Na przykład, spełnienie jakiekolwiek z poniższych kryteriów może prowadzić do niskiej oceny jakości badań:

i. badania mają poważne wady,
ii. istnieje istotna wariancja wyników między badaniami,
iii. estymator efektu psychoterapii ma bardzo szeroki przedział ufności.

### 2.2.3. Dokonanie rekomendacji

W trzeciej kolejności, po dokonaniu oceny jakości pojedynczych badań oraz całej bazy dowodów, należy ocenić spełnienie szeregu kolejnych kryteriów, które umożliwią dokonanie rekomendacji. Ocena jest opisywana jako trzystopniowa, ale w zasadzie ma cztery poziomy:

a. bardzo silna rekomendacja - na ten moment określana raczej jako „aspiracyjna", ponieważ niewiele psychoterapii spełni te kryteria. Wszystkie kryteria do spełnienia to:

i. wysokiej jakości dowody, że psychoterapia ma klinicznie istotny pozytywny wpływ w przypadku danego zaburzenia na (A) na objawy oraz (B) funkcjonowanie, a (C) efekty te utrzymują się przynajmniej 3 miesiące po zakończeniu psychoterapii,
ii. przynajmniej jedno dobrze przeprowadzone badanie w warunkach naturalnych (inny kontekst niż rygorystyczne badania laboratoryjne),
iii. relatywnie niskie ryzyko szkodliwości,
iv. rozsądny nakład zasobów (efektywność kosztowa).

b. silna rekomendacja. Kryteria do spełnienia to:

i. umiarkowane lub wysokiej jakości dowody, że psychoterapia ma klinicznie istotny pozytywny wpływ w przypadku danego zaburzenia na (A) na objawy i/lub (B) funkcjonowanie,
ii. pozytywna równowaga między korzyściami a możliwym ryzykiem szkodliwości.

c. słaba rekomendacja wskazuje, że dowody na jej skuteczność istnieją, ale są one niskiej jakości; nie jest intencją tej oceny, aby zniechęcać do korzystania z danej metody. Powinno być spełnione przynajmniej jedno kryterium:

i. niskiej lub bardzo niskiej jakości dowody, że psychoterapia ma klinicznie istotny pozytywny wpływ w przypadku danego zaburzenia na (A) na objawy i/lub (B) funkcjonowanie,
ii. umiarkowane lub wysokiej jakości dowody na istotną statystycznie (ale nie istotną klinicznie) efektywność.

Nie może być wydana żadna rekomendacja, jeśli nie została dokonana ocena dowodów ze względu na brak systematycznego przeglądu lub dostępne badania nie spełniają minimalnych kryteriów dla słabej rekomendacji.

## 2.3. Podsumowanie modelu oceny skuteczności psychoterapii i tworzenia rekomendacji Tolina i współpracowników (2025)

**Krok 1: Analiza systematycznych przeglądów badań**

- Nominacja metody i wskazanie bazy dowodów, tj. istniejącego przeglądu (lub przeglądów) systematycznego lub jego przeprowadzenie
- Powołanie zespołu oceniającego dowody
- Ocena jakości przeglądu (np. AMSTAR)
- Ocena przydatności klinicznej (PICOTS)
- Ocena ryzyka tendencyjności

**Krok 2: Ocena dowodów przez zespół oceniający w oparciu o kryteria GRADE**

- Ocena jakości poszczególnych badań
- Ocena całości zebranych dowodów
- Dokonanie rekomendacji (bardzo silnej, silnej, słabej) lub braku rekomendacji

*Opracowanie powstało na podstawie artykułu Tolina i wsp. (2015). Przygotowali je członkowie Koalicji dla Psychoterapii (maj 2025): Przemysław Bąbel, Grzegorz Iniewicz, Joachim Kowalski, Maciej Pilecki oraz Rafał Styła.*

## Literatura cytowana


Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Medicine, 6(7), e1000100.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology, 7, 10.

Thorlund, K., Imberger, G., Johnston, B. C., Walsh, M., Awad, T., Thabane, L., ... & Wetterslev, J. (2012). Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PloS One, 7(7), e39471.

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. Clinical Psychology: Science and Practice, 22(4), 317.