

Jak czytać metaanalizy badań nad skutecznością i nie zablądzić.

Wprowadzenie dla osób praktykujących psychoterapię

Joachim Kowalski¹

*Instytut Psychologii Polskiej Akademii Nauk,
Pracownia Psychopatologii Eksperymentalnej*

<https://orcid.org/0000-0001-6281-7401>

Streszczenie

Cel: Celem artykułu jest zilustrowanie praktycznych zagadnień związanych z przygotowaniem oraz interpretacją przeglądów systematycznych i metaanaliz dotyczących badań klinicznych nad skutecznością psychoterapii. Tekst pełni funkcję praktycznego przewodnika i mapy, które ułatwiają orientację w sposobie powstawania tych opracowań oraz pozwalają krytycznie oceniać ich wyniki.

Tezy: 1) **Znaczenie przeglądów systematycznych i metaanaliz.** Metaanalizy i przeglądy systematyczne są kluczowymi metodami syntezy danych w psychologii i stanowią podstawę decyzji klinicznych opartych na dowodach naukowych. W obszarze psychoterapii liczba publikowanych prac przeglądowych sięga setek rocznie. 2) **Zróżnicowana jakość badań przeglądowych.** Przeglądy różnią się jakością metodologiczną i poziomem ryzyka tendencyjności, co wpływa na pewność formułowanych wniosków oraz możliwość ich zastosowania w praktyce. 3) **Etapy przygotowania i raportowania.** Artykuł opisuje formalne kroki tworzenia przeglądów systematycznych i metaanaliz, w tym definiowanie pytania badawczego, znaczenie prerejestracji oraz stosowanie standardów raportowania. 4) **Wskaźniki efektywności i pewności wyników.** Omówiono najczęściej stosowane wskaźniki efektu (np. standaryzowane różnice średnich, iloraz szans, liczbę potrzebnych interwencji, procent remisji lub indeksy zmiany oparte na punktach odcięcia) oraz pewności co do uzyskanych wyników, jak miary tendencyjności i heterogeniczności. 5) **Elementy graficzne i analizy dodatkowe stosowane w metaanalizach.** Przedstawiono graficzne reprezentacje wyników metaanaliz oraz metody analityczne ukierunkowane na redukcję tendencyjności.

¹ Adres do korespondencji: jkowalski@psych.pan.pl.

Konkluzje: Przeglądy systematyczne i metaanalizy stanowią fundament praktyki klinicznej bazującej na dowodach i odgrywają istotną rolę w formułowaniu zaleceń terapeutycznych w psychoterapii. Jednak ich interpretacja wymaga świadomości procesów stojących za ich powstawaniem oraz umiejętności krytycznej oceny jakości i ograniczeń tych prac.

Słowa kluczowe: metaanaliza, przegląd systematyczny, badania kliniczne, metapsychologia

Badania kliniczne dostarczają informacji na temat bezpieczeństwa i efektywności psychoterapii, ograniczając wpływ zniekształceń poznawczych, które negatywnie oddziałują na decyzje podejmowane w codziennej praktyce klinicznej. Chodzi o takie tendencje poznawcze jak iluzoryczna korelacja, efekt pewności wstecznej, błąd potwierdzenia, efekt autorytetu, podstawowy błąd atrybucji i wiele innych. Badania kliniczne, których celem jest systematyczne gromadzenie danych za pomocą metody naukowej, są częściowo wolne od niektórych z tych tendencji. Częściowo oraz niektórych, ale należy podkreślić, że mają w tym względzie znaczącą przewagę nad niesystematycznymi obserwacjami. Wypowiedzi typu „Nie potrzebuję żadnych badań – ja widzę, że u mnie w gabinecie to działa” nie padną z ust osoby, która posiada wiedzę w zakresie psychologii poznawczej i społecznej, szczególnie zniekształceń poznawczych w praktyce klinicznej (prace przeglądowe w: Bowes i in., 2020 czy Lilienfeld i in., 2014) czy dotyczącą rozbieżności między ocenami klinicystów co do swojej efektywności, a faktyczną efektywnością (np. Hannan i in., 2005; Kraus i in., 2011; Walfish i in., 2012).

Istnieje wiele rodzajów badań klinicznych terapii psychologicznych², z których wszystkie mają istotne znaczenie dla budowania rzetelnej wiedzy o użyteczności terapii (APA Presidential Task Force..., 2006). W praktyce bazującej na dowodach (*evidence-based practice*, EBP) uznaje się, że randomizowane badania kliniczne z grupą kontrolną (*randomized controlled trial*, RCT) oraz ich syntezy są najistotniejszym źródłem danych o efektywności interwencji (np. Burns i in., 2011; Evans, 2003). W przypadku psychoterapii również (APA Presidential Task Force, 2006; Dozois i in., 2014). Dodatkowo podkreśla się znaczenie RCT obarczonych niskim ryzykiem tendencji (np. Jobst i in., 2016), szczególnie w przypadku badań nad wyizolowanymi problemami klinicznymi (Philips i Falckenström, 2021).

² Definicja za Tolinem i współpracownikami (2025): „Terapia psychologiczna to interwencja polegająca na specyficznych oddziaływaniach pomiędzy osobą lub osobami a specjalistą/tką zdrowia psychicznego lub osobą wyznaczoną przez specjalistę/tkę. Intencją tych oddziaływań jest zaangażowanie poznawczych, emocjonalnych, behawioralnych lub interpersonalnych procesów, a celem modyfikacja zdrowia lub funkcjonowania, a których podstawowe założenia dotyczące procedur i mechanizmów są oparte na naukowej psychologii i spójne z naukowym sposobem rozumienia zjawisk”. W niniejszym artykule – ze względu na mnogość definicji i dyskusji wokół nich – gdy mowa o psychoterapii, należy przyjąć, że chodzi o terapię psychologiczną w przytoczonym wyżej znaczeniu.

Jednakże proste wyszukiwanie w bazie PubMed frazy *psychotherapy AND (trial OR case)* pokazuje, że od 2013 roku publikowanych jest ponad 4000 prac naukowych rocznie zawierających te terminy. Nawet wykluczając znaczącą większość jako niezwiązane z interesującą nas tematyką, szybko zorientujemy się, że bycie na bieżąco z literaturą na temat badań nad psychoterapią przekracza możliwości jednej osoby. Tym samym staniemy przed tym samym problemem, przed którym stanęli twórcy współczesnej koncepcji metaanalizy (Shadish i Lecy, 2015): jak zagregować i zsyntetyzować ogromną ilość danych, gdy tradycyjna metoda narracyjnego przeglądu³ badań przestaje wystarczać? Warto tutaj wspomnieć, że jedna z pierwszych współczesnych metaanaliz powstała w celu podsumowania wyników kilkuset badań nad efektywnością psychoterapii (Smith i Glass, 1977).

Systematyczny przegląd (*systematic review*) to praca, która ma na celu jak najbardziej wyczerpujące zebranie, wyekstrahowanie i syntezę danych na jakiś temat. Zaś metaanaliza, czyli „analiza analiz”, to statystyczna agregacja efektów wielu pojedynczych badań w celu integracji płynących z nich wniosków (Glass, 1976). Obecna praca odnosi się zarówno do przeglądów systematycznych, jak i metaanaliz, ale dla ułatwienia przyjęto w niniejszej pracy, że określenie „metaanaliza” odnosi się do obu rodzajów prac, jako że nieodzownym elementem metaanalizy niezależnych badań jest przeprowadzenie systematycznego przeglądu.

W badaniach nad psychoterapią integracja wniosków dotyczy przede wszystkim ich efektywności, ale również innych zjawisk, przykładowo efektów niepożądanych (np. Moritz i in., 2019) czy cech procesu terapeutycznego, które wpływają na efektywność tego procesu (np. Ciharova i in., 2024). Poza agregacją dużej ilości danych i integracją wniosków płynących z wielu badań, tworzenie metaanaliz ma również kilka innych zalet. Jedną z nich jest dokładniejsza i bardziej bezstronna ocena uzyskiwanych w badaniach efektów, gdyż badacze mają tendencję do prezentowania uzyskiwanych przez siebie wyników w pozytywnym świetle, nawet jeśli same dane niekoniecznie takie pozytywne spojrzenie by uzasadniały (Cuijpers i Cristea, 2016; Stoll i in., 2020). Kolejną jest to, że pozwala ona osiągnąć większą moc statystyczną (tj. większe prawdopodobieństwo wykrycia w badaniu faktycznie istniejącego efektu) przez agregację wielu obserwacji oraz zawężenie błędu pomiaru wyników (Cohn i Becker, 2003). Dodatkową zaletą jest przejrzystość prezentowania problemów badawczych i usystematyzowany sposób prezentacji wyników tego rodzaju badań.

Niemniej jednak, należy wspomnieć również o wadach metaanaliz. Jedną z kluczowych jest przeszacowywanie uzyskiwanych w nich wielkości efektów, mimo prób miarkowania tendencyjności wynikających z nieuwzględniania wszystkich badań (Kvarven i in., 2020). Co więcej, wyszukiwanie ograniczone do tytułów i abstraktów w bazie PubMed frazy *psychotherapy AND (meta-analysis OR systematic review)* pokazuje, że tylko w zeszłym roku opublikowano 284 takie prace. Oznacza to, że bycie na bieżąco z metaanalizami na temat psychoterapii

³ Narracyjny przegląd to metoda syntezy badań, w której opisuje się kolejno wyniki przedstawione w poszczególnych pracach. Zdecydowanym ograniczeniem tej metody jest, że niektóre wyniki mogą być uwypuklane bez odpowiedniego uzasadnienia, np. w oparciu o preferencje badacza (Higgins i in., 2024).

jest trudnym i czasochłonnym zadaniem, a zadanie to staje się trudniejsze, gdy weźmiemy pod uwagę, że różnią się one między sobą jakością.

Dlatego też celem niniejszej pracy jest syntetyczne omówienie procesu przygotowywania metaanaliz, interpretacji ich wyników, z uwzględnieniem wskaźników efektywności, wskaźników dotyczących jakości danych oraz elementów graficznych. Kolejnym celem jest pokazanie znaczenia metaanaliz w praktyce klinicznej oraz tworzeniu zaleceń terapeutycznych (*treatment guidelines*). Ze względu na obszerność poruszanych wątków pewna skrótowość jest nieunikniona. Osoby, które chciałyby w sposób bardziej wyczerpujący zrozumieć zagadnienie agregacji wiedzy, mogą sięgnąć po pozycje książkowe (np. Higgins i in., 2024; Rothstein i in., 2005). Niniejsza praca ma być mapą oraz przewodnikiem, zaadresowanymi do osób praktykujących psychoterapię, które chciałyby, zgodnie z duchem EBP, wykorzystywać dowody naukowe (Spring, 2007). Szczególnie przydatna może być dla tych, którzy czują, że może istnieć rozbieżność między ich przygotowaniem do zawodu a umiejętnościami rozumienia badań naukowych. Cel pracy zostanie osiągnięty jeśli osoba czytająca metaanalizę, będzie w stanie zrozumieć jej cel, uzyskane wyniki, ich ograniczenia, płynące z tego wszystkiego wnioski – i przy tym *nie zablądzić*.

Jak są tworzone metaanalizy?

Pytanie badawcze, prerejestracja i standardy raportowania

Pytanie badawcze – czego chcemy się dowiedzieć?

Odpowiednio zadane pytanie jest kluczowym elementem metaanalizy, bowiem to od niego zależy, jak i jakiego rodzaju danych będziemy poszukiwać, a następnie interpretować. Najpopularniejszym oraz rekomendowanym (np. Tolin i in., 2015) sposobem jest posłużenie się akronimem PICOTS. Szczegółowe rozwinięcie akronimu przedstawione zostało w tabeli 1 (s. 33), a poniżej pokazany jest przykład jego zastosowania w kontekście psychoterapii depresji. Należy dodać, że „S” z tego akronimu może być również interpretowane jako *study design* czy *type*, czyli informacja o rodzaju interesujących nas badań (np. ilościowe) (por. Methley i in., 2014).

Parafrazując Chmielowskiego, osoba nieobeznana z EBP mogłaby rzec, że „psychoterapia depresji, jaka jest, każdy widzi”. Jednakże będziemy opierać się na zupełnie innych dowodach, gdy chcemy wiedzieć jak skuteczna jest:

- a. indywidualna i ambulatoryjna (S) psychoterapia poznawcza (I) adolescentów z pierwszym epizodem depresyjnym (P) w porównaniu do psychoterapii wspierającej (C) w redukcji nasilenia trudności interpersonalnych (O) w okresie roku od zakończenia terapii (T)?
- b. grupowa i prowadzona na oddziale psychiatrycznym (S) psychoterapia poznawcza (I) osób z lekooporną depresją (P) w porównaniu do oczekiwania na terapię (C) w redukcji nasilenia myśli i częstości prób „S” (O) w okresie tuż po zakończeniu leczenia (T)?

Powyższy przykład dwóch pytań badawczych, w którym elementami wspólnymi są tylko główny problem kliniczny oraz metoda psychoterapii, pozwala

unaocznąć wagę dokładnego sformułowania pytania badawczego. Zostało to zauważone już w latach 90. XX wieku, z praktyczną myślą, że dokładne sformułowanie pytania przekłada się na trafne decyzje w praktyce klinicznej (Richardson i in., 1995 za: Davies, 2011).

Tabela 1

Rozwinięcie i objaśnienie akronimu PICOTS (na podstawie Daview, 2011 i Tolin i in., 2015)

Znaczenie	Wyjaśnienie
P <i>Population</i>	Populacja osób, która brała udział w badaniu interwencji; np. osoby z konkretnym rozpoznaniem (np. osoby z fobią społeczną), określonym stopniem zaawansowania (np. osoby po pierwszym epizodzie psychozy), danymi charakterystykami (osoby, których rodzice stosowali przemoc fizyczną) czy określonych okolicznościach życiowych (np. uchodźcy wojenni)
I <i>Intervention</i>	Interwencja, która była przedmiotem badania; np. psychoterapia poznawczo-behawioralna, psychoterapie z komponentą treningu regulacji emocji, internetowe psychoterapie bez wsparcia psychoterapeuty
C <i>Comparator</i>	Warunek kontrolny, do którego porównywane są efekty stosowanej interwencji, np. nieaktywny (lista oczekujących) lub aktywny (farmakoterapia, psychoterapia wspierająca)
O <i>Outcome</i>	Oceniane efekty interwencji; np. te związane z nasileniem objawów klinicznych (np. depresja, lęk), fizjologią (np. aktywność mózgu), funkcjonowaniem (np. poziom codziennej aktywności) czy wskaźnikami społecznymi i epidemiologicznymi (np. koszty dla służby zdrowia)
T <i>Timeline</i>	Długość trwania leczenia albo okres, w którym mierzone są efekty oddziaływania; np. terapia trwająca 3 miesiące czy trwająca rok, pomiar efektów od razu po zakończeniu terapii czy po kilku miesiącach od jej zakończenia
S <i>Setting</i>	Setting psychoterapii, np. terapia indywidualna czy grupowa, terapia ambulatoryjna czy na oddziale psychiatrycznym

Jakie badania są zbierane w metaanalizach?

W metaanalizach mogą być syntetyzowane wyniki pochodzące z badań o różnych metodach, np. można przeprowadzić metaanalizę badań korelacyjnych. W badaniach nad efektywnością psychoterapii mogą to być metaanalizy badań kohortowych albo mierzących zmianę między początkiem i końcem terapii (*pretest-posttest design*), jednak najczęściej będą to RCT. W tym schemacie badawczym losowo przydziela się uczestników do dwóch lub więcej warunków, które różnią się między sobą rodzajem oferowanego oddziaływania (lub brakiem oddziaływania w przypadku warunków kontrolnych polegających na oczekiwaniu na terapię lub monitorowaniu klinicznym). Randomizacja, element RCT odróżniający tę metodę od innych badań klinicznych, ma na celu zminimalizowanie różnic między grupami badawczymi. Daje to możliwość interpretacji uzyskanych w ten sposób wyników jako powstałych na skutek relacji przyczynowo-skutkowej. W tej koncepcji osoby losowo przydzielone do warunków różnią się tylko faktem przydziału, więc to on odpowiada za obserwowane między osobami różnice. Warto

tutaj nadmienić, że często stosowany w badaniach nad psychoterapią schemat pomiaru przed i po interwencji jest obciążony większym ryzykiem błędu systematycznego, gdyż nie pozwala na kontrolowanie efektów związanych z upływającym czasem czy byciem monitorowanym przez specjalistów, a w przypadku aktywnych warunków kontrolnych (o czym poniżej), efektów związanych z oczekiwaniami czy niespecyficznymi czynnikami leczącymi.

W RCT przynajmniej jedna grupa powinna być warunkiem eksperymentalnym, czyli grupą poddawaną psychoterapii, oraz przynajmniej jedna – warunkiem kontrolnym. Istnieją różne warunki kontrolne w RCT. Powszechnie spotykanymi są terapia jak zazwyczaj (*treatment as usual*, TAU), oczekiwanie na terapię (*waiting list*) i terapia placebo. TAU jest trudna do dokładnego zdefiniowania, gdyż potrafi znacząco różnić się między badaniami (por. Watts i in., 2015). Oczekiwanie na terapię jako warunek kontrolny polega na poinformowaniu pacjentów przydzielonych do tej grupy o konieczności oczekiwania na leczenie. Pacjenci z tej grupy są badani na początku i końcu okresu oczekiwania, co stanowi punkt odniesienia dla osób, które w tym czasie przechodzą psychoterapię. Terapia placebo oznacza leczenie pozbawione „aktywnych” elementów, które według założeń terapii miałyby wpływać na samopoczucie i stan zdrowia pacjenta. Jej efekty są związane jedynie z oczekiwaniami i nastawieniem pacjenta. Stosowanie tego określenia w badaniach nad psychoterapią, w których nie da się pominąć niespecyficzných efektów związanych z kontaktem ze specjalistą (Kirsch, 2005) i w których różne warunki mogą wywoływać zróżnicowane efekty oczekiwań (Boot i in., 2013), jest krytykowane. Kolejnym rodzajem warunku kontrolnego jest zastosowanie innej uznanej albo zbliżonej efektywnością metody terapeutycznej (*active control*), co pozwala na bezpośrednie porównywanie ich rezultatów.

Szacunki dotyczące efektywności psychoterapii zależą od rodzaju użytego warunku kontrolnego. Na ogół, efektywność psychoterapii w porównaniu do oczekiwania na terapię jest większa niż efektywność w porównaniu do aktywnych warunków kontrolnych (Kowalski i in., 2024; Michopoulos i in., 2021). Dlatego informacja o warunku, do którego porównywana jest psychoterapia jest kluczowa dla zrozumienia syntezy wyników badań.

Warto tutaj nadmienić, że bezpośrednie porównania efektów uzyskanych w różnych badaniach jest obciążone poważnymi ograniczeniami i nie należy wprost wnioskować o tym, że jeśli terapia A jest skuteczniejsza od B ($A > B$), a B skuteczniejsza od C ($B > C$), to automatycznie $A > C$ (tzw. tranzytywizm; Baker i Kramer, 2002). Zaś badania nie mniejszej skuteczności (*non-inferiority*) oraz wyższości (*superiority*) interwencji wymagają wzięcia pod uwagę dodatkowych czynników, nie zaś samego bezpośredniego porównania (Leon, 2011). Istnieją metody syntezy danych, które pozwalają szacować niebezpośrednie wielkości efektów, np. metaanalizy sieciowe, ale ze względu na ograniczoną przestrzeń ten wątek nie będzie w tej pracy rozwijany (zainteresowane osoby mogą sięgnąć np. po: Caldwell i in., 2005; Hutton i in., 2015; Salanti, 2012). Niemniej jednak, wnioskowanie o większej skuteczności na podstawie zestawienia dwóch wielkości efektów uzyskanych z syntezy różnych badań jest po prostu błędne. Przykładem takiego błędu może być zestawienie efektu małego RCT psylocybiny z metaanalizami badań nad psychoterapią i farmakoterapią (Davis i in., 2021), gdzie

dodatkowym poziomem błędu, poza naiwnym założeniem o tranzytywizmie, jest zestawienie efektu małego RCT z efektami metaanaliz, które są uśrednionym szacunkiem, ale o wiele węższym przedziale ufności.

W wielu RCT jest stosowany także pomiar trwałości rezultatów w dłuższym okresie po zakończeniu terapii (*follow-up*). Zazwyczaj jest to przedział od 3 do 12 miesięcy, ale można napotkać pomiary *follow-up* po 2 czy 5 latach od zakończenia terapii. W metaanalizach trwałości efektów psychoterapii zazwyczaj wspólnie analizuje się wyniki z określonych przedziałów czasowych, np. do 6 miesięcy, 6–12 miesięcy i powyżej roku.

W metaanalizach podsumowuje się również wyniki badań, w których nie stosowano randomizacji, a co za tym idzie – możliwość wnioskowania przyczynowego na ich podstawie jest ograniczona (np. Harrer i in., 2023). Przykładem nie-losowego przydziału do grup może być analiza profilu objawów pacjentów i przydzielanie ich do interwencji, której metoda jest „bardziej dopasowana” do tego profilu (np. Levi i in., 2016). Innym typem badania bez randomizacji są badania kohortowe z wielokrotnym pomiarem, w których nie przydziela się uczestników do różnych warunków lub następuje to bez ingerencji badacza.

Prerejestracja

W ostatnich latach prowadzona jest dyskusja o rzetelności i stabilności efektów (lub: replikowalności, np. Nosek i in., 2022) obserwowanych w badaniach psychologicznych (Open Science Collaboration, 2015) i nauce ogólnie (Ioannidis, 2005). Jedną z ważnych praktyk, która ma pomóc uzyskiwać solidne wyniki badań, jest prerejestracja. Jej główne założenie to opisanie metody badania (hipotez, narzędzi, procedur, itd.) przed jego rozpoczęciem z taką dokładnością, aby przenieść nacisk z eksploracji na testowanie hipotez (Nosek i in., 2018). Celem jest próba uniknięcia np. stawiania hipotez po tym, jak znane są wyniki, efektu pewności wstecznej czy po prostu manipulowania parametrami analiz statystycznych, aby za cenę rzetelności dowieść istnienia postulowanych efektów (np. Chan i in., 2004).

W przypadku metaanaliz, znaczenie prerejestracji jest dwójakie. Po pierwsze, na poziomie indywidualnych badań prerejestracja oznacza, że takie badanie możemy uznać za obarczone mniejszym ryzykiem tendencyjności związanym np. z selektywnym raportowaniem. Możemy też odnaleźć w repozytoriach rekordy wskazujące na to, że takie badanie było przygotowywane. Co za tym idzie – możemy uwzględnić jego (nie)obecność w metaanalizie, zmniejszając efekt szuflady (o którym poniżej). Po drugie, na poziomie samej metaanalizy, prerejestracja pozwala na dokładniejsze opracowanie pytania badawczego oraz uniknięcie tendencyjności związanych z podejmowaniem decyzji analitycznych (np. które badania wykluczyć), gdy znane są już wyniki. Dodatkowo, pozwala uniknąć dublowania wysiłków badaczy, gdyż proces przygotowania przeglądu systematycznego i metaanalizy potrafi być długi, a obecność prerejestracji stanowi wyraźny sygnał dla innych badaczy, że praca danego rodzaju jest w przygotowaniu.

Mimo że w badaniach nad interwencjami psychologicznymi i behawioralnymi (Riehm i in., 2015) oraz w metaanalizach z zakresu psychologii (Sandoval-

Lentisco i in., 2025) prerejestracja jest dość niedawną i jeszcze nie powszechnie stosowaną praktyką, to została uznana za niezbędną w standardach tworzenia przeglądów i metaanaliz oddziaływań zdrowotnych.

Standardy tworzenia i raportowania przeglądów

Istnieje wiele schematów dotyczących tworzenia i raportowania metaanaliz. Tutaj wspomniane zostaną te, które są ogólnodostępne i uznawane za standard. Celem istnienia tego rodzaju standardów jest dbałość o jakość przygotowywanych metaanaliz oraz klarowność prezentowania ich wyników. Ma to skutkować wysoką jakością oraz zmniejszonym ryzykiem tendencyjności.

W tworzeniu przeglądów zdecydowanie wiodącą rolę ma podręcznik inicjatywy Cochrane (Higgins i in., 2024), który w wyczerpujący sposób opisuje proces przygotowywania syntezy badań nad interwencjami zdrowotnymi, w tym psychoterapią. W przypadku standardów raportowania wyników przeglądów najczęściej wykorzystywane są PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*, Liberati i in., 2009). PRISMA to wyczerpująca lista elementów, które powinny zostać uwzględnione w opisie metaanalizy od tytułu, przez strategię wyszukiwania, ekstrakcję danych, ocenę ryzyka tendencyjności, aż po wnioski oraz informacje o źródłach finansowania badania. Przykładowo, standardy PRISMA określają, że praca, która jest przeglądem systematycznym albo metaanalizą, powinna mieć to określone już w tytule, co skutkuje tym, że łatwiej znaleźć takie badania przez zawężenie pól wyszukiwania w bazie danych.

Jak interpretować wyniki metaanaliz?

Wskaźniki efektywności, ocena ryzyka tendencyjności, wizualizacja danych

Wskaźniki efektywności interwencji

Najczęściej używanymi wskaźnikami efektywności w metaanalizach są standaryzowane różnice średnich (ang. *standardized mean differences*, SMD), liczba potrzebnych interwencji (ang. *number needed to treat*, NNT), szacunki ilorazu szans (ang. *odds ratio*, OR) oraz proporcja osób, u których doszło do remisji lub znaczącej poprawy.

Ten ostatni najczęściej jest wyliczany na trzy sposoby. Każdy z nich ma swoje ograniczenia, o których warto wspomnieć. Pierwszym jest wskaźnik rzetelnej zmiany (ang. *reliable change index*, RCI; Jacobson i Truax, 1991), w którym najważniejszymi parametrami są wariancja wyników narzędzia oraz jego stabilność czasowa. Jego założeniem jest wyliczenie punktu odcięcia, dla którego istnieje znacząca pewność, że obserwowaną zmianę można przypisać efektowi oddziaływania, a nie błędowi pomiaru danego narzędzia. Istnieje wiele punktów krytyki dotyczącej tego wskaźnika (McAleavey, 2024), do których należą np. trudność wykazania istotnej zmiany (tj. konserwatywność), odnoszenie grupowej wariancji do wyników pojedynczego pacjenta czy utrata informacji przez dychotomizację

ciągłej zmiennej. Drugim sposobem jest ocena poprawy na podstawie punktu odcięcia ustalonego dla danego narzędzia. W tej metodzie zakłada się, że osoby, których wynik wypadła poniżej określonego punktu odcięcia, mają nie-kliniczne nasilenie danego objawu. Dzieli ono niektóre ograniczenia z RCI, np. dychotomizację ciągłej zmiennej. Bardziej specyficznym ograniczeniem jest to, że nie bierze ono pod uwagę rzetelności narzędzia, co może prowadzić do sytuacji, w których minimalna zmiana w punktacji, np. z jednego punktu powyżej progu odcięcia do jednego punktu poniżej, prowadzi do wniosku o zaistnieniu znaczącej poprawy funkcjonowania (np. Andersson i in., 2025). Ostatnim sposobem jest ocena remisji na podstawie niespełniania kryteriów diagnostycznych po zakończeniu terapii. Ta metoda, podobnie jak wcześniejsza, może skutkować oceną, w której przejście z relatywnie niskiego nasilenia objawów do „subklinicznego” poziomu jest traktowane jako znacząca poprawa, mimo braku odzwierciedlenia w rzeczywistym poziomie funkcjonowania pacjenta. Warto również wspomnieć o wskaźnikach z „rodziny” *Minimal Important Difference*, które bywają wykorzystywane w badaniach psychoterapeutycznych (np. Copay i in., 2007).

Standaryzowane różnice średnich (SMD)

SMD to wskaźniki, w których wylicza się różnicę między dwoma pomiarami (np. między dwoma grupami albo przed i po terapii wewnątrz jednej grupy) i dzieli ją przez zsumowane odchylenie standardowe (Lakens, 2013). Najpopularniejszą miarą wielkości efektu stosowaną w metaanalizach jest *g* Hedgesa (która jest zmodyfikowanym *d* Cohena z poprawką na mały rozmiar próby, więcej w: Hedges, 2025).

SMD w badaniach interwencji najczęściej zawierają się w przedziale od -1 do 0^4 , co jest zależne np. od użytego warunku kontrolnego czy rodzaju zaburzenia, którego dotyczy psychoterapia (np. Cuijpers i in., 2025), ale również mogą wykraczać poza ten zakres, gdyż są bezwymiarowe (Hedges, 2025). Tradycyjnie przyjmuje się (za Cohenem), że wielkości efektów z różnych przedziałów interpretuje się jako małe ($d = -0,2$), średnie ($d = -0,5$) i duże ($d = -0,8$), jednak jest to podział arbitralny i nie ma uzasadnienia, żeby trzymać się tych interpretacji w sposób ścisły (Lakens, 2013).

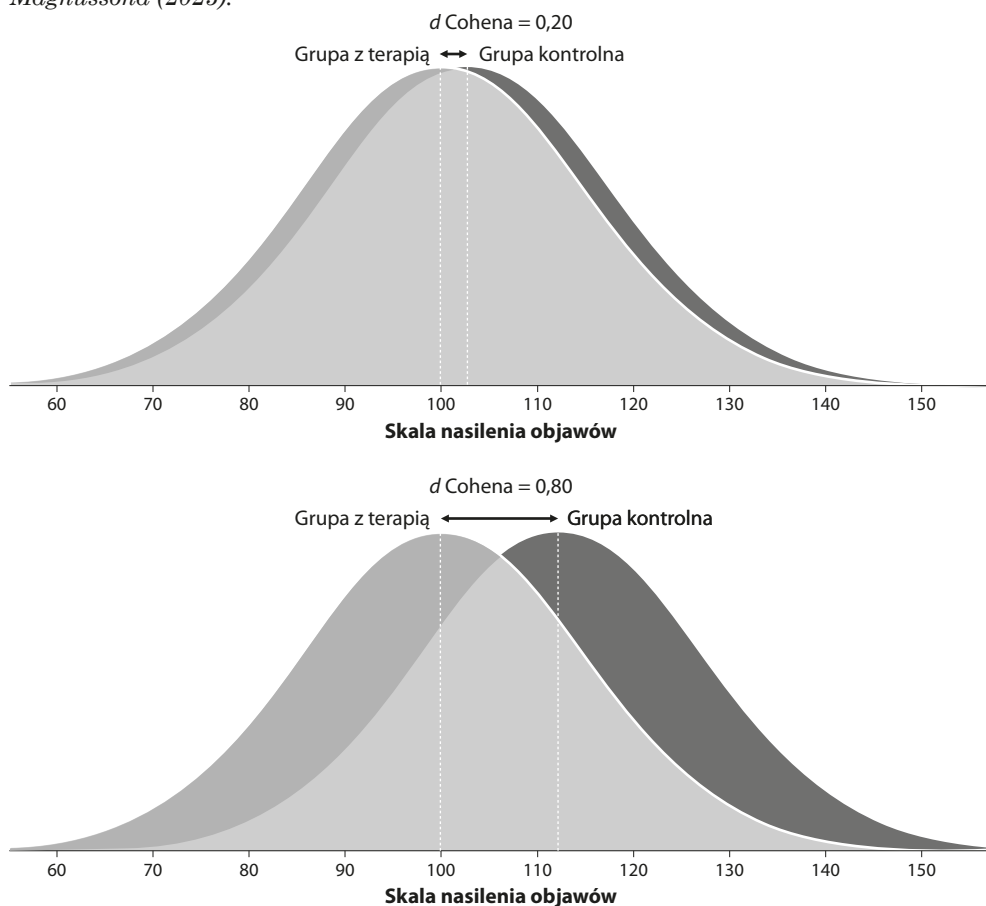
Bardziej intuicyjnym sposobem rozumienia wskaźników SMD może być analiza zachodzenia na siebie rozkładów wyników dwóch grup, np. biorącej udział w terapii oraz kontrolnej. Na rysunku 1 (s. 38) przedstawione zostały dwa rozkłady zmiennych o wielkości efektu $d = 0,80$ oraz $d = 0,20$. Dla „dużego” efektu rozkłady zachodzą na siebie w około 69%, a około 79% osób biorących udział w terapii ma wyniki powyżej średniego wyniku grupy kontrolnej. Dla efektu „małego” rozkłady nachodzą na siebie w 92%, a około 58% grupy z terapią ma wyniki powyżej średniego wyniku grupy kontrolnej. Należy jednak zaznaczyć, że

⁴ lub od 0 do 1, w zależności od tego, która z dwóch grup jest grupą odniesienia (eksperymentalna czy kontrolna). Oznacza to, że interpretując wielkości efektów, należy zwracać uwagę na przedstawiony przy nich opis, a nie tylko znak, żeby poprawnie interpretować kierunek zależności między grupami.

ten sposób interpretacji jest możliwy jedynie przy spełnieniu założeń o normalności rozkładu i równości wariancji grup (Hedges, 2025). Czytelnicy mogą na własną rękę sprawdzać, w jaki sposób można interpretować d Cohena za pomocą tej aplikacji: <https://rpsychologist.com/cohend/> (Magnusson, 2025).

Rysunek 1

Wizualizacja dwóch wielkości efektów oddziaływania terapeutycznego, d Cohena = 0,2 (efekt mały) oraz 0,8 (efekt duży) przy założeniu, że wyniki obu grup mają rozkład normalny, ich rozkłady były identyczne przed rozpoczęciem terapii, a łączone odchylenie standardowe wynosi $SD = 15$. Przygotowane za pomocą programu opracowanego przez Magnussona (2025).



Liczba potrzebnych interwencji (NNT)

Kolejnym wskaźnikiem skuteczności jest szacunkowa liczba osób, która powinna zostać poddana psychoterapii, w porównaniu z interwencją kontrolną, aby osiągnąć zamierzony efekt terapeutyczny u jednej osoby. Tym efektem może

być uniknięcie jakiegos niekorzystnego zdarzenia (np. próby samobójczej) lub uzyskanie zadowalającej poprawy lub remisji (Hodgson i in., 2011).

NNT oblicza się przez iloraz jedności oraz różnicy stosunków efektów (poprawa vs brak poprawy, itp.) w grupie kontrolnej i grupie eksperymentalnej, według wzoru $NNT = 1 / (\text{stosunek efektów w grupie kontrolnej} - \text{stosunek efektów w grupie eksperymentalnej})$. Czyli w sytuacji, w której znaczącą poprawę obserwujemy u wszystkich osób z grupy eksperymentalnej, a u żadnej osoby z grupy kontrolnej, NNT wynosi 1. Jeśli poprawę obserwujemy u połowy osób z grupy eksperymentalnej, a u żadnej osoby z grupy kontrolnej, to $NNT = 2$. Jeśli poprawę obserwujemy u połowy grupy badanej oraz u 25% grupy kontrolnej, to $NNT = 4$.

Niestety, nie ma sposobu na dokładne oszacowanie NNT na podstawie wielkości efektu bez znajomości opisanych tu stosunków (por. Furukawa i Leucht, 2011), ale Furukawa (1999) zaproponował metodę szacowania NNT na podstawie d Cohena, co ma znaczenie dla praktyków, dla których NNT może być bardziej intuicyjnym wskaźnikiem efektywności terapii niż standaryzowana różnica średnich. Kalkulator bazujący na tej metodzie jest uwzględniony w przywołanej w poprzednim podrozdziale aplikacji.

Komplementarnym wskaźnikiem do NNT jest *Number Needed to Harm*, czyli wyliczany w ten sam sposób iloraz, ale zdarzeń niepożądanych. Tego rodzaju wskaźnik pozwala na ocenę profilu bezpieczeństwa interwencji psychoterapeutycznej.

Ryzyko względne i iloraz szans

Kolejne wskaźniki, które są wykorzystywane do przedstawiania wyników metaanaliz, to ryzyko względne (*relative risk*, RR) i iloraz szans (*odds ratio*, OR). Ich szczególnym zastosowaniem jest przedstawianie dychotomicznych efektów terapii, np. osiągnięcie remisji vs brak remisji, jako stosunku prawdopodobieństw lub szans⁵. Ryzyko względne jest stosunkiem prawdopodobieństw wystąpienia zdarzeń w dwóch grupach w odniesieniu do możliwych do zaobserwowania efektów, a iloraz szans jest stosunkiem szans na wystąpienie zdarzenia do szans jego niewystąpienia w dwóch grupach (np. Andrade, 2015; Schmidt i Kohlmann, 2008). W sytuacji, w której a oznacza osiągnięcie efektu terapeutycznego w grupie eksperymentalnej, b brak tego efektu w grupie eksperymentalnej, c wystąpienie efektu w grupie kontrolnej, a d brak efektu w grupie kontrolnej, to ryzyko względne wylicza się według wzoru $RR = (a / [a+b]) / (c / [c+d])$, a iloraz szans jako $OR = (a/b) / (c/d)$.

Częściej spotykanym estymatem w badaniach nad psychoterapią jest OR, ale ze względu na zdecydowaną różnicę między tymi dwoma wskaźnikami, gdy frekwencja obserwowanych zdarzeń jest wysoka (Schmidt i Kohlmann, 2008)

⁵ Intuicyjnym przykładem ułatwiającym zrozumienie różnicy między pojęciami ryzyka i szansy jest kostka do gry. Ryzyko jest wartością prawdopodobieństwa, że dane zdarzenie będzie miało miejsce, czyli ryzyko wyrzucenia 6 na kostce do gry to $1/6 \approx 0.17$, zaś szansa to stosunek prawdopodobieństw wystąpienia i nie wystąpienia zdarzenia, czyli szanse wyrzucenia 6 na kostce do gry to 1 do 5, czyli 0.20 (Andrade, 2015).

oraz mniej intuicyjną interpretację OR, warto unaocznić różnicę między tymi dwoma wskaźnikami na przykładach.

- Przykład 1: 80% pacjentów w grupie z terapią osiąga remisję, zaś w grupie kontrolnej 50%. W tym wypadku ryzyko względne wynosi $RR = 1,6$, co można interpretować jako 60% większe ryzyko na osiągnięcie remisji dla osób uczestniczących w terapii w porównaniu z grupą kontrolną, zaś $OR = 4$, co można interpretować jako cztery razy większe szanse na osiągnięcie remisji wśród osób uczestniczących w terapii (szansa w tej grupie wynosi 4 do 1) w stosunku do szans osób z grupy kontrolnej (szansa 1 do 1).
- Przykład 2: Jeśli 80% osób w grupie z terapią osiąga remisję, a tylko 20% w grupie kontrolnej, to $RR = 4$, co oznacza cztery razy większe ryzyko remisji u osób w grupie z terapią w porównaniu z grupą kontrolną, a $OR = 16$, co oznacza 16 razy większą szansę osiągnięcia remisji u osób w grupie z terapią (szansa wynosi 4 do 1) w stosunku do szans osób w grupie kontrolnej (szansa 1 do 4).

Przedziały ufności oraz przedziały predykcji

Dla uśrednionych wskaźników jak SMD czy OR w metaanalizach przedstawia się również przedziały ufności (ang. *confidence intervals*, CI) oraz przedziały predykcji (ang. *prediction intervals*, PI), najczęściej dla poziomu 95%. Ogólnie, im węższy przedział tym bardziej precyzyjnie jesteśmy w stanie określić szacowany efekt.

Przedział ufności oznacza – na najprostszym przykładzie – że przy 100 takich samych badaniach z wyliczonymi CI, 95 z tych przedziałów zawiera prawdziwą wartość efektu w populacji (co oznacza, że istnieje ryzyko, że przedział wyliczony w „naszym” badaniu jest jednym z tych 5%, które nie zawierają parametru populacji). CI jest więc interpretowany jako zakres, w którym zawiera się „prawdziwa” wielkość efektu w badanej populacji, szacunkowa *pewność* co do efektu. CI, które w przypadku SMD stykają się z 0 albo w przypadku OR z 1 są interpretowane jako nieistotne statystycznie. Interpretacja ta bazuje na możliwości, że faktyczny efekt w populacji jest zerowy.

Przedział predykcji oznacza – ponownie na najprostszym przykładzie – wartości wielkości efektów, które możemy uzyskać w 95 na 100 badań podobnych do tych uwzględnionych w metaanalizie, które moglibyśmy przeprowadzić w przyszłości (IntHout i in., 2016). PI jest więc interpretowany jako zakres, w którym z dużym prawdopodobieństwem będzie zawierać się wielkość efektu w kolejnym badaniu z populacji badań ujętej w metaanalizie, czyli szacunkowa *zmiennność* efektu między badaniami. PI, które w przypadku SMD stykają się z 0 albo w przypadku OR z 1 są interpretowane jako wskaźnik wysokiej zmienności analizowanych wyników.

Przykładowo, wynik metaanalizy, w którym 95% CI dla SMD = 0,60 wynosi 0,20 do 1,00; ale 95% PI wynosi –0,10 do 1,30 oznacza, że efekt interwencji jest statystycznie istotny, ale istnieje duża zmienność między poszczególnymi badaniami, która może skutkować tym, że kolejne badanie „w serii” może mieć zerowy bądź odwrotny efekt niż średni efekt dotychczasowych badań.

Analiza w podgrupach oraz metaregresja

Analiza wyników metaanalizy w podziale na podgrupy pozwala na określenie, czy jakaś charakterystyka badań moduluje uzyskane w badaniu efekty (Deeks i in., 2023). Przykładowo, w metaanalizie która agreguje wiele różnych badań można spodziewać się zróżnicowanych efektów terapeutycznych w zależności od tego, czy efekt ten jest wyliczany na podstawie badań o wysokim lub niskim ryzyku tendencyjności, badań terapii indywidualnej lub grupowej, badań z udziałem pacjentów z diagnozą depresji lub zaburzeń lękowych, a nawet na jakim kontynencie badanie było przeprowadzane (wszystkie te analizy oraz inne można znaleźć np. w pracy Anderssona i in., 2025).

Istotne jest, żeby rodzaj i liczba analizowanych podgrup były ograniczone, uzasadnione teoretycznie oraz określone przed przeprowadzeniem analiz (a najlepiej prerejestrowane), a liczba uwzględnionych w metaanalizie badań wystarczająca do tego typu analizy (Deeks i in., 2023). Przykładem obrazującym nieroztropne analizowanie podgrup może być pozornie humorystyczny przykład z badań medycznych, gdzie dokonano podziału pacjentów według znaku zodiaku i wykazano, że położenie słońca na tle układów gwiazd w momencie urodzenia jest związane ze zmniejszonym działaniem ochronnym aspiryny w prewencji zawału serca (Sleight, 2000).

Metaregresja jest rozwinięciem idei analizy w podgrupach (Deeks i in., 2023), tyle że zamiast jednej zmiennej kategoryjnej można analizować wiele zmiennych kategoryjnych oraz ciągłych na raz (odrębną kwestią jest czy dysponuje się odpowiednią liczbą badań do przeprowadzenia takiej analizy). Analogicznie do regresji liniowej, charakterystyki badań służą za zmienne przewidujące, a zmienną przewidywaną jest wskaźnik efektywności (SMD albo OR).

Przeprowadzenie metaregresji pozwala zaadresować problem dużego zróżnicowania badań uwzględnionych w metaanalizie (więcej o tym poniżej), dzięki możliwości zidentyfikowania jaka charakterystyka badania jest powiązana z obserwowaną heterogenicznością oraz może modulować obserwowane efekty (Baker i in., 2009). Za przykład może posłużyć metaanaliza 15 podejść do psychoterapii depresji, w której szacunkowa efektywność różnych metod różniła się w zależności od tego, jakie było ryzyko tendencyjności analizowanych badań, rodzaj grupy kontrolnej albo czy badanie było przeprowadzone w kraju globalnego Zachodu czy też nie (Cuijpers i in., 2020).

Jak oceniamy jakość metaanaliz?

Analizy ryzyka tendencyjności i heterogeniczności

Badania psychoterapii potrafią się między sobą znacząco różnić, między innymi ze względu na badane populacje, metody terapii, długość oddziaływania, setting, w którym terapia jest prowadzona, itd., ale też jakość zastosowanych metod badawczych. W związku z tym, istotnym elementem metaanalizy jest również analiza, w jaki sposób to zróżnicowanie może wpływać na interpretację uzyskanych wyników.

Tendencyjność (*bias*) w badaniach klinicznych i ich metaanalizach oznacza występowanie systematycznego błędu, który może zakłócać uzyskane wyniki.

Przy czym należy odróżnić tendencyjność od niesystematycznego błędu pomiaru. Przykładem takiego błędu może być np. zmienność obserwowana w pomiarach kwestionariuszowych, w których istotę wpisana jest niedokładność. Tendencyjność zaś to rodzaj systematycznego błędu, gdzie estymowane wielkości efektu będą odbiegały w podobny, systematyczny sposób od realnego efektu. Przykładem takiej tendencyjności może być np. efekt oczekiwań badaczy, który – jeśli faktycznie zniekształca wyniki – będzie skutkował „podbijaniem” efektywności danej interwencji w pozornie niezależnych od siebie badaniach.

Ocena ryzyka tendencyjności badań w metaanalizie

W przypadku oceny badań mówimy o ryzyku tendencyjności (*risk of bias*, Boutron i in., 2023) ze względu na to, że nie mamy informacji o tym, czy dane źródło tendencyjności faktycznie mogło ujawnić się w danym badaniu. Na przykład brak zaślepienia badaczy nie oznacza automatycznie, że wyniki są tendencyjne, ale ryzyko tendencyjności jest podwyższone w porównaniu do analogicznego badania, w którym zostali oni zaślepieni. W celu lepszego zobrazowania – nowa metaanaliza badań nad psychoterapią depresji nie wykazała istotnych różnic między oceną nasilenia objawów przez osoby uczestniczące w psychoterapii, a taką oceną wykonaną przez zaślepionych klinicystów (Miguel i in., 2025). Może to oznaczać, że – niezgodnie z intuicją – oceny dokonywane przez pacjentów nie są obciążone większą tendencyjnością (związaną np. z efektem oczekiwań) niż oceny klinicystów.

Najczęściej wykorzystywanym narzędziem do oceny ryzyka tendencyjności randomizowanych badań uwzględnionych w metaanalizie jest *the Cochrane tool for assessing risk of bias in randomized trials*, obecnie w swojej aktualizowanej wersji, czyli RoB 2 (Sterne i in., 2019), a w przypadku badań nierandomizowanych ROBINS-I (Sterne i in., 2016). Za pomocą tych narzędzi ocenia się np. ryzyko tendencyjności wynikające z: nieadekwatnego przeprowadzenia losowego przydziału do grup (czy nieadekwatnego opisu takiej procedury), braku zaślepienia, wybiórczego prezentowania danych dotyczących uczestników badania albo narzędzi pomiarowych (Higgins i in., 2023; Sterne i in., 2023). Warto również brać pod uwagę ryzyko tendencyjności związane z konfliktem interesów, efektem przynależności badaczy (*allegiance bias*; Cuijpers i Cristea, 2016; Munder i in., 2013), albo źródłami finansowania badań. W ostatnim czasie te źródła tendencyjności były szczególnie dyskutowane w kontekście badań nad efektywnością psychodelików (Buchman i Rosenbaum, 2024; Lemarchand i in., 2024). Badacze starają się również wziąć poprawkę na to, które z badań nie zostały udostępnione szerszej publiczności (*publication bias* lub efekt szuflady). Finałnie, można ocenić, jak skonstruowana została sama metaanaliza (Boutron i in., 2023), gdyż decyzje, podejmowane przez jej autorów również mogą skutkować większym ryzykiem tendencyjności.

Ocena i analiza ryzyka tendencyjności samej metaanalizy

Istnieje kilka narzędzi pozwalających oceniać przeglądy systematyczne i metaanalizy pod względem ryzyka tendencyjności (np. ROBIS, Whiting i in., 2016,

system zaproponowany przez JBI, Aromataris i in., 2015, czy narzędzie ROB-ME do oceny tendencyjności wynikającej z brakujących badań, Page i in., 2023), ale zdecydowanie najpopularniejszym narzędziem jest AMSTAR (Shea i in., 2007a), obecnie jego druga wersja (Shea i in., 2017), czyli *A Measurement Tool to Assess systematic Reviews*. AMSTAR-2 pozwala na ocenę zarówno metaanaliz RCT, jak i badań nierandomizowanych. To narzędzie bierze pod uwagę wiele aspektów przeglądów i metaanaliz, od sformułowania pytania badawczego i prerejestracji, przez wybór badań, ocenę tendencyjności tych badań po ocenę potencjalnego konfliktu interesów samych autorów danego przeglądu. Badania dotyczące pierwszej wersji AMSTAR wykazały zadowalającą zgodność oceniających, spójność wewnętrzną i łatwość dokonywania oceny (Shea i in., 2007b, 2009).

Skwantyfikowanie ryzyka tendencyjności poszczególnych badań jest trudne, dlatego ocenia się je przede wszystkim jakościowo albo wprowadza dodatkowe analizy (tzw. *sensitivity analyses*, czyli analizy pokazujące, na ile uzyskany wynik jest wrażliwy na zmianę parametrów), np. opisane powyżej analizy w podgrupach. W przypadku źródeł tendencyjności związanych z brakującymi publikacjami (w tym *publication bias*, Sedgwick, 2015) stosuje metody takie jak analiza wykresu lejkowego (Egger i in., 1997, ale też krytyka w Lau i in., 2006), test Eggera (Egger i in., 1997), analizę *trim-and-fill*, czyli niejako poprawkę na brakujące w analizie badania (Duval i Tweedie, 2000), czy też oszacowanie *fail-safe N* (Rosenthal, 1979), czyli liczby badań o zerowym efekcie, która sprawiłaby, że obserwowany w metaanalizie efekt będzie nieistotny (można myśleć o tym jako o „teście na rozmiar szuflady”).

Ocena i analiza heterogeniczności w metaanalizie

Statystyczna analiza heterogeniczności bazuje na testach sprawdzających, do jakiego stopnia różnice pomiędzy poszczególnymi badaniami wynikają z innych czynników niż losowy, niesystematyczny błąd (Deeks i in., 2023). Zazwyczaj w tego rodzaju analizach przedstawia się test Q Cochrańa albo obliczany na podstawie Q i liczby badań test I^2 (Higgins i in., 2003). Wynik testu I^2 to szacunkowa część zmienności uzyskanych wyników, która jest spowodowana heterogenicznością, a nie losowym błędem, i jest wyrażona w procentach (Deeks i in., 2023). Mimo że raczej rzadko raportowany w ten sposób test I^2 powinien być interpretowany razem z np. przedziałem ufności, gdyż bardzo szeroki przedział może sprawiać, że sensowna interpretacja testu nie jest możliwa (Borenstein i in., 2017). Przykładowo, $I^2 = 15\%$ może być interpretowane jako niska heterogeniczność, ale 95% CI dla tego estymatu I^2 od 0% do 90% oznacza, że faktyczna heterogeniczność zawiera się w przedziale od bardzo niskiej do bardzo wysokiej. Innym wskaźnikiem, może nawet bardziej intuicyjnym, jest opisany powyżej przedział predykcji (IntHout i in., 2016).

Heterogeniczność w badaniach psychoterapii (i oddziaływań zdrowotnych ogólnie) jest nie do uniknięcia (Higgins i Thompson, 2002). Ale wynika to również z tego, jak zaprojektowana jest metaanaliza, np. zadania szerokiego pytania badawczego czy włączenia do analizy zróżnicowanych pomiarów efektywności. Bazując na przykładzie – metaanaliza opisująca małą liczbę badań zróżnicowanych metod psychoterapii (np. w różnych settingach i różnych podejściach terapeutycznych)

na temat zróżnicowanego problemu klinicznego (co w zasadzie odnosi się do większości diagnoz klinicznych) da „ogólną” odpowiedź o skuteczności tych psychoterapii, ale jednocześnie zróżnicowanie tych badań sprawi, że pewność co do uzyskanych wyników będzie zdecydowanie ograniczona. W związku z zaobserwowaniem wysokiej heterogeniczności badacze powinni podjąć decyzję o zmianie strategii analitycznej, np. w ogóle nie przeprowadzać metaanalizy, zidentyfikować i wykluczyć badania będące źródłem heterogeniczności (obserwacje odstające, *outliers*), zidentyfikować zróżnicowane efekty za pomocą metaregresji, czy też przeprowadzić metaanalizę bazującą na modelu z efektem losowym. Więcej informacji na ten temat można znaleźć w odpowiednim rozdziale podręcznika grupy Cochrane (Deeks i in., 2023).

Jak to wszystko czytać, żeby zrozumieć? Wizualizacja wyników i praktyczny przykład

Wykres leśny

Wykres leśny jest graficzną reprezentacją wyników metaanalizy (rysunki 2a i 2c, s. 46). Przedstawione są na nim uwzględnione badania oraz ogólny wynik metaanalizy w formie kolejnych rzędów, gdzie powinny być umiejscowione informacje o tym: a) jakie badanie przedstawia dany punkt, b) przedstawienie jaki efekt został w nim uzyskany wraz z c) CI najczęściej 95%, d) zarówno w formie graficznej, jak i tekstowej. Kolejne badania są przedstawiane jako punkty z „wąsami” CI, zaś ogólny efekt metaanalizy najczęściej w formie rombu, którego wierzchołki wyznaczają granice przedziału ufności dla tej analizy. Dodatkowo na takim wykresie umieszczane są inne informacje, jak np. ocena ryzyka tendencyjności, wielkość próby, waga danego badania dla całego efektu metaanalizy (najczęściej reprezentowana jako wielkość punktu przedstawiającego efekt danego badania), analiza heterogeniczności (często, niestety, bez przedziału ufności), itd. Zaś na poziomej osi zaprezentowane są wartości wielkości efektu (SMD, OR, itd.) z oznaczeniem, który efekt faworyzuje grupę eksperymentalną, a który grupę kontrolną oraz oznaczonym punktem zerowym (Andrade, 2020; Sedgwick, 2015a). Wizualna interpretacja wykresu z przedziałami ufności pozwala również na ocenę istotności statystycznej, gdyż możemy stwierdzić, że dwa niezależne od siebie efekty różnią się istotnie statystycznie ($p \leq 0,05$), gdy ich CI zachodzą na siebie w mniejszej części niż połowa uśrednionego między nimi przedziału, zaś $p \leq 0,01$, gdy nie zachodzą na siebie wcale (Cumming i Finch, 2005)⁶.

Wykres lejkowy

Wykres lejkowy (ang. *funnel plot*, rysunki 2b i 2d, s. 46) przedstawia rozrzut punktowy, na którym punkty odpowiadają poszczególnym badaniom. Na osi X

⁶ Ten sposób interpretacji jest wystarczająco dokładny, gdy obie porównywane grupy są większe niż 10 obserwacji, zaś stosunek szerszego CI do węższego nie przewyższa 2 (Cumming i Finch, 2005).

prezentowane są wielkości efektu, a na osi Y najczęściej błąd standardowy (Sterne i Egger, 2001), zaś skośne linie odpowiadają granicom przedziałów ufności dla określonych wartości błędu standardowego. Wykres ten ma na celu ocenę, czy w przeprowadzanej metaanalizie obserwujemy zwiększone ryzyko tendencyjności wynikające z 1. *selection/reporting bias*, czyli różnych okolicznościach skutkujących nieuwzględnieniem badania w metaanalizie (Sedgwick, 2015b) oraz 2. przeszacowania wielkości efektów przez badania o małej próbie i/lub wysokim ryzyku tendencyjności (Sedgwick i Marston, 2015). Nie wnioskujemy o zwiększonym ryzyku tendencyjności, gdy punkty są rozrzucone symetrycznie względem osi wyznaczonej przez średnią wielkość efektu, a rozrzut zawęża się wraz z malejącym błędem standardowym, tj. badania na mniejszych próbach dają bardziej zróżnicowane efekty. Zaś ryzyko tendencyjności jest oceniane na podstawie asymetrii wykresu lejkowego, tj. im bardziej „skrzywiony” rozrzut, tym należy liczyć się z większym ryzykiem tendencyjności. Co więcej, istnieją metody statystyczne pozwalające formalnie przeanalizować ryzyko brakujących badań lub „uzupełnić” wykres lejkowy, np. wspomniana wyżej metoda *trim-and-fill*. Szczególnie, że ocena wykresów lejkowych „na oko” nie daje rzetelnych efektów (np. Simmonds, 2015). Dodatkowo, nie można wykluczyć, że obserwowana na wykresie asymetria nie jest efektem „brakujących” badań, a np. wynikiem heterogeniczności metaanalizowanych badań czy czynnika losowego (Afonso i in., 2024; Sterne i in., 2005).

Praktyczny przykład

Za przykład obrazujący wykresy leśny i lejkowy posłuży wizualizacja metaanalizy porównująca efektywność terapii poznawczo-behawioralnej (CBT) oraz psychodynamicznej (PDT) w leczeniu osób z fobią społeczną, która została opisana w pracy Kowalskiego (2024). Analiza została przeprowadzona za pomocą JASP v0.18.3 (JASP Team, 2024).

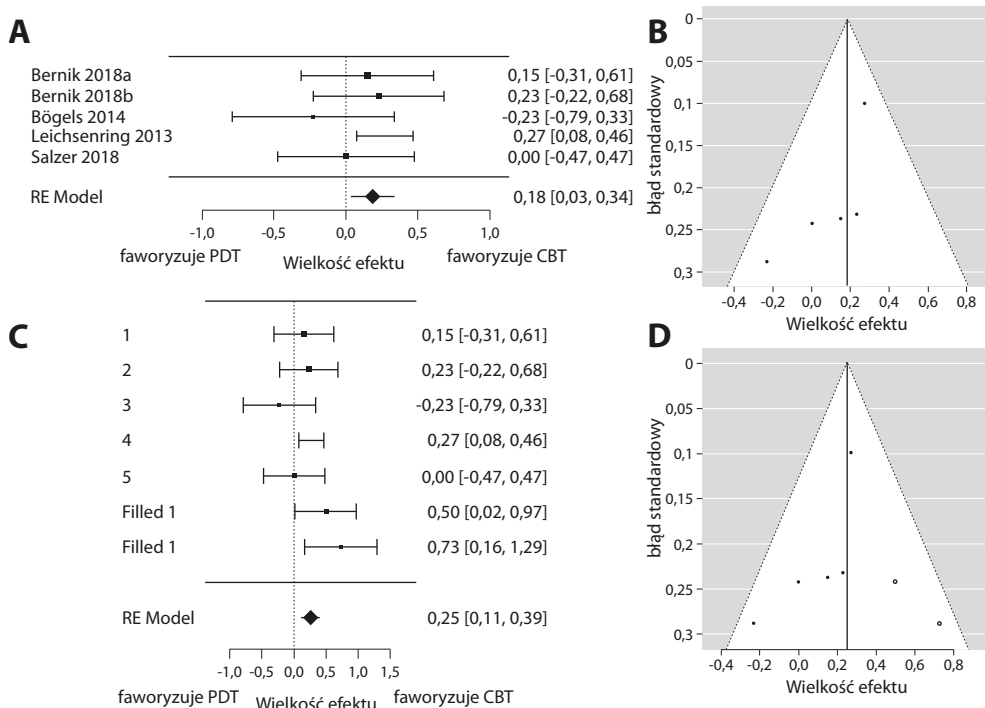
Na rysunku 2a (s. 46) widzimy podsumowanie wyników 5 badań, które porównywały ze sobą te dwie terapie. Wszystkie badania poza jednym wykazały nieistotne statystycznie różnice między dwoma terapiami (przedziały ufności przekraczają zero), a tylko jedno, ale o największej próbie, wykazało istotną różnicę na korzyść CBT. Ogólny wynik metaanalizy to mały efekt na korzyść CBT, $SMD = 0,18$; 95% CI 0,03–0,34, ale z przedziałem predykcji przekraczającym zero: 95% PI –0,06 do 0,42, i jednocześnie obserwujemy w tej metaanalizie $I^2 = 3\%$ (95% CI 0–87%). Czyli choć oszacowanie I^2 wynosi zaledwie 3%, bardzo szeroki przedział ufności (0–87%) oznacza, że nie możemy wykluczyć istotnej heterogeniczności, najpewniej wynikającej z małej liczby badań. Ponadto, test rangowanej korelacji na asymetrię wykresu lejkowego (rys. 2b, s. 46; Begg i Mazumdar, 1994) wykazał istotne ($\tau = -1,0$; $p = 0,017$) ryzyko asymetrii, wskazując, że mniejsze badania w tej analizie są skrzywione w stronę wykazywania zerowego albo faworyzującego PDT efektu (np. ze względu na efekt przynależności badaczy, Munder i in., 2013).

W związku z tym przeprowadzono analizę *trim-and-fill*, która uzupełniła metaanalizę o dwa hipotetyczne badania, gdzie CBT była bardziej skuteczna. Ta poprawka na asymetrię wykresu lejkowego zmieniła ogólny wynik metaanalizy, $SMD = 0,25$ (95% CI 0,11–0,39), co zostało szczegółowo przedstawione na

rysunkach 2c i 2d. Zaś sam wynik tej metaanalizy może być interpretowany jako wskazujący na wyższą efektywność CBT w leczeniu fobii społecznej w porównaniu do PDT, ale z małą wielkością efektu.

Rysunek 2

Wykresy leśny (A) i lejkowy (B) dla metaanalizy porównującej skuteczność psychoterapii poznawczo-behawioralnej i psychodynamicznej w leczeniu fobii społecznej. Oraz wykresy leśny (C) i lejkowy (D) po zastosowaniu procedury trim-and-fill.



Należy zwrócić uwagę, że poza podwyższonym ryzykiem tendencyjności wykazany w analizie wykresu lejkowego przedstawiona metaanaliza ma inne problemy – przede wszystkim małą liczbę uwzględnionych badań oraz, najprawdopodobniej jako skutek tego, wysoką heterogeniczność i szeroki przedział predykcji. Należy więc traktować ten przykład ilustracyjnie, jako zobrazowanie tego, jakie informacje można uzyskać, analizując wykresy leśny i lejkowy.

Czy ma to jakieś zastosowanie praktyczne?

Rola metaanaliz w praktyce klinicznej i tworzeniu zaleceń terapeutycznych

Prace syntetyzujące wyniki badań mają kluczowe znaczenie dla EBP oraz praktyki klinicznej w ogóle (np. Spring, 2007). Należy też wziąć pod uwagę zjawisko

luki między nauką a praktyką (np. Baker i in., 2008) czy opisane we wprowadzeniu zniekształcenia poznawcze, które przekładają się na to, że osoby praktykujące psychoterapię nie polegają w swojej pracy na dowodach naukowych w tak dużym stopniu jak np. na intuicji (Gyani i in., 2014). Dlatego poniżej przytoczone zostaną argumenty za korzyściami wynikającymi ze znajomości i stosowania w praktyce wyników badań.

Jednym z kluczowych argumentów może być osiąganie lepszych efektów terapeutycznych. Np. w badaniu IAPT (*Improving Access to Psychological Therapies*) kohortowe dane z brytyjskich centrów zdrowia psychicznego wykazały, że osoby z nasilonymi objawami lękowymi, które otrzymywały rekomendowaną w zaleceniach terapię, częściej osiągały poprawę niż osoby, które wzięły udział w terapii nie zalecanej w przypadku tego problemu (Gyani i in., 2013). Opieranie się w praktyce na syntezie danych na temat efektywności psychoterapii w danym problemie pozwalana trafniejsze dobieranie metod terapeutycznych do problemu danego pacjenta (np. Kowalski i in., 2023), dostosowanie długości, intensywności albo settingu tego leczenia. Pozwala też na „szycie terapii na miarę”, czyli dopasowania do specyficznych mechanizmów danego problemu u danego pacjenta (np. Nye i in., 2023), oraz monitorowanie efektywności tego szycia. Dodatkowo, korzystanie z dowodów w pracy klinicznej może mieć pozytywny efekt dla etycznej strony prowadzonej praktyki, gdyż udzielenie informacji i psychoedukacja na temat efektywności interwencji pozwala w większym stopniu na uzyskanie świadomej zgody (Blease i in., 2018).

Kolejnym argumentem jest synergia z wymogiem ustawicznego kształcenia. Sięganie po dowody naukowe w kształceniu i superwizji mogą usprawniać pracę terapeutyczną (Mallard-Swanson i in., 2021; Seegan i in., 2023). Literatura naukowa oferuje odpowiedzi na pytania nurtujące osoby praktykujące psychoterapię, a co za tym idzie może być źródłem wsparcia i zasobów w podejmowaniu decyzji. Odnosząc się do przykładów z doświadczenia autora, zostało mu przedstawione pytanie o zastosowanie terapii z przedłużoną ekspozycją w przypadku osoby z doświadczeniem psychozy i wątpliwości co do niekorzystnych efektów tej metody w zakresie innych objawów niż te związane z zaburzeniami po traumie. Tymczasem, dostępne przeglądy dowodów naukowych, nawet jeśli nie oferują definitywnej odpowiedzi, wskazują na bezpieczeństwo stosowania przedłużonej ekspozycji w tej populacji pacjentów (Deleuran i in., 2024; Grubaugh i in., 2017), a sama procedura nie nasila, a prawdopodobnie zmniejsza nasilenie, innych objawów psychopatologicznych (van Minnen i in., 2015). Polem wspólnego działania dla osób praktykujących psychoterapię a osobami prowadzącymi badania może być, między innymi, formułowanie właśnie tego rodzaju istotnych dla praktyki pytań.

Kolejnym ważnym sposobem wykorzystywania syntezy dowodów naukowych są rekomendacje i zalecenia terapeutyczne (por. Kowalski i in., 2024), które formułują towarzystwa zawodowe albo agendy rządowe. Zalecenia formułowane są na bazie dostępnych dowodów, które następnie oceniane są pod kątem pewności z jaką można się na tych dowodach opierać (tj. ich jakości). Tutaj znaczenie ma ocena ryzyka tendencyjności, ale też niespójności i braku precyzji. Najczęściej stosowanym narzędziem do oceny jakości jest GRADE (*Grading of Recommendations Assessment, Development and Evaluation*; Schünemann i in., 2025), które, jest np. zalecane w ocenianiu psychoterapii bazujących na dowodach (Tolin i in., 2015).

Tego rodzaju zalecenia formułowane są m.in. przez Amerykańskie Towarzystwo Psychologiczne (APA, b.d.), Amerykańskie Towarzystwo Psychiatryczne (APA, 2020) czy brytyjski *National Institute for Health and Care Excellence* (NICE, 2014). Przykładowo, APA psychiatryczne posiada w swoich strukturach komitet zajmujący się koordynowaniem procesów tworzenia rekomendacji terapeutycznych. Członkowie APA z doświadczeniem klinicznym i naukowym przygotowują tekst rekomendacji we współpracy z komitetem oraz doradcami wyspecjalizowanymi w tworzeniu przeglądów systematycznych. Tekst jest recenzowany zarówno przez władze towarzystwa, jak i ekspertów, którzy nie brali udziału w przygotowaniu danych zaleceń. Brytyjskie NICE również skodyfikowało proces tworzenia rekomendacji terapeutycznych, w których są uwzględniane nie tylko skrupulatnie oceniane dowody naukowe, lecz także ekonomiczna ocena proponowanych strategii prowadzenia leczenia osób z danym zaburzeniem. NICE poddaje przygotowywane rekomendacje konsultacjom z interesariuszami (np. grupami pacjentów czy organizacjami pozarządowymi), a także ma możliwość poddania rekomendacji zewnętrznej recenzji merytorycznej (*peer review*).

Podsumowanie

Niniejszy tekst miał na celu przybliżenie osobom praktykującym psychoterapię zagadnień istotnych dla wykorzystywania metaanaliz badań nad psychoterapią w podejmowaniu decyzji klinicznych. Opisano teoretyczne oraz proceduralne podstawy sporządzania syntez badań. Dużą uwagę przywiązano do aspektów formalnych, pozwalających na czytanie tego rodzaju prac badawczych z większym zrozumieniem, ale też większą świadomością ich ograniczeń. Poruszona mnogość wątków i pobieżny sposób w jaki niektóre z nich zostały potraktowane oznacza, że niniejsza praca w żaden sposób nie wyczerpuje istoty zagadnienia, a wręcz może stanowić furtkę do dalszego, samodzielnego zgłębiania tematu. Dlatego też starano się odnosić do najbardziej podstawowych, relevantnych, ale też aktualnych tekstów, aby artykuł mógł stanowić przewodnik do samodzielnego zgłębiania literatury przedmiotu. Powodzenia na szlaku!

Podziękowania

Autor serdecznie dziękuje Julii Szymańskiej, Weronice Browarczyk, Przemysławowi Łukasiewiczowi oraz Magdalenie Pietruch za cenne uwagi do pierwszej wersji artykułu.

Bibliografia

Afonso, J., Ramirez-Campillo, R., Clemente, F. M., Büttner, F. C., Andrade, R. (2024). The perils of misinterpreting and misusing “publication bias” in meta-analyses: An education

- review on funnel plot-based methods. *Sports Medicine*, 54(2), 257–269. <https://doi.org/10.1007/s40279-023-01927-9>
- American Psychiatric Association (2020). Development process for practice guidelines of the American Psychiatric Association – revised. <https://www.psychiatry.org/getmedia/0b96df17-66a7-4f49-8159-d6522615f047/APA-Guideline-Development-Process.pdf>
- American Psychological Association (b.d.). APA clinical practice guideline development. www.apa.org/about/offices/directorates/guidelines/clinical-practice
- Andersson, E., Aspvall, K., Schettini, G., Kraepelien, M., Särholm, J., Wergeland, G. J., Öst, L. G. (2025). Efficacy of metacognitive interventions for psychiatric disorders: A systematic review and meta-analysis. *Cognitive Behaviour Therapy*, 54(2), 276–302. <https://doi.org/10.1080/16506073.2024.2434920>
- Andrade, C. (2015). Understanding relative risk, odds ratio, and related terms: As simple as it can get. *The Journal of Clinical Psychiatry*, 76(7), 857–861. <https://doi.org/10.4088/jcp.15f10150>
- Andrade, C. (2020). Understanding the basics of meta-analysis and how to read a forest plot: As simple as it gets. *The Journal of Clinical Psychiatry*, 81(5), artykuł 20f13698. <https://doi.org/10.4088/JCP.20f13698>
- APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *The American Psychologist*, 61(4), 271–285. <https://doi.org/10.1037/0003-066x.61.4.271>
- Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., Tungpunkom, P. (2015). Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *JBIM Evidence Implementation*, 13(3), 132–140. <https://doi.org/10.1097/xeb.0000000000000055>
- Baker, S. G., Kramer, B. S. (2002). The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C? *BMC Medical Research Methodology*, 2, artykuł 13. <https://doi.org/10.1186/1471-2288-2-13>
- Baker, T. B., McFall, R. M., Shoham, V. (2008). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest*, 9(2), 67–103. <https://doi.org/10.1111/j.1539-6053.2009.01036.x>
- Baker, W. L., Michael White, C., Cappelleri, J. C., Kluger, J., Coleman, C. I., From the Health Outcomes, Policy, and Economics (HOPE) Collaborative Group. (2009). Understanding heterogeneity in meta-analysis: The role of meta-regression. *International Journal of Clinical Practice*, 63(10), 1426–1434. <https://doi.org/10.1111/j.1742-1241.2009.02168.x>
- Begg, C. B., Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50(4), 1088–1101.
- Blease, C., Kelley, J. M., Trachsel, M. (2018). Informed consent in psychotherapy: Implications of evidence-based practice. *Journal of Contemporary Psychotherapy*, 48, 69–78. <https://doi.org/10.1007/s10879-017-9372-9>
- Boot, W. R., Simons, D. J., Stothart, C., Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo

- effects. *Perspectives on Psychological Science*, 8(4), 445–454. <https://doi.org/10.1177/1745691613491271>
- Borenstein, M., Higgins, J. P., Hedges, L. V., Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8(1), 5–18. <https://doi.org/10.1002/jrsm.1230>
- Boutron, I., Page, M. J., Higgins, J. P. T., Altman, D. G., Lundh, A., Hróbjartsson, A. (2023). Chapter 7: Considering bias and conflicts of interest among the included studies. W: Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., Welch, V. A. (red.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*. Cochrane. Pobrane 21 maja 2025 z: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-07>
- Bowes, S. M., Ammirati, R. J., Costello, T. H., Basterfield, C., Lilienfeld, S. O. (2020). Cognitive biases, heuristics, and logical fallacies in clinical practice: A brief field guide for practicing clinicians and supervisors. *Professional Psychology: Research and Practice*, 51(5), 435–445. <https://doi.org/10.1037/pro0000309>
- Buchman, D., Rosenbaum, D. (2024). Psychedelics in PeRil: The commercial determinants of health, financial entanglements and population health ethics. *Public Health Ethics*, 17(1–2), 24–39. <https://doi.org/10.1093/phe/phae002>
- Burns, P. B., Rohrich, R. J., Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery*, 128(1), 305–310. <https://doi.org/10.1097/prs.0b013e318219c171>
- Caldwell, D. M., Ades, A. E., Higgins, J. P. T. (2005). Simultaneous comparison of multiple treatments: Combining direct and indirect evidence. *BMJ*, 331(7521), 897–900. <https://doi.org/10.1136/bmj.331.7521.897>
- Chan, A. W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA*, 291(20), 2457–2465. <https://doi.org/10.1001/jama.291.20.2457>
- Ciharova, M., Karyotaki, E., Miguel, C., Walsh, E., de Ponti, N., Amarnath, A., van Ballegooijen, W., Riper, H., Arroll, B., Cuijpers, P. (2024). Amount and frequency of psychotherapy as predictors of treatment outcome for adult depression: A meta-regression analysis. *Journal of Affective Disorders*, 359, 92–99. <https://doi.org/10.1016/j.jad.2024.05.070>
- Cohn, L. D., Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, 8(3), 243–253. <https://doi.org/10.1037/1082-989x.8.3.243>
- Copay, A. G., Subach, B. R., Glassman, S. D., Polly Jr., D. W., Schuler, T. C. (2007). Understanding the minimum clinically important difference: A review of concepts and methods. *The Spine Journal*, 7(5), 541–546. <https://doi.org/10.1016/j.spinee.2007.01.008>
- Cuijpers, P., Cristea, I. A. (2016). How to prove that your therapy is effective, even when it is not: A guideline. *Epidemiology and Psychiatric Sciences*, 25(5), 428–435. <https://doi.org/10.1017/s2045796015000864>
- Cuijpers, P., Harrer, M., Miguel, C., Ciharova, M., Papola, D., Basic, D., Botella, C., Cristea, I., de Ponti, N., Donker, T., Driessen, E., Franco, P., Gómez-Gómez, I., Hamblen, J.,

- Jiménez-Orenga, N., Karyotaki, E., Keshen, A., Linardon, J., Motrico, E., Matbouri-ahi, M., Panagiotopoulou, O. M., Pfund, R. A., Plessen, C. Y., Riper, H., Schnurr, P. P., Sijbrandij, M., Toffolo, M. B. J., Tong, L., van Ballegooijen, W., van der Ven, E., van Straten, A., Wang, Y., Furukawa, T. A. (2025). Cognitive behavior therapy for mental disorders in adults: A unified series of meta-analyses. *JAMA Psychiatry*, 82(6), 563–571. <https://doi.org/10.1001/jamapsychiatry.2025.0482>
- Cuijpers, P., Karyotaki, E., de Wit, L., Ebert, D. D. (2020). The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychotherapy Research*, 30(3), 279–293. <https://doi.org/10.1080/10503307.2019.1649732>
- Cumming, G., Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60(2), 170–180. <https://doi.org/10.1037/0003-066x.60.2.170>
- Davies, K. S. (2011). Formulating the evidence based practice question: A review of the frameworks. *Evidence Based Library and Information Practice*, 6(2), 75–80. <https://doi.org/10.18438/B8WS5N>
- Davis, A. K., Barrett, F. S., May, D. G., Cosimano, M. P., Sepeda, N. D., Johnson, M. W., Finan, P. H., Griffiths, R. R. (2021). Effects of psilocybin-assisted therapy on major depressive disorder: A randomized clinical trial. *JAMA Psychiatry*, 78(5), 481–489. <https://doi.org/10.1001/jamapsychiatry.2020.3285>
- Deeks, J. J., Higgins, J. P. T., Altman, D. G. (2023). Chapter 10: Analysing data and undertaking meta-analyses. W: J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, V. A. Welch (red.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*. Cochrane. Pobrane 21 maja 2025 z: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-10>
- Deleuran, D. H. K., Skov, O., Bo, S. (2024). Prolonged exposure for posttraumatic stress disorder in patients exhibiting psychotic symptoms: A scoping review. *Clinical Psychology & Psychotherapy*, 31(4), artykuł e3027. <https://doi.org/10.1002/cpp.3027>
- Dozois, D. J. A., Mikail, S. F., Alden, L. E., Bieling, P. J., Bourgon, G., Clark, D. A., Drapeau, M., Gallson, D., Greenberg, L., Hunsley, J., Johnston, C. (2014). The CPA Presidential Task Force on Evidence-Based Practice of Psychological Treatments. *Canadian Psychology/Psychologie canadienne*, 55(3), 153–160. <https://doi.org/10.1037/a0035767>
- Duval, S., Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>
- Egger, M., Smith, G. D., Schneider, M., Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, 12(1), 77–84. <https://doi.org/10.1046/j.1365-2702.2003.00662.x>
- Furukawa, T. A. (1999). From effect size into number needed to treat. *The Lancet*, 353(9165), artykuł 1680. [https://doi.org/10.1016/s0140-6736\(99\)01163-0](https://doi.org/10.1016/s0140-6736(99)01163-0)

- Furukawa, T. A., Leucht, S. (2011). How to obtain NNT from Cohen's d: Comparison of two methods. *PloS One*, 6(4), artykuł e19070. <https://doi.org/10.1371/journal.pone.0019070>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8. <https://doi.org/10.3102/0013189X005010003>
- Grubaugh, A. L., Veronee, K., Ellis, C., Brown, W., Knapp, R. G. (2017). Feasibility and efficacy of prolonged exposure for PTSD among individuals with a psychotic spectrum disorder. *Frontiers in Psychology*, 8, artykuł 977. <https://doi.org/10.3389/fpsyg.2017.00977>
- Gyani, A., Shafran, R., Layard, R., Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, 51(9), 597–606. <https://doi.org/10.1016/j.brat.2013.06.004>
- Gyani, A., Shafran, R., Myles, P., Rose, S. (2014). The gap between science and practice: How therapists make their clinical decisions. *Behavior Therapy*, 45(2), 199–211. <https://doi.org/10.1016/j.beth.2013.10.004>
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61(2), 155–163. <https://doi.org/10.1002/jclp.20108>
- Harrer, M., Cuijpers, P., Schuurmans, L. K., Kaiser, T., Buntrock, C., van Straten, A., Ebert, D. (2023). Evaluation of randomized controlled trials: A primer and tutorial for mental health researchers. *Trials*, 24(1), artykuł 562. <https://doi.org/10.1186/s13063-023-07596-3>
- Hedges, L. V. (2025). Effect sizes for experimental research. *British Journal of Mathematical and Statistical Psychology*. <https://doi.org/10.1111/bmsp.12389>
- Higgins, J. P., Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Higgins, J. P., Thompson, S. G., Deeks, J. J., Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., Sterne, J. A. C. (b.d.). Chapter 8: Assessing risk of bias in a randomized trial. W: J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, V. A. Welch (red.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*. Cochrane. Pobrane 21 maja 2025 z: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-08>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., Welch, V. A. (red.). (2024). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5*. Pobrane 21 maja 2025r. z: www.training.cochrane.org/handbook
- Hutton, B., Salanti, G., Caldwell, D. M., Chaimani, A., Schmid, C. H., Cameron, C., Ioannidis, J. P., Straus, S., Thorlund, K., Jansen, J. P., Mulrow, C., Catalá-López, F., Gøtzsche, P. C., Dickersin, K., Boutron, I., Altman, D. G., Moher, D. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Annals of Internal Medicine*, 162(11), 777–784. <https://doi.org/10.7326/m14-2385>

- IntHout, J., Ioannidis, J. P. A., Rovers, M. M., Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6, artykuł e010247. <https://doi.org/10.1136/bmjopen-2015-010247>
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), artykuł e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Jacobson, N. S., Truax, P. (1991). Clinical significance: A statistical approach to denning meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037//0022-006x.59.1.12>
- JASP Team (2024). JASP (Version 0.18.3) [oprogramowanie komputerowe].
- Jobst, A., Brakemeier, E. -L., Buchheim, A., Caspar, F., Cuijpers, P., Ebmeier, K. P., Falkai, P., Jan van der Gaag, R., Gaebel, W., Herpertz, S., Kurimay, T., Sabaß, L., Schnell, K., Schramm, E., Torrent, C., Wasserman, D., Wiersma, J., Padberg, F. (2016). European Psychiatric Association Guidance on psychotherapy in chronic depression across Europe. *European Psychiatry*, 33(1), 18–36. <https://doi.org/10.1016/j.eurpsy.2015.12.003>
- Kirsch, I. (2005). Placebo psychotherapy: Synonym or oxymoron? *Journal of Clinical Psychology*, 61(7), 791–803. <https://doi.org/10.1002/jclp.20126>
- Kowalski, J. (2024). Possible errors in a meta-analysis on the efficacy of psychodynamic therapy in social anxiety disorder (Qiqi Zhang et al., 2022). *Psychiatry Research*, 342, artykuł 116174. <https://doi.org/10.1016/j.psychres.2024.116174>
- Kowalski, J., Blaut, A., Dragan, M., Farley, D., Pankowski, D., Sanna, K., Śliwerski, A., Wiśniowska, J. (2024). *Systematyczny narracyjny przegląd metaanaliz badań nad skutecznością psychoterapii poznawczo-behawioralnej i zaleceń terapeutycznych opublikowanych między 2010 a 2023*. Polskie Towarzystwo Terapii Poznawczej i Behawioralnej.
- Kowalski, J., Elżanowski, A., Śliwerski, A. (2023). A review of selected psychotherapies for PTSD, their efficacy and treatment guidelines in adults. *Psychiatria Polska*, 58(2), 315–328. <https://doi.org/10.12740/pp/onlinefirst/157105>
- Kraus, D. R., Castonguay, L., Boswell, J. F., Nordberg, S. S., Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research*, 21(3), 267–276. <https://doi.org/10.1080/10503307.2011.563249>
- Kvarven, A., Strømland, E., Johannesson, M. (2020). Comparing meta-analyses and pre-registered multiple-laboratory replication projects. *Nature Human Behaviour*, 4(4), 423–434. <https://doi.org/10.1038/s41562-019-0787-z>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, artykuł 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., Olkin, I. (2006). The case of the misleading funnel plot. *BMJ*, 333(7568), 597–600. <https://doi.org/10.1136/bmj.333.7568.597>
- Lemarchand, C., Chopin, R., Paul, M., Braillon, A., Cosgrove, L., Cristea, I., Fried, E. I., Turner, E. H., Naudet, F. (2024). Fragile promise of psychedelics in psychiatry. *BMJ*, 387, artykuł e080391. <https://doi.org/10.1136/bmj-2024-080391>

- Leon, A. C. (2011). Comparative effectiveness clinical trials in psychiatry: Superiority, noninferiority, and the role of active comparators. *Journal of Clinical Psychiatry*, 72(10), 1344–1349. <https://doi.org/10.4088/jcp.10m06089whi>
- Levi, O., Bar-Haim, Y., Kreiss, Y., Fruchter, E. (2016). Cognitive-behavioural therapy and psychodynamic psychotherapy in the treatment of combat-related post-traumatic stress disorder: A comparative effectiveness study. *Clinical Psychology & Psychotherapy*, 23(4), 298–307. <https://doi.org/10.1002/cpp.1969>
- 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. *Journal of Clinical Epidemiology*, 62(10), 1–34. <https://doi.org/10.1016/j.jclinepi.2009.06.006>
- Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., Latzman, R. D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, 9(4), 355–387. <https://doi.org/10.1177/1745691614535216>
- Magnusson, K. (2025). *Interpreting Cohen's d effect size: An interactive visualization (Version 2.6.0)*. R Psychologist. Pobrane 21 maja 2025 z: <https://rpsychologist.com/cohend>
- Mallard Swanson, K., Song, J., Beristianos, M., Aajmain, S., Lane, J. E. M., Landy, M. S. H., Suvak, M. K., Shields, N., Monson, C. M., Stirman, S. W. (2021). A glimpse into the “black box”: Which elements of consultation in an EBP are associated with client symptom change and therapist fidelity? *Implementation Research and Practice*, 2. <https://doi.org/10.1177/26334895211051791>
- McAleavey, A. A. (2024). When (not) to rely on the reliable change index: A critical appraisal and alternatives to consider in clinical psychology. *Clinical Psychology: Science and Practice*, 31(3), 351–366. <https://doi.org/10.1037/cps0000203>
- Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, 14(1), artykuł 579. <https://doi.org/10.1186/s12913-014-0579-0>
- Michopoulos, I., Furukawa, T. A., Noma, H., Kishimoto, S., Onishi, A., Ostinelli, E. G., Ciharova, M., Miguel, C., Karyotaki, E., Cuijpers, P. (2021). Different control conditions can produce different effect estimates in psychotherapy trials for depression. *Journal of Clinical Epidemiology*, 132, 59–70. <https://doi.org/10.1016/j.jclinepi.2020.12.012>
- Miguel, C., Harrer, M., Karyotaki, E., Plessen, C. Y., Ciharova, M., Furukawa, T. A., Cristea, I. A., Cuijpers, P. (2025). Self-reports vs clinician ratings of efficacies of psychotherapies for depression: A meta-analysis of randomized trials. *Epidemiology and Psychiatric Sciences*, 34, artykuł e15. <https://doi.org/10.1017/s2045796025000095>
- Moritz, S., Nestoriuc, Y., Rief, W., Klein, J. P., Jelinek, L., Peth, J. (2019). It can't hurt, right? Adverse effects of psychotherapy in patients with depression. *European Archives of Psychiatry and Clinical Neuroscience*, 269(5), 577–586. <https://doi.org/10.1007/s00406-018-0931-1>

- Munder, T., Brüttsch, O., Leonhart, R., Gerger, H., Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review*, 33(4), 501–511. <https://doi.org/10.1016/j.cpr.2013.02.002>
- National Institute for Health and Care Excellence (2014). *Developing NICE guidelines: The manual*. Pobrane 21 maja 2025 z: <https://www.nice.org.uk/process/pmg20/chapter/introduction>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Nye, A., Delgadillo, J., Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*, 91(7), 389–397. <https://doi.org/10.1037/ccp0000820>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), artykuł aac4716. <https://doi.org/10.1126/science.aac4716>
- Page, M. J., Sterne, J. A. C., Boutron, I., Hróbjartsson, A., Kirkham, J. J., Li, T., Lundh, A., Mayo-Wilson, E., McKenzie, J. E., Stewart, L. A., Sutton, A. J., Bero, L., Dunn, A. G., Dwan, K., Elbers, R. G., Kanukula, R., Meerpohl, J. J., Turner, E. H., Higgins, J. P. T. (2023). ROB-ME: A tool for assessing risk of bias due to missing evidence in systematic reviews with meta-analysis. *BMJ*, 383, artykuł e076754. <https://doi.org/10.1136/bmj-2023-076754>
- Philips, B., Falkenström, F. (2021). What research evidence is valid for psychotherapy research? *Frontiers in Psychiatry*, 11, artykuł 625380. <https://doi.org/10.3389/fpsy.2020.625380>
- Riehm, K. E., Azar, M., Thombs, B. D. (2015). Transparency of outcome reporting and trial registration of randomized controlled trials in top psychosomatic and behavioral health journals: A 5-year follow-up. *Journal of Psychosomatic Research*, 79(1), 1–12. <https://doi.org/10.1016/j.jpsychores.2015.04.010>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Rothstein, H. R., Sutton, A. J., Borenstein, M. (red.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.
- Sandoval-Lentisco, A., Tortajada, M., López-Nicolás, R., López-López, J. A., Wagenmakers, E. J., Sánchez-Meca, J., Hardwicke, T. E. (2025). Preregistration of psychology meta-analyses: A cross-sectional study of prevalence and practice. *Advances in Methods and Practices in Psychological Science*, 8(1), artykuł 25152459241300113. <https://doi.org/10.1177/25152459241300113>
- Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next

- generation evidence synthesis tool. *Research Synthesis Methods*, 3(2), 80–97. <https://doi.org/10.1002/jrsm.1037>
- Schmidt, C. O., Kohlmann, T. (2008). When to use the odds ratio or the relative risk? *International Journal of Public Health*, 53(3), 165–167. <https://doi.org/10.1007/s00038-008-7068-3>
- Schünemann, H. J., Higgins, J. P. T., Vist, G. E., Glasziou, P., Akl, E. A., Skoetz, N., Guyatt, G. H. (2024). Chapter 14: Completing ‘Summary of findings’ tables and grading the certainty of the evidence. W: J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, V. A. Welch (red.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.5.1*. Cochrane. Pobrane 21 maja 2025 z: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-14>
- Sedgwick, P. (2015a). How to read a forest plot in a meta-analysis. *BMJ*, 351, artykuł H4028. <https://doi.org/10.1136/bmj.h4028>
- Sedgwick, P. (2015b). What is publication bias in a meta-analysis?. *BMJ*, 351, artykuł h4419. <https://doi.org/10.1136/bmj.h4419>
- Sedgwick, P., Marston, L. (2015). How to read a funnel plot in a meta-analysis. *BMJ*, 351, artykuł h4718. <https://doi.org/10.1136/bmj.h4718>
- Seegan, P. L., Miller, L., Young, A. S., Parrish, C., Cullen, B., Reynolds, E. K. (2023). Enhancing quality of care through evidence-based practice: Training and supervision experiences. *American Journal of Psychotherapy*, 76(3), 100–106. <https://doi.org/10.1176/appi.psychotherapy.20220015>
- Shadish, W. R., Lecy, J. D. (2015). The meta-analytic big bang. *Research Synthesis Methods*, 6(3), 246–264. <https://doi.org/10.1002/jrsm.1132>
- 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, artykuł 10. <https://doi.org/10.1186/1471-2288-7-10>
- Shea, B. J., Bouter, L. M., Peterson, J., Boers, M., Andersson, N., Ortiz, Z., Ramsay, T., Bai, A., Shukla, V. K., Grimshaw, J. M. (2007). External validation of a measurement tool to assess systematic reviews (AMSTAR). *PloS One*, 2(12), artykuł e1350. <https://doi.org/10.1371/journal.pone.0001350>
- Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., Henry, D. A., Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), 1013–1020. <https://doi.org/10.1016/j.jclinepi.2008.10.009>
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., 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, artykuł j4008. <https://doi.org/10.1136/bmj.j4008>
- Simmonds, M. (2015). Quantifying the risk of error when interpreting funnel plots. *Systematic Reviews*, 4, artykuł 24. <https://doi.org/10.1186/s13643-015-0004-8>

- Sleight, P. (2000). Debate: Subgroup analyses in clinical trials: Fun to look at-but don't believe them! *Trials*, 1(1), 25–27. <https://doi.org/10.1186/cvm-1-1-025>
- Smith, M. L., Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32(9), 752–760. <https://doi.org/10.1037/0003-066X.32.9.752>
- Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. *Journal of Clinical Psychology*, 63(7), 611–631. <https://doi.org/10.1002/jclp.20373>
- Sterne, J. A. C., Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54(10), 1046–1055. [https://doi.org/10.1016/s0895-4356\(01\)00377-8](https://doi.org/10.1016/s0895-4356(01)00377-8)
- Sterne, J. A. C., Becker, B. J., Egger, M. (2005). The funnel plot. W: H. R. Rothstein, A. J. Sutton, M. Borenstein, M. (red.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (s. 73–98). Wiley. <https://doi.org/10.1002/0470870168.ch5>
- Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. -W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Piggott, T. D., Ramsay, C. R., Rothstein, H. R., Sandhu, L., Santaguida, P. L., Schünemann, H. J., Shea, B., Shrier, I., Tugwell, P., Turner, L., Valentine, J. C., Waddington, H., Waters, E., Wells, G. A., Whiting, P. F., Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, artykuł i4919. <https://doi.org/10.1136/bmj.i4919>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. -Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., McAleenan, A., Reeves, B. C., Shepperd, S., Shrier, I., Stewart, L. A., Tilling, K., White, I. R., Whiting, P. F., Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366, artykuł l4898. <https://doi.org/10.1136/bmj.l4898>
- Sterne, J. A. C., Hernan, M. A., McAleenan, A., Reeves, B. C., Higgins, J. P. T. (2023). Chapter 25: Assessing risk of bias in a non-randomized study. W: J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, V. A. Welch (red.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*. Cochrane. Pobrane 21 maja 2025 z: www.training.cochrane.org/handbook
- Stoll, M., Mancini, A., Hubenschmid, L., Dreimüller, N., König, J., Cuijpers, P., Barth, J., Lieb, K. (2020). Discrepancies from registered protocols and spin occurred frequently in randomized psychotherapy trials—a meta-epidemiologic study. *Journal of Clinical Epidemiology*, 128, 49–56. <https://doi.org/10.1016/j.jclinepi.2020.08.013>
- 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), 317–338. <https://doi.org/10.1037/h0101729>
- Tolin, D. F., Grasso, D., Boness, C. L., Beck, J. G., Keane, T. M., Leichenring, F., Olatunji, B. O., Otto, M. W., Weinand, J. (2025). A proposed definition of psychological treatment and its relation to empirically supported treatments. *Clinical Psychology: Science and Practice*, 32(3), 213–225. <https://doi.org/10.1037/cps0000220>

- van Minnen, A., Zoellner, L. A., Harned, M. S., Mills, K. (2015). Changes in comorbid conditions after prolonged exposure for PTSD: A literature review. *Current Psychiatry Reports*, 17(3), artykuł 549. <https://doi.org/10.1007/s11920-015-0549-1>
- Walfish, S., McAlister, B., O'Donnell, P., Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports*, 110(2), 639–644. <https://doi.org/10.2466/02.07.17.pr0.110.2.639-644>
- Watts, S. E., Turnell, A., Kladnitski, N., Newby, J. M., Andrews, G. (2015). Treatment-as-usual (TAU) is anything but usual: A meta-analysis of CBT versus TAU for anxiety and depression. *Journal of Affective Disorders*, 175, 152–167. <https://doi.org/10.1016/j.jad.2014.12.025>
- Whiting, P., Savović, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., Churchill, R., ROBIS group (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, 69, 225–234. <https://doi.org/10.1016/j.jclinepi.2015.06.005>