

Jolanta Matera
Joanna Czapska

Zarys metody przeglądu systematycznego w naukach społecznych



POMOC TECHNICZNA
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



INSTYTUT
BADAN
EDUKACYJNYCH

UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO



Projekt realizowany w ramach konkursu dotacji organizowanego przez Ministerstwo Infrastruktury i Rozwoju, współfinansowanego ze środków Unii Europejskiej w ramach Programu Operacyjnego Pomoc Techniczna.

Jolanta Matera
Joanna Czapska

Zarys metody przeglądu systematycznego w naukach społecznych



Warszawa 2014



POMOC TECHNICZNA
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



INSTYTUT
BADAN
EDUKACYJNYCH

UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO



Projekt realizowany w ramach konkursu dotacji organizowanego przez Ministerstwo Infrastruktury i Rozwoju, współfinansowanego ze środków Unii Europejskiej w ramach Programu Operacyjnego Pomoc Techniczna.

Redakcja naukowa: dr Michał Sitek

Recenzenci:

Elisabeth Buk-Berge, PhD
prof. dr hab. Jarosław Górniak

Autorzy: Jolanta Matera, Joanna Czapska

Suplement: Michał Sitek

Redakcja: Katarzyna Maria Grabarczyk
Korekta: Agnieszka Rytel
Skład: Katarzyna Lemańska

Wydawca:

Instytut Badań Edukacyjnych

ul. Górczewska 8

01-180 Warszawa

tel. (22) 241 71 00; www.ibe.edu.pl

© Copyright by: Instytut Badań Edukacyjnych, Warszawa 2014

Druk:

QPrint Center

ul. Bema 65

01-244 Warszawa

Wydanie I

ISBN 978-83-61693-58-1

Publikacja opracowana w ramach projektu: *Metoda przeglądów systematycznych jako narzędzie polityki opartej na dowodach (evidence based policy), przydatnej dla pozyskiwania środków z funduszy UE oraz wdrażania i realizacji projektów*, realizowanego przez Instytut Badań Edukacyjnych w ramach Konkursu Dotacji - „Fundusze europejskie na poziomie NSS” VI edycja, organizowanego przez Ministerstwo Infrastruktury i Rozwoju, wspólnie finansowanego ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego (Program Operacyjny Pomoc Techniczna). Nr projektu: POPT.03.01.00-00-383/13. Nr umowy o dofinansowanie: DKS /BDG II /POPT/75/14.

Egzemplarz bezpłatny

Spis treści

Wprowadzenie	5
Przeglądy systematyczne a fundusze europejskie.....	11
1.....	16
Definicja przeglądu systematycznego	16
Mity na temat przeglądów systematycznych	21
Zalety i ograniczenia metody	24
2.....	28
Definiowanie pytania przeglądowego.....	28
Powołanie zespołu badaczy i ekspertów	33
Opracowanie protokołu.....	34
3.....	36
Strategia wyszukiwania literatury	36
Selekcja źródeł – faza wstępna	39
Selekcja tekstów według kryteriów włączenia i wyłączenia	42
4.....	43
Ekstrakcja danych z badań włączonych do przeglądu.....	43
Ocena krytyczna	45
Synteza wyników badań	48
5.....	56
Ocena stroniczości publikacji	56
Opracowanie raportu lub artykułu	56
Upowszechnianie wyników	57
Zakończenie	58
Suplement	60
Glosariusz	72
Bibliografia	76
Netografia.....	80

Wprowadzenie

Tło

Niniejszy tekst sytuuje się w szerokiej dziedzinie *evidence based (informed) policy*. Pojęcie *evidence* (dowody) bywa używane wymiennie z pojęciem *wiedza (knowledge-based policy and practice in education)* i oznacza m.in. wiedzę uzyskaną w wyniku przeglądów systematycznych (European Commission, 2007).

Celem niniejszej publikacji jest opis metody przeglądu systematycznego – obiektywnej i powtarzalnej oceny badań/programów społecznych – dostarczającej wiarygodnych danych na rzecz świadomego kształtowania polityk społecznych, w tym edukacyjnych. Znaczenie dowodów naukowych w kontekście *evidence-based policy* (polityk korzystających z najlepszych dowodów w podejmowaniu decyzji) zostało podkreślone w publikacji *Evidence in Education. Linking research and policy* (OECD, 2007).

Zarys metody przeglądu systematycznego w naukach społecznych to w Polsce pionierska publikacja dotycząca przeglądów systematycznych w zastosowaniu do badań społecznych, ze wskazaniem na badania edukacyjne. Jest to jednocześnie ważny krok w popularyzacji tej metody. Ze względu na wąski zakres czasowy projektu (trzy miesiące), w którego ramach powstała niniejsza publikacja, jak również przyjęty format broszurowy, zagadnienie przedstawiono szkicowo, jako opis metody i przyczynek do przyszłych polskich publikacji poświęconych przeglądom systematycznym badań podstawowych lub programów UE. Zdajemy sobie jednak sprawę, że obecnie (wrzesień, 2014) nasz *Zarys* – choć nie wyczerpuje wszystkich aspektów dotyczących procesu przeglądu systematycznego – stanowi źródło wstępnych informacji zarówno dla przyszłych autorów przeglądów systematycznych, jak i użytkowników, w tym decydentów, szczególnie w obszarze polityki edukacyjnych.

Przeglądy systematyczne (PS) syntetyzują wyniki badań podstawowych lub programów (interwencji) za pomocą przejrzystych metod: wyszukiwania, oceny, wyodrębniania i analizy danych, w sposób odpowiedzialny i podlegający aktualizacji. Przegląd tego typu zapewnia dostęp do lepszych jakościowo dowodów naukowych

na dany temat, niż rezultat analizy pojedynczego badania/programu; jako badanie wtórne, PS jest metodą zarządzania wiedzą dotyczącą danej kwestii badawczej w celu uzyskania możliwie najbardziej rzetelnych implikacji dla polityk i praktyk społecznych na poziomach krajowym i lokalnym. Przeglądy systematyczne z założenia dostarczają w sposób przystępny rzetelnych dowodów użytkownikom spoza środowiska naukowego: decydentom i profesjonalistom w wielu dziedzinach życia społecznego, w tym edukacji (nauczycielom, urzędnikom, pracownikom socjalnym, prawnikom i in.).

W społeczeństwie opartym na wiedzy warunkiem zwiększania efektywności podejmowania decyzji publicznych jest rozwój systematycznych metod zarządzania jej zasobami i upowszechniania dowodów naukowych. Ruch *evidence-informed policy* stanowi przejaw ogólnego zainteresowania możliwościami oparcia polityki społecznej na solidnej i przystępnej bazie dowodowej.

Takiemu podejściu sprzyja potrzeba systematyzacji wiedzy edukacyjnej, rozproszonej pomiędzy praktyką a nauką. Z jednej strony mamy bowiem do czynienia z metodami tworzenia wiedzy o skutecznych i odpowiednich praktykach edukacyjnych (nauczania i uczenia się), teoriami na temat edukacji, jej funkcji i celów, wiedzą praktyczną nauczycieli, z drugiej – daleki od ideału i nie zawsze metodologicznie staranny stan badań edukacyjnych. Rosnącą popularność podejścia *evidence-based* wiązać można z nowymi możliwościami, jakie dają technologie informacyjno-komunikacyjne, w tym zwłaszcza łatwość dostępu do informacji naukowej.

Jeśli zatem edukacja ma opierać się na solidnych, naukowych podstawach, potrzebne są systematycznie zebrane, powtarzalne i dające się uogólnić dowody, oceniające, np. skutki stosowanych praktyk lub interwencji. Biorąc pod uwagę tysiące badań przeprowadzonych w Polsce w dziedzinie edukacji, co najmniej w kilku obszarach tematycznych można by wyodrębnić całe korpusy dowodów – wyników badań, stanowiących zasób i źródło profesjonalnych działań.

Systematyczne łączenie wyników różnych badań ma długą historię, jednak popularność tej metody znacznie wzrosła w latach dziewięćdziesiątych XX wieku. W ostatnich dwóch dekadach osiągnięto znaczne postępy w metodach statystycznych wspierających porównywanie wyników różnych badań (np. metaanaliza).

Na przełomie lat siedemdziesiątych i osiemdziesiątych XX wieku grupa naukowców z Oxfordu pod przewodnictwem lekarza Archibalda Cochrane'a opracowała podstawy

medycyny „opartej na dowodach”, zapoczątkowując metodologię przeglądów systematycznych dotyczących skuteczności interwencji medycznych. W roku 1993 powstała instytucja o nazwie The Cochrane Collaboration (www.cochrane.org) – międzynarodowa organizacja badaczy, wykładowców, praktyków i użytkowników zaangażowanych w tę formę zarządzania skumulowaną, rzetelną i dostępną wiedzą, jaką są przeglądy systematyczne. Przeglądy Cochrane’a przeważnie, choć nie wyłącznie, koncentrują się na randomizowanych badaniach eksperymentalnych (RCT), w tym na styku zdrowia i edukacji.

Siostrzaną organizacją wobec The Cochrane Collaboration jest instytucja The Campbell Collaboration, z siedzibą w Stanach Zjednoczonych, która adaptowała metodologię Cochrane’a, dostosowując ją do nauk społecznych, w tym edukacji. The Campbell Collaboration (C2) przeprowadza, archiwizuje, aktualizuje i udostępnia przeglądy systematyczne dotyczące interwencji społecznych, stale udoskonalając metody przeglądowe.

Wspomniane powyżej dwie organizacje wpłynęły na powstanie wielu podobnych inicjatyw, co pokazuje, że politycy i decydenci z wielu krajów i przedstawiciele różnych dyscyplin naukowych widzą potrzebę wspierania decyzji politycznych dowodami naukowymi. W ostatnich latach powstały liczne publikacje na temat przeglądów systematycznych, szczególnie w Wielkiej Brytanii i Stanach Zjednoczonych, co potwierdza duże zainteresowanie tą formą przekazu wiedzy naukowej. W Polsce podobne inicjatywy dopiero pojawiają się na horyzoncie nauk społecznych, lecz w dziedzinie edukacji brakuje zarówno przeglądów systematycznych, jak i tworzenia wokół ich produkcji odpowiedniej infrastruktury – przede wszystkim repozytoriów, baz danych czy agencji pośredniczących w przepływie informacji naukowej między badaczami, praktykami i politykami. Podobnie, jak to jest przyjęte w krajach zachodnich, również w Polsce każdy zainteresowany odbiorca powinien mieć dostęp do bazy dowodów uzyskanych z wielu przeglądów systematycznych, dotyczących użyteczności interwencji.

Wprawdzie badania polityk publicznych (*policy research*) są obecne w Polsce w obszarze polityki gospodarczej, jednak w dziedzinie edukacji jest ich niewiele (Federowicz i Sitek, 2012). Na uwagę zasługuje krakowskie Centrum Ewaluacji i Analiz Polityk Publicznych (<http://www.ceapp.uj.edu.pl/>), jako ośrodek wiedzy

prowadzący prace badawczo-rozwojowe w zakresie metodologii ewaluacji, ze szczególnym uwzględnieniem nowoczesnych metod badania efektów netto programów i projektów oraz ustalania związków przyczynowych pomiędzy interwencją publiczną a obserwowanymi efektami.

Zastosowanie metody przeglądów systematycznych można postrzegać na trzech płaszczyznach użyteczności badań. Po pierwsze, na poziomie akademickim, przeglądy systematyczne mogą być zastosowane przez magistrantów i doktorantów wybierających tę metodę (zob. np. Boland, Cherry i Dickson, 2014). Wiąże się to z zapewnieniem materiałów, podręczników, organizacją szkoleń, studiów podyplomowych etc. Po drugie, samoistną wartość mają przeglądy systematyczne publikacji naukowych przydatnych w przygotowaniu interwencji publicznych. Po trzecie, płaszczyzna użyteczności PS dotyczy ewaluacji oraz metaewaluacji. Ewaluacja, rozumiana tu jako „użycie zróżnicowanych metod badań społeczno-ekonomicznych do systematycznego zbierania danych, analizy, interpretacji, oceny i informowania o jakości i wartości programów publicznych” (Olejniczak, 2008, s. 22), może włączyć metody przeglądów systematycznych, w tym wypracowane syntezy danych. Ponadto, przegląd systematyczny ewaluacji jest jedną z odmian metaewaluacji (*meta-evaluation*), badania wtórnego, którego przedmiotem są wcześniejsze ewaluacje, poddane analizie porównawczej, podsumowaniu obserwacji i wniosków (z reguły dotyczących analogicznych programów), zawierającego rekomendacje oraz oceniającego zastosowane techniki i poprawność metodologii (Olejniczak, 2008).

W typologii ewaluacji istotne miejsce ma ewaluacja jako badanie naukowe (*evaluation as research*), koncentrujące się raczej na poszerzaniu wiedzy, wyjaśnianiu zjawisk społecznych, weryfikowaniu teorii (Olejniczak, 2008). Odbiorcami tego typu ewaluacji są decydenci tworzący programy, liderzy, społeczność akademicka. Istnieje też rodzaj ewaluacji zorientowanych na cele (*objectives-oriented*), czyli koncentrujących się na specyfikacji celów programu i pomiarze efektów w odniesieniu do tychże celów. Ich odbiorcami są głównie liderzy i politycy decydujący o strategiach i politykach. Ewaluacje zorientowane na zarządzanie (*management-oriented*), analizujące procesy i działania programu, mają dostarczać technicznych informacji decydującym i menedżerom programu.

Przeglądy systematyczne źródłem dowodów dla polityków i praktyków edukacji

Rządy i instytucje wielu krajów rozwijają w różnym stopniu politykę tworzenia i wykorzystywania zastanej wiedzy, inwestując w infrastrukturę umożliwiającą te działania (Gough i in., 2013) oraz badania, takie jak przeglądy systematyczne, prowadzone wyłącznie na danych zastanych.

Od pewnego czasu, w wyniku współpracy instytucji rządowych i badawczych, zwiększono środki finansowe na działalność ośrodków wiedzy wspierających wykorzystanie dowodów naukowych i infrastruktury na rzecz gromadzenia i upowszechniania dowodów naukowych. W obszarze edukacji powstają takie inicjatywy jak: duńska organizacja informacyjna Danish Clearinghouse for Education (<http://edu.au.dk/en/research/research-areas/danish-clearinghouse-for-educational-research/>), ośrodek londyński EPPI-Centre, Portal Dowodów Edukacyjnych (Educational Evidence Portal) czy europejska organizacja sieciowa na rzecz świadomych polityk i praktyk edukacyjnych: Evidence Informed Policy and Practice in Education in Europe (zob. projekt EIPPEE, www.eippee.eu). Tytuł projektu EIPPEE: „Dowody na rzecz świadomie kształtowanej polityki i praktyki edukacyjnej w Europie” (*Evidence Informed Policy and Practice in Education in Europe – EIPPEE*) odnosi się do nowo utworzonej sieci instytucji (*networking*) udostępniającej informacje o dowodach uzyskanych z wszelkiego rodzaju badań, podejmowanych w każdej części świata w dziedzinie edukacji. Projekt EIPPEE (2011–2013) powstał jako kontynuacja projektu EIPEE (2010–2011) finansowanego przez Komisję Europejską i zwieńczonego raportem końcowym: *Evidence informed policymaking in Education in Europe* (2011).

Działalność ośrodków wiedzy, związanych z podejściem *evidence-based (informed) policy*, koncentruje się wokół tworzenia i upowszechniania rezultatów przeglądów systematycznych dotyczących wpływu i skutków konkretnych rodzajów interwencji lub programów. Przykładem takich organizacji są: w dziedzinie zdrowia – The Cochrane Collaboration, w dziedzinach polityki społecznej, w tym edukacyjnej – The Campbell Collaboration, ośrodki informacyjne w dziedzinie edukacji finansowane przez rząd brytyjski What Works Centres oraz z funduszy federalnych Stanów Zjednoczonych – What Works Clearinghouse (http://ies.ed.gov/ncee/wwc/publications_reviews.aspx),

Best Evidence Encyclopedia i inne. Celem wymienionych ośrodków jest lepsze wykorzystanie dowodów poprzez udostępnianie odpowiednio przetworzonej wiedzy z wielu badań.

Zakres tego, co uznaje się za dowody naukowe na rzecz kształtowania polityki i praktyki, jest szeroki i może obejmować zarówno wiedzę ekspercką, dane z badań ilościowych i jakościowych, jak również wyniki analiz statystycznych. Wysoka jakość polityki, którą rozumieć należy jako zbiór decyzji dotyczących określonego obszaru, w dużej mierze zależy od wysokiej jakości informacji pochodzących z tych źródeł (Office, 1999, za: Górniak i Mazur, 2011).

Podejmowanie świadomych decyzji na skalę społeczną wymaga od decydentów uwzględnienia szeregu czynników, w tym wartości i priorytetów politycznych, dostępności zasobów i informacji z badań i innych form będących źródłem dowodów. Wyniki badań są tylko jednym z tych czynników, choć niewątpliwie znaczącym:

W politykach państwa główną bolączką pozostaje niedostateczny dorobek analityczny. Decyzje publiczne często oparte na wątplych podstawach w zakresie analiz, ewaluacji, dowodów wynikających z badań czy porównań międzynarodowych doświadczeń. Są potrzebne standardy w zakresie projektowania, wykonywania i ewaluowania polityk publicznych. Najważniejsze działania w zakresie projektowania, wykonywania i ewaluowania polityk publicznych powinny być objęte obowiązkową ewaluacją, a administracja publiczna musi się nauczyć integrować różne przedsięwzięcia badawcze i ewaluacyjne (Zybała, 2012a).

Zwiększenie rangi wykorzystania dowodów naukowych podczas procesu podejmowania decyzji w obszarze edukacji wydaje się konieczne. Dowody powinny być solidne, jawne, aktualne, dostępne oraz istotne, tzn. odnoszące się do potrzeb decydentów i praktyków edukacji. Podejście *evidence-based policy* zakłada, że decyzje polityczne winny uwzględniać racjonalną analizę informacji o dostępnych wynikach naukowych, ponieważ polityka bazująca na rzetelnych dowodach jest bardziej efektywna.

Szczególnego znaczenia w tym kontekście nabiera potrzeba właściwej realizacji polityk/programów publicznych. Instytucje publiczne powinny korzystać z wyników badań naukowych w celu precyzyjnego odpowiadania na potrzeby społeczeństwa (Górniak, 2008). Polityka publiczna i jej składowe elementy: „1. Identyfikacja problemu

i szacowanie zasadności podjęcia interwencji publicznej na rzecz jego rozwiązania, 2. Określenie zasad i celów projektowanej polityki publicznej, 3. Wskazanie instrumentarium i zasobów koniecznych dla osiągnięcia ustanowionych celów oraz ocena ich zasadności” (Górniak i Mazur, 2011, s. 12), w sposób naturalny wydają się nieodzownie związane z bezstronnym, naukowym wy tłumaczeniem zasadności ich stosowania. Zatem, obiektywny i przejrzysty sposób syntezy wyników w przeglądach systematycznych sprawia, że wpisują się one w postulat prawidłowej realizacji polityk publicznych. Równie ważną funkcją, jaką może pełnić niniejsza publikacja, jest wykorzystanie przedstawionych informacji do planowania, pozyskiwania i oceny projektów unijnych, w tym projektów mających na celu dostarczanie wiedzy w zakresie użyteczności interwencji (Górniak i Mazur, 2011, s. 10).

Pisząc niniejszą broszurę, podjęliśmy próbę przybliżenia odbiorcom najważniejszych etapów przeglądu systematycznego i zarazem przetransponowania tej metody na płaszczyznę polskich badań, w tym ewaluacji programów edukacyjnych.

Mamy również nadzieję, że lepsze wykorzystanie dowodów naukowych pomoże trafniej zdefiniować cele projektów, uzasadnić podejmowane działania oraz pozyskiwać środki z funduszy UE.

Przeglądy systematyczne a fundusze europejskie

Wraz z rozwojem ruchu na rzecz podejmowania decyzji politycznych na podstawie dowodów naukowych (*evidence-based policy and practice in education*) również w Polsce pojawiła się potrzeba zastosowania metodologii przeglądu systematycznego, umożliwiającej uzyskanie najbardziej solidnych dowodów w danym aspekcie polityki i praktyki edukacyjnej.

Należy podkreślić, że zarządzanie publicznymi finansami wymusza dokładną analizę wydawanych środków. Metoda przeglądów systematycznych, integrująca wyniki wszystkich rzetelnych badań naukowych, stanowi skuteczne narzędzie ekonomicznego gospodarowania środkami publicznymi, w tym pochodzącymi z funduszy europejskich. Jest ona jedną z propozycji naukowego wsparcia optymalnego wykorzystania dotacji ze środków UE.

Wagę polityk publicznych realizowanych w Polsce, na podstawie obiektywnych, trafnych i rzetelnych wyników badań naukowych, podkreślają nie tylko publikacje

naukowe (Górniak, 2008; Górniak i Mazur, 2011) lecz także decydenci. O konieczności prowadzenia polityki opartej na dowodach w polskich i europejskich warunkach mówiono na konferencji *Polityka Spójności Oparta na Wynikach* (Gdańsk, 7–8 lipca 2011). Minister Rozwoju Regionalnego i Infrastruktury Elżbieta Bieńkowska zwróciła uwagę na zależność prawidłowej realizacji strategii i programów od procesu decyzyjnego opartego na wiarygodnych dowodach i faktach.

Podobnie rzecz się ma z funduszami europejskimi, po które tak chętnie sięgają polscy beneficjenci. Efektywne wydatkowanie publicznych pieniędzy, w tym także pochodzących z Unii Europejskiej, współgra ze świadomymi decyzjami politycznymi podejmowanymi na poziomie centralnym, regionalnym i lokalnym, a także z przemyślanymi koncepcjami projektów realizowanych przez beneficjentów w ramach programów UE. Na gruncie polskim, zgodnie z zapisami rozporządzenia Rady Unii Europejskiej zawierającego przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego i zobowiązującego państwa członkowskie UE do przeprowadzania ewaluacji projektów finansowanych z programów pomocowych - *Rozporządzenie Rady (WE) nr 1083/2006 z dnia 11 lipca 2006* – opracowano zarówno dokumenty systemowe, metodologiczne, jak i ekspertyzy oraz plany dotyczące ewaluacji środków finansowych płynących z funduszy UE. Za definicję ewaluacji przyjęto ustalenie Krajowej Jednostki Oceny, według której ewaluacja to „osąd wartości interwencji publicznej dokonany przy uwzględnieniu odpowiednich kryteriów (skuteczności, efektywności, użyteczności, trafności i trwałości) i standardów. Osąd dotyczy zwykle potrzeb, jakie muszą być zaspokojone w wyniku interwencji oraz wyprodukowanych efektów. Ewaluacja oparta jest na specjalnie w tym celu zebranych i zinterpretowanych informacjach za pomocą odpowiedniej metodologii” (*Wytyczne w zakresie ewaluacji programów operacyjnych na lata 2007–2013*, s. 4). Interwencja publiczna, wdrażana w celu poprawy problemowej sytuacji, finansowana ze środków UE, powinna być przeprowadzona świadomie, rzetelnie, na podstawie obiektywnych przesłanek pochodzących z wiarygodnych i pewnych źródeł, np. z badań naukowych. W ramach działań na rzecz ewaluacji przy Ministerstwie Infrastruktury i Rozwoju powstała baza badań ewaluacyjnych, zawierająca opracowania ewaluacyjne dotyczące planowanych i przeprowadzonych interwencji publicznych, także w obszarze edukacji (np. *Ewaluacja działań podejmowanych na rzecz edukacji w ramach EFS*

i EFRR w województwie opolskim z 2012 roku, Ocena wpływu projektów zrealizowanych w ramach Poddziałania 9.1.2 PO KL na wzrost jakości edukacji w województwie lubuskim z 2013 roku czy raport Ocena rozwoju szkolnictwa zawodowego w województwie podkarpackim w kontekście Działania 9.2 PO KL z 2012 roku).

Zarówno decydentom, jak i beneficjentom rezultaty przeglądów systematycznych, jako szczególny rodzaj ewaluacji, mogą przydać się w podejmowaniu trafnych decyzji dotyczących m.in. kierunków i formy udzielanego wsparcia czy obszaru konkretnej interwencji.

Metody przeglądu systematycznego, zgodnie z założeniami projektu, znajdują swoje zastosowanie na następujących etapach:

1. Planowanie wydatkowania funduszy ze środków UE.

Zgodnie z dokumentem strategicznym *Europa 2020* jednym z pięciu celów UE jest efektywniejsza inwestycja we wtórne badania naukowe, co przyczyni się do osiągnięcia wzrostu gospodarczego, traktowanego jako priorytet dla całej UE. Wykonywanie przeglądów systematycznych na decyzyjnym etapie planowania wydatkowania funduszy UE, np. na potrzeby pisania strategii krajowych, regionalnych i lokalnych, zwiększa prawdopodobieństwo trafnych i zgodnych z rzeczywistym zapotrzebowaniem kierunków, do których powinno popłynąć wsparcie funduszowe w latach 2014–2020. Ponadto wykorzystanie danych z zastanych badań naukowych eliminuje nakłady finansowe na nowe badania, zwiększając tym samym efektywność polityk publicznych.

Przykład: Przegląd systematyczny dotyczący zapotrzebowania na konkretne kierunki edukacji zawodowej w regionie precyzyjniej obrazuje problem w porównaniu z pojedynczym badaniem i pozwala wyciągnąć wnioski pomocne w układaniu strategii rozwoju regionalnego w obszarze edukacji zawodowej.

2. Aplikowanie o środki z UE, realizacji i wdrażania projektów finansowanych ze środków UE.

Beneficjenci ubiegający się o sfinansowanie projektów ze środków UE, zarówno w trybie konkursowym, jak i systemowym, wypełniają we wniosku rubrykę dotyczącą uzasadnienia potrzeby realizacji projektu, w tym wskazują problem, na który odpowiedź

stanowi cel główny projektu, przedstawiają sytuację problemową, do której odnosi się projekt (z uwzględnieniem danych statystycznych odnoszących się do obszaru realizacji projektu), jak również powołują się na analizy statystyczne uzasadniające cel główny.

Zdecydowana większość ubiegających się o środki powołuje się na pojedyncze badanie naukowe bądź kilka doniesień badawczych, co nie zawsze koreluje z rzeczywistym zapotrzebowaniem na daną interwencję lub nawet ją wyklucza. Wcześniejsze przeprowadzenie przeglądu systematycznego pozwala na precyzyjne określenie sytuacji problemowej, sformułowanie celu głównego i celów szczegółowych, dostarczenie odpowiedniego wsparcia, a na dalszym etapie (realizacji projektu) skierowanie odpowiednich działań do grup rzeczywiście ich potrzebujących.

3. Ocena projektów zrealizowanych ze środków unijnych.

Ocena (ewaluacja) projektów finansowanych ze środków UE może dokonywać się np. tematycznie i dotyczyć porównywalnych interwencji.

Przykład: Przegląd systematyczny kilku badań prowadzonych w ramach projektów unijnych dotyczących edukacji ustawicznej dorosłych umożliwi ewaluację tego typu interwencji oraz wskaże potrzebę bądź zaniechanie przyszłych działań.

Powyższe etapy dotyczą dwóch grup:

1. Decydenci (na poziomie centralnym, np. instytucje zarządzające, pośredniczące i wdrażające, oraz na poziomie regionalnym i lokalnym, np. organy samorządowe) i środowisko naukowe (uczelnie/wydziały nauczające o funduszach UE, naukowcy).
2. Beneficjenci ubiegający się o dofinansowanie z UE swoich projektów.

Jak wykazano w publikacji: *W kierunku polityki opartej na dowodach* (Górnjak i Mazur, 2011), polska rzeczywistość potrzebuje zarówno specjalistów od przeprowadzania badań na potrzeby polityk publicznych, jak i urzędników świadomych wagi dowodów naukowych w procesie decyzyjnym w administracji publicznej. Dla środowiska urzędniczego w 2012 roku został opracowany poradnik zatytułowany: *Ewaluacja. Poradnik dla pracowników administracji publicznej*, przeprowadzający przez kolejne etapy ewaluacji przy planowaniu i wdrażaniu interwencji publicznych. Metoda

przeглядów systematycznych, jako jedna z możliwości metod badań naukowych, może spełniać podobne funkcje, tj. służyć jako narzędzie wsparcia.

Niezależnie od etapu, typu beneficjenta i formy wsparcia, metody przeglądu systematycznego pozwalają na efektywne wydawanie publicznych pieniędzy i skuteczne interwencje; ułatwiają także podejmowanie świadomych decyzji – na podstawie wiarygodnych i obiektywnych dowodów naukowych uzyskanych z przeglądu – zarówno na poziomie centralnym, regionalnym, jak i lokalnym. Tym samym wpisują się w założenia właściwego kształtowania i realizacji polityk publicznych opartych na „dobrym zapleczu dowodowym, które należy kształtować w wyniku systematycznych badań i analiz” (Górniak i Mazur, 2011, s. 10).

Niniejszy tekst podzielono na pięć części, z których każda zawiera trzy rozdziały. Pierwsza część poświęcona jest definicji przeglądu systematycznego i jego cechom dystynktywnym, a także zagadnieniom dotyczącym tej metody, takim jak mity na temat przeglądów systematycznych oraz zalety i ograniczenia PS. W kolejnych częściach i rozdziałach opisano najważniejsze etapy realizacji przeglądu systematycznego, takie jak: formułowanie pytania badawczego, dobór i selekcja badań, wydobycie danych z tekstów zakwalifikowanych do przeglądu, ocena krytyczna, synteza, określenie rodzajów obciążeń (stronniczości) oraz upowszechnianie wyników przeglądu.

W polskim obszarze edukacji brakuje konwencjonalnych przeglądów systematycznych, niekiedy utożsamianych z metaanalizą. Tymczasem przegląd systematyczny jest pojęciem szerszym – może, ale nie musi zawierać metaanalizę (stanowiącą jedynie metodę syntezy danych ilościowych). Eklektyczny PS (*mixed metod review*) łączy syntezy badań jakościowych (np. metaetnografia) z ilościowymi (np. metaanaliza). W dziedzinie edukacji jednak, w wielu przypadkach, przegląd systematyczny nie włącza badań opartych na statystyce, jeśli pytanie badawcze odnosi się do kwestii badanych za pomocą metod jakościowych lub innych (Gough, Olivier i Thomas, 2012). Tylko w niektórych PS potrzebna i uzasadniona jest metaanaliza, stąd dodatkową wartość publikacji stanowi dołączony *Suplement*, opisujący bardziej szczegółowo tę metodę agregowania badań ilościowych. Do tekstu dodano *Glosariusz*, przydatny w lekturze oryginalnych źródeł zamieszczonych w *Bibliografii*.

Przekazujemy niniejszą publikację z nadzieją, że przedstawiona w zarysie struktura przeglądu systematycznego przybliży polskim odbiorcom podstawowe zagadnienia związane z jego metodologią.

Definicja przeglądu systematycznego
Mity na temat przeglądów systematycznych
Zalety i ograniczenia metody

Definicja przeglądu systematycznego

*Jeśli widzę dalej niż inni, to dlatego,
że stałem na ramionach olbrzymów.*

Isaac Newton

Jeśli rzetelnie przeprowadzony przegląd systematyczny zintegruje z sukcesem wyniki z zestawu wszystkich rzetelnych metodologicznie badań, które są adekwatne do pytania badawczego, to czytelnik przeglądu faktycznie może odnieść wrażenie, że „wspiął się na ramiona olbrzymów” i ma najpełniejszy obraz tego, jak się rzeczy mają w zbadanym aspekcie.

Czym zatem jest badanie retrospektywne o nazwie „przegląd systematyczny” i jakie są jego właściwości? Definicję przeglądu można ująć lapidarnie: „Przegląd systematyczny to przegląd literatury naukowej z zastosowaniem jawnych, szczegółowo udokumentowanych i powtarzalnych metod” (Petticrew i Roberts, 2006, s. 283). Zważywszy jednak na złożoność samego procesu tworzenia przeglądu, podana definicja nie jest wyczerpująca. Więcej szczegółów podaje inny opis:

Przegląd systematyczny jest to przegląd podporządkowany jasno sprecyzowanemu pytaniu, z zastosowaniem jawnych metod identyfikacji, selekcji i krytycznej oceny istotnych badań oraz zbioru i analizy danych z badań zakwalifikowanych do tego przeglądu (Booth, Papaioannou i Sutton, 2012, s. 271) [tłum. – J.M.].

Powyzsza definicja rowniez jest niepełna – potrzebujemy szerszego ujęcia, które odsłoni cele PS:

Przeglądy systematyczne to przeglądy literatury naukowej z zastosowaniem ściśle określonego zestawu metod, w celu zredukowania systematycznych błędów (obciążeń) za pomocą zidentyfikowania, oceny i syntezy wyników wszystkich istotnych badań podstawowych (bez względu na schemat badania), aby uzyskać odpowiedź na konkretne pytanie badawcze (lub zestaw pytań badawczych). W procesie realizacji tego zadania określa się z góry i szczegółowo metody, podobnie jak w każdym badaniu socjologicznym. Pod tym względem przeglądy systematyczne różnią się od tradycyjnych, narracyjnych artykułów przeglądowych (Petticrew i Roberts, 2006, s. 9–10) [tłum. – J.M.].

Tradycyjne artykuły przeglądowe, jakkolwiek stanowią podsumowanie aktualnego stanu wiedzy na dany temat, przedstawiają jednak źródła subiektywnie wyselekcjonowane przez autorów w sposób niesystematyczny, niewyczerpujący i niejawnny, stąd też potencjalnie obciążający błędem tendencyjności wyniki oraz wnioski końcowe. Cechy dystynktywne narracyjnych artykułów przeglądowych oraz przeglądów systematycznych przedstawiono poniżej.

Tabela 1. Różnice metodologiczne między artykułami przeglądowymi a przeglądami systematycznymi (oprac. na podstawie: Boland i in., 2014, s. 13).

Artykuły przeglądowe	Przeglądy systematyczne
<p>Pytanie może, ale nie musi być jasno skonkretyzowane. Skoncentrowanie na ogólnym, szeroko zakreślonym temacie.</p>	<p>Pytanie jasno sformułowane i konkretne (aspektowe). Zawsze wymagane. Pytanie przeglądowe złożone z kilku konkretnych elementów (kryteriów) w celu uzyskania odpowiedzi w rezultacie przeglądu systematycznego.</p>
<p>Sporządzenie protokołu zazwyczaj niewymagane.</p>	<p>Sporządzenie protokołu rekomendowane lub niezbędne.</p>

Artykuły przeglądowe	Przeglądy systematyczne
<p>Brak jawnej i rygorystycznej metodologii. Metody niejawne i nieudokumentowane.</p>	<p>Postępowanie według jawnej i rygorystycznej metodologii – jawne i udokumentowane metody:</p> <ul style="list-style-type: none"> • wyszukiwania literatury; • kryteriów włączenia i wyłączenia badań; • oceny jakości (rzetelności); • syntezy wyników badań.
<p>Wyszukiwanie materiałów: brak ustalonych kryteriów wyszukiwania. Niekoniecznie wyczerpujące i rozległe. Zasadniczo polega tylko na danych z publikacji. Strategie wyszukiwania mogą bazować na doświadczeniu eksperta.</p>	<p>Wyszukiwanie materiałów: wyczerpujące i z zastosowaniem odpowiedniego bilansu tzw. czułości i precyzji. Prowadzone w elektronicznych bazach danych, poprzez wyszukiwanie ręczne z bibliografii istotnych prac naukowych i czasopism punktowanych oraz z doniesień naukowych i dokumentów. Niekiedy wyszukiwana jest tzw. szara literatura/prace naukowe. Wszechstronne, jawne i udokumentowane metody wyszukiwania.</p>
<p>Zdefiniowanie kryteriów włączenia i wyłączenia: niekonieczne; brak selekcji badań pod względem typu badania.</p>	<p>Zdefiniowanie kryteriów włączenia i wyłączenia: konieczne. Można ograniczyć wybór do określonego typu badań (np. włączyć tylko dane jakościowe i/lub dane z badań o wyższej jakości metodologicznej).</p>
<p>Selekcja tekstów na ogół wykonana przez jednego badacza, oparta na istotnych pracach naukowych i własnym doświadczeniu.</p>	<p>Systematyczne sprawdzanie i selekcja tytułów, abstraktów i tekstów, zazwyczaj sprawdzone przez innego badacza.</p>
<p>Niewymagana ocena jakości badań.</p>	<p>Wymagana ocena jakości badań.</p>

Artykuły przeglądowe	Przeglądy systematyczne
Ekstrakcja danych.	Ekstrakcja danych.
Nie ma jasno określonej metody syntezy .	Jasno określona metoda analizy i syntezy, np. metaanaliza, metaetnografia, synteza narracyjna lub jakościowa.
Termin realizacji: przeprowadzone relatywnie szybko.	Termin realizacji: czasochłonne ze względu na wymagany rygor.
Powtarzalność: niełatwo powtarzalne.	Powtarzalne wskutek jawnych metod.
Podatne na błędy systematyczne.	Minimalizowanie potencjalnych błędów systematycznych.
Wyniki i wnioski mogą wydawać się niejasne i podane w nieprzystępny sposób dla decydentów i innych odbiorców spoza środowiska badaczy.	Wyniki i wnioski są jasne i istotne dla odbiorców (odpowiedź na pytanie PS), podane w przystępnej formie dla użytkowników spoza środowiska naukowego (decydenci i in.).

Naukowcy prowadzący przeglądy systematyczne używają określonych, jawnych metod, mając na celu wykluczenie różnego rodzaju obciążeń (*biases*), a przez to wytworzenie bardziej solidnych wyników, które mogą być wykorzystane przez polityków w procesie podejmowania decyzji. Ten aspekt jest szczególnie ważny, ponieważ obciążenia mogą prowadzić do przeszacowania lub niedoszacowania efektów, a tym samym – do błędnych wniosków. Obciążenia te mogą być konsekwencją błędów popełnionych w analizowanych badaniach bądź w samym procesie przeglądu badań. Stąd zarówno staranność w dokumentowaniu podejmowanych działań, jak i przejrzyste kryteria doboru analizowanych źródeł dowodów.

Celem przeglądu systematycznego jest podsumowanie najlepszych dostępnych badań dających odpowiedź na konkretne, szczegółowo sformułowane pytanie

badawcze. Pytanie przeglądowe składa się z kilku konkretnych elementów – kryteriów, precyzujących cel przeglądu. Podsumowanie odbywa się za pomocą syntezy wyników badań włączonych do przeglądu systematycznego.

Przeгляд systematyczny charakteryzuje się przejrzystymi procedurami w celu znalezienia, oceny i syntezy wyników badań związanych z postawionym pytaniem przeglądowym. Procedury są zdefiniowane wcześniej, w celu zapewnienia przejrzystości i powtarzalności procesu przeglądu systematycznego. Praktyka ta ma również na celu zminimalizowanie stronniczości.

Badania podstawowe włączone do przeglądu są sprawdzane na podstawie spełniania kryteriów jakości metodologicznej, aby możliwe było łączenie ich wyników. Podsumowując, przegląd systematyczny badań podstawowych musi zawierać:

- jawne kryteria włączenia i wyłączenia badań;
- jawną strategię wyszukiwania doniesień badawczych;
- systematyczne kodowanie i analizę jakości włączonych badań;
- syntezę wyników badań ilościowych i/lub syntezę badań jakościowych (np. metaanaliza; metaetnografia lub oba rodzaje syntezy).

Przeگłady systematyczne w obszarze nauk społecznych, prowadzone w The Campbell Collaboration (C2), mają zazwyczaj zasięg międzynarodowy, odbywają się według standardów wypracowanych w C2, muszą także zawierać zapis systematycznego wyszukiwania niepublikowanych raportów i doniesień w celu uniknięcia stronniczości publikacji.

Przeگłady systematyczne są zróżnicowane, ich typologia wyznaczana jest przez: pytanie badawcze, założenia teoretyczne, głębię i zakres dociekań, schemat przeglądu oraz zastosowane metody. Podobnie jak badania podstawowe, badania wtórne w formie przeglądów systematycznych odzwierciedlają szeroki wybór kwestii badawczych i metod. Aby uzyskać pełniejszy i bardziej spójny obraz, niż przedstawiony w poszczególnym badaniu, a także zrozumieć lepiej problem, należy wziąć pod uwagę badania zajmujące się tym zagadnieniem w różnych ujęciach i kontekstach. Jeśli chcemy przewidzieć skutki konkretnej polityki lub interesują nas przyczyny problemu, potrzebujemy przeglądu podobnych badań i połączenia ich wyników, zapewniając dokładniejszy pomiar (Gough i in., 2013).

Mity na temat przeglądów systematycznych

Spotkać można szereg błędnych opinii na temat przeglądów systematycznych, stąd w niektórych publikacjach (np. Petticrew, 2001; Petticrew i Roberts, 2006; Littell, Corcoran i Pillai, 2008) pojawiła się potrzeba rozwiania pewnych mitów. Najbardziej upowszechnione, błędne przekonanie powstało w związku z faktem, że PS biorą swój początek w naukach medycznych.

Mit: Przeglądy systematyczne muszą zawsze zawierać randomizowane kliniczne badania kontrolne (randomized controlled trials - RCT), a tym samym badają jedynie skuteczność interwencji medycznych i muszą stosować tylko modele biomedyczne.

Jest to oczywiście nieporozumienie, gdyż przeglądy systematyczne na świecie są szeroko wykorzystywane jako pomoc w podejmowaniu decyzji na podstawie dowodów naukowych w obrębie polityk i praktyk w różnych obszarach życia społecznego. Jakkolwiek przegląd systematyczny powstał w dziedzinie opieki zdrowotnej, to wzrasta korzystanie z tej metody w dziedzinie nauk społecznych, zwłaszcza w psychologii i edukacji (zob. Campbell Collaboration, EPPI-Centre).

Poza tym, przeglądy systematyczne nie mają jednego preferowanego modelu badawczego, gdyż podejmują tak różnorodną tematykę, jak reklama, rolnictwo, archeologia, astronomia, biologia, zoologia, entomologia, chemia, kryminologia, ekologia, edukacja, prawo, psychologia. Metoda przeglądów systematycznych może odpowiadać na bardzo różne pytania badawcze, niekoniecznie odnoszące się jedynie do efektów interwencji. Mit ten ponadto odzwierciedla obawy niektórych badaczy, że synteza w przeglądzie systematycznym dotyczy jedynie badań RCT; ten rodzaj badań jest uznawany w nauce za „złoty standard” w badaniach zależności przyczynowo-skutkowych. W rzeczywistości w różnych przeglądach systematycznych stosuje się wiele rodzajów badań. Brytyjskie wytyczne do przeprowadzania przeglądów systematycznych uwzględniają np. badania jakościowe. Natomiast przeglądy dotyczące skuteczności interwencji nie ograniczają się wyłącznie do badań RCT, lecz mogą zawierać badania innego typu, w tym badania obserwacyjne i badania przypadków. Metodologia systematycznego przeglądu ma na celu m.in. ograniczenie (zminimalizowanie)

stronniczości, natomiast zakres uwzględnianych schematów badawczych należy do autorów przeglądu (recenzentów).

Wyjaśnienia wymagają również kolejne mity, zauważone przez Petticrew (2001), które przedstawiono poniżej.

Mit: Przeglądy systematyczne są takie same, jak narracyjne (tradycyjne), jedynie o większej objętości.

O cechach dystynktywnych obu rodzajów przeglądów powiedziano w rozdziale dotyczącym definicji PS. Warto jednak zaznaczyć, że przeglądy systematyczne, wskutek przemyślanego doboru i selekcji badań, mogą rozpatrywać mniejszą liczbę badań niż przeglądy tradycyjne po części dlatego, że przegląd systematyczny stosuje bardziej rygorystyczne kryteria włączenia badań do analizy. Poza tym przeglądy systematyczne wymagają pracy zespołowej, większego nakładu czasu i środków finansowych oraz ogromnej liczby doniesień naukowych w celu ich selekcji – stąd różnica jest nie tyle objętościowa, ile jakościowa.

Mit: Przeglądy systematyczne nie mają znaczenia w realnym świecie.

Od lat 1970. badacze prowadzą przeglądy systematyczne na temat skuteczności interwencji w kwestiach medycznych, politycznych i społecznych, takich jak: zapobieganie aktom wandalizmu, przestępczości, narkomanii, przemocy w rodzinie, przemocy wobec dzieci i in., co pokazuje, że przeglądy znajdują zastosowanie w wielu obszarach polityki społecznej. Prowadzone rzetelnie i odpowiedzialnie – dostarczają wiarygodnych dowodów na wsparcie określonych działań pomocowych.

Mit: Przeglądy systematyczne wymagają koniecznie syntezy statystycznej.

Ten mit pochodzi z niewiedzy na temat metod wykorzystywanych w przeglądach systematycznych. Statystyczna synteza wyników badań (metaanaliza) ma wiele zalet. W ostatnich latach zaobserwować można znaczny postęp w metodach statystycznych wykorzystywanych do tego celu. Niektóre przeglądy nie agregują jednak danych

ilościowych, lecz podsumowują narracyjnie rezultaty podstawowych badań jakościowych, opisując ich metody i wyniki (synteza narracyjna; metaetnografia). Inne przeglądy, które przyjęły podejście ilościowe, integrują dane, przekształcając je z każdego badania w skali pomiarowej i łącząc wyniki statystycznie (metaanaliza). Jest jednak wiele przeglądów systematycznych, w których nie należy stosować metaanalizy. Wskazówki na temat zasadności syntezy wyników badań poprzez metaanalizę znaleźć można w poświęconych tej metodzie podręcznikach (np. Cooper, 2010).

Mit: Przeglądy systematyczne muszą być wykonywane wyłącznie przez ekspertów.

W zespołach realizujących te badania wtórne specjaliści mogą być w mniejszości, tzn. oprócz wkładu naukowego przynajmniej dwóch badaczy, w konsultacjach uczestniczą także doradcy (konsultanci): praktycy zawodowi, konsumenci, przedstawiciele odpowiednich grup odbiorców.

Mit: Przeglądy systematyczne mogą być wykonywane bez specjalistów w zakresie informacji naukowej.

Z pozoru jest możliwe prowadzenie strategii wyszukiwania bez pomocy specjalistycznej, naukowcy jednak nie mają wystarczającego doświadczenia technicznego we wszechstronnym wyszukiwaniu informacji, rezultaty ich wyszukiwania mogą być mniej precyzyjne, o tzw. mniejszej czułości i precyzji wyników, mniej szczegółowe i wolniej osiąmane niż te wykonane przez specjalistów z zakresu informacji naukowej.

Wykonanie dobrego przeglądu systematycznego wymaga umiejętności w zakresie projektowania strategii wyszukiwania i organizowania źródeł, słowem – w zarządzaniu wiedzą. Specjaliści w wyszukiwaniu informacji naukowej muszą współpracować z badaczami, np. w opracowaniu listy słów kluczowych potrzebnych do całościowej strategii wyszukiwania w wielu odpowiednich źródłach, a także pracując jako organizatorzy i analitycy artykułów i dokumentów; muszą skutecznie zarządzać procesem wyszukiwania, sprawdzania, dokumentowania i archiwizacji materiałów naukowych.

Mit: Przeglądy systematyczne zastępują badania podstawowe.

Przeglądy nie wykluczają potrzeby dalszych badań podstawowych, np. w sytuacji tzw. pustego przeglądu (*empty review*), gdy synteza jest niemożliwa z powodu rozbieżności wyników, słabej jakości znalezionych badań lub braku badań kwalifikujących się do syntezy. Takie rezultaty też warto rejestrować i sporządzać abstrakty do baz danych, ponieważ wiedza o tym, że brakuje badań odpowiadających na daną kwestię, jest pomocna dla badaczy starających się o granty. W ten sposób PS są skutecznym sposobem identyfikacji badań, których brakuje. Mogą również być pomocne w uniknięciu niepotrzebnych nakładów na nowe badania podstawowe, gdy np. przeprowadzona metaanaliza wskazuje na skuteczność interwencji.

Zalety i ograniczenia metody

Zamiast chaotycznie przeszukiwać literaturę badawczą, dokonując bardziej lub mniej świadomie wyboru prac, które popierają nasze wcześniejsze przekonania, wybierzmy systematyczne podejście naukowe do procesu poszukiwania dowodów, uzyskując pewność, że otrzymane przez nas dowody są tak kompletne i reprezentatywne, jak to tylko możliwe (Ben Goldacre, za: Gough i in., 2013, s. 4) [tłum. – J.M.].

Powyższe słowa Bena Goldacre'a, badacza, lekarza oraz entuzjasty przeglądów systematycznych, akcentują ich największy atut, czyli uzyskanie najbardziej rzetelnych dowodów naukowych stanowiących odpowiedź na pytanie badawcze. Zaletą systematycznego podejścia do przeglądu jest zapobieganie osobistym preferencjom, („bardziej lub mniej świadomym”) upodobaniom czy uprzedzeniom autorów przeglądu. Kolejną zaletą przeglądów systematycznych (w odróżnieniu od tradycyjnych, niesystematycznych) jest fakt, że otrzymane dowody są „tak kompletne i reprezentatywne, jak to tylko możliwe”. Istotnie, w hierarchii dowodów naukowych rezultaty analizy i syntezy przeglądu systematycznego mają najwyższą pozycję.

Metoda naukowej syntezy badań jest potrzebna nie tylko decydom; służy także ulepszeniu rozwiązań oraz praktyk zawodowych w danej dziedzinie, np. edukacji.

Synteza tak solidnych dowodów przyczynia się do rozwoju wiedzy, pomaga odkryć, co wiemy, a czego nie wiemy, co się sprawdza, a co nie; prawidłowa synteza wyników daje najbardziej wiarygodne oszacowanie skuteczności danej interwencji i umożliwia znalezienie luk w wiedzy, które wyznaczają potrzebę nowych badań podstawowych.

Ponadto, rezultaty przeglądu systematycznego mogą wskazać, jak funkcjonuje wdrożenie danej polityki, programu, techniki lub interwencji (działania pomocowego) w różnych podgrupach i informować o potencjalnych szkodach, czy niebezpieczeństwach (Booth i in., 2012).

Niektóre syntezy mogą pokazać wady określonych działań politycznych lub oszacować koszty danej interwencji, a tym samym – jej opłacalność. Innym powodem syntezy wyników badań jest sprawdzenie ich stałości, przy czym warto podkreślić, że konflikt między oceniającymi tę stałość jest cenny, a niezgodność wniosków może pokazać, w jakim kontekście (warunkach, miejscu) dana polityka społeczna może się sprawdzić, w jakich okolicznościach dany program edukacyjny może przynieść najlepsze rezultaty czy też – w naukach medycznych – jaka dawka leku jest najbardziej skuteczna.

Rzetelnie przeprowadzona synteza wskazuje słabe strony dowodów i potrzebę nowych badań, lecz czasami może odsłonić niepożądany rezultat, stanowiący wyzwanie dla ustalonych, mocno podtrzymywanych opinii. Nawet wtedy rozczarowujący wynik syntezy ma zaletę – jest nią możliwość wzięcia pod uwagę innych czynników.

Atuty przeglądowej syntezy dowodów polegają na tym, że stanowią cenną pomoc w procesie podejmowania decyzji. Do ogólnych atutów przeglądu systematycznego niewątpliwie należą:

- koncentrowanie się na konkretnych, porównywalnych interwencjach;
- uzyskanie bardziej wiarygodnych informacji na temat wpływu i przyczynowości dzięki bardziej rygorystycznemu projektowi badania wtórnego;
- najbardziej solidna podstawa do uogólnienia w porównaniu z wnioskami z pojedynczego badania (lub ekspertyzy);
- możliwość zbadania korelacji pomiędzy cechami poszczególnego programu a jego efektami.

Przeгляд systematyczny, jak każda metoda, ma ograniczenia. Przede wszystkim jest żmudnym, czasochłonnym procesem – potencjalnie może trwać zbyt długo i nie dostarczyć decydom dowodów w oczekiwanym przez nich terminie. Przeglądy systematyczne wymagają fachowej oceny krytycznej i bardzo dobrej znajomości metodologii badań społecznych oraz innych kompetencji ważnych w analizowaniu dowodów.

Na etapie syntezy rekomenduje się łączenie jednorodnych badań (homogenicznych), zdarzają się jednak przeglądy integrujące badania heterogeniczne i czasami, nawet pomimo wysokiego poziomu metodologicznego, nie dają odpowiedzi na postawione pytanie.

Podstawowym problemem jest dostępność badań. W praktyce, przeglądy są ograniczone do prac napisanych w określonym języku lub językach, co w dużej mierze wyklucza badania nieopisane w literaturze wydawanej w analizowanych językach. Współczesne nauki społeczne są zdominowane przez język angielski, co powoduje między innymi nadreprezentowanie badań przeprowadzonych w krajach anglosaskich. Prowadzi to do problemów z uogólnieniem wyników (trafności zewnętrznej) na inne populacje, interwencje czy konteksty społeczno-instytucjonalne – inne niż te, które były objęte analizowanymi badaniami.

Kolejnym ograniczeniem jest jakość dostępnych badań. Wiele z nich opisanych w publikacjach naukowych może nie spełniać kryteriów jakościowych lub brakuje w nich podstawowych informacji niezbędnych do opisanania badania. Problem ten ma swoją specyfikę w Polsce i w różnym stopniu dotyczy poszczególnych dyscyplin naukowych, ale jest obecny także w innych krajach. Może się okazać, że badań wstępnie kwalifikujących się do danego przeglądu systematycznego jest zbyt mało, aby w pełni wykorzystać potencjał metody PS. Z kolei, uwzględnienie wielu słabych badań może prowadzić do słabo uzasadnionych wniosków (*garbage in garbage out*).

Bardziej uniwersalnym problemem są obciążenia wynikające z selektywnego publikowania wyników badań (*publication bias*) (zob. Rothstein i in. 2005). Artykuły naukowe zwykle nie odzwierciedlają wiernie wszystkich przeprowadzonych badań i analiz. Autorzy badań, które nie przyniosły spodziewanych lub interesujących wyników, nie zawsze opisują uzyskane wyniki w artykułach. W analizach tego problemu zwraca się też uwagę na kwestię przywiązywania przez naukowców zbyt dużej wagi do istotności statystycznej uzyskanych wyników. Także redakcje czasopism

naukowych podchodzą selektywnie do nadsyłanych prac. Prowadzi to zwykle do przeszacowania uzyskanych efektów statystycznych oraz nadreprezentowania jednych badań i niedoreprezentowania innych.

Problemy wiążą się także z konceptualizacją przeglądu. Powinien on bowiem nawiązywać do teorii danego zjawiska, będącej zwykle punktem wyjścia konkretnych badań. W przypadku łączenia różnych badań problematyczna może być interpretacja wyników, zwłaszcza w przypadku łączenia bardzo różnych badań, w których są badane populacje, różnorodne interwencje czy różne sposoby pomiaru efektów. Z tego względu przegląd systematyczny powinien mieć ściśle określony cel i mocne założenia teoretyczne (Littell i in., 2008).

<p style="text-align: center;">Definiowanie pytania przeglądowego Powołanie zespołu badaczy i ekspertów Opracowanie protokołu</p>
--

Definiowanie pytania przeglądowego

Przed rozpoczęciem pracy nad definiowaniem pytania należy sprawdzić, kto i w jaki sposób będzie wykorzystywał wyniki przeglądu oraz czy temat przeglądu nie został już zarejestrowany w międzynarodowych bazach rejestracyjnych, np. w Międzynarodowym Prospektywnym Rejestrze Przeglądów Systematycznych – PROSPERO – globalnej inicjatywie prowadzonej przez ośrodek upowszechniania przeglądów systematycznych z uniwersytetu w Yorku: CRD, University of York, <http://www.crd.york.ac.uk/prospero>, lub w rejestrze The Campbell Collaboration (C2), http://www.campbellcollaboration.org/systematic_reviews/index.php?page=2.

Pytanie badawcze wyznacza zakres przeglądu, dlatego proces definiowania stanowi kluczowy etap. Najlepiej zdefiniowane pytania powstają w wyniku dyskusji z potencjalnymi użytkownikami przeglądów. Konsultacje z interesariuszami powinny być prowadzone na wczesnym etapie, aby poznać ich potrzeby informacyjne. Na przykład, rodziców może interesować przegląd na tematy: Jakiego rodzaju opieki przedszkolnej łączą się z najlepszymi wynikami dzieci? lub: Jak opieka przedszkolna wpływa na życie rodziców i dzieci? (Petticrew i Roberts, 2006).

Jeśli przegląd dotyczy skuteczności interwencji, pytanie można opracować, stosując model **PICO** – opracowany w dziedzinie medycyny na potrzeby pytań klinicznych o skuteczność – ułatwiający rozważenie następujących elementów pytania przedstawionych w tabeli poniżej.

Tabela 2. Opis modelu **PICO**, opracowanie własne.

Population (Populacja) – cechy badanej grupy, terminy z nią powiązane, wiek, płeć, status ekonomiczno-społeczny, wykształcenie etc.

Intervention (Interwencja) – działanie pomocowe, programy edukacyjne, terapeutyczne, społeczne, reformowanie edukacji, zmiany w określonym celu poprawy sytuacji/danego fenomenu.

Comparison (Porównanie) – czynnik porównawczy; porównanie skuteczności i kosztów, sposobów rozwiązania problemu; w badaniach obserwacyjnych grupa porównawcza.

Outcomes (Wyniki) – rezultat interwencji; co chcemy zmierzyć, sposób pomiaru.

Dla odbiorców przeglądów systematycznych ważna jest często nie tylko informacja o skuteczności danej interwencji. Nie wystarczy powiedzieć, że dana polityka społeczna czy interwencja sprawdziła się lub nie – odbiorcy chcą poznać szerszy kontekst: jak dana interwencja została zrealizowana; co wspomagało lub utrudniło jej wpływ, jaki proces wdrażania przeprowadzono, i jakie czynniki mogły wpłynąć na jej sukces lub porażkę. W związku z tym, Petticrew i Roberts (2006) zaproponowali dla przeglądów w dziedzinie nauk społecznych model **PICOC**, uwzględniający dany kontekst pytania (lokalizacja, kraj/e, obszar/y).

Tabela 3. Przykład zastosowania modelu PICOC (oprac. na podstawie: Booth i in., 2012, s. 56).

Population (próba) = dzieci i młodzież w wieku szkolnym 5–16 lat. Chłopcy i dziewczęta. Wszystkie etapy edukacji obowiązkowej.

Intervention/Exposure = picie alkoholu wśród niepełnoletnich (w przypadku braku interwencji można w to miejsce modelu dać czynnik narażający na szkodliwe działanie).

Comparison = alternatywa; niepijący alkoholu.

Outcomes = wpływ społeczny lub ekonomiczny. Skutki obejmują m.in. perspektywy ekonomiczne (poziom dochodów), rozwój emocjonalny, osiągnięcia edukacyjne i zawodowe, antyspołeczne lub okrutne zachowanie, wykorzystanie seksualne, choroby przenoszone drogą płciową, ciężę wśród nastolatków, wzrost kosztów opieki zdrowotnej (włącznie z nagłymi wypadkami).

Context = zasadniczo UK, ale możliwe jest włączenie międzynarodowych badań, jeśli wyniki można zastosować do lokalnych warunków.

Powyższy przykład pokazuje, że czas spędzony na określaniu zakresu pytania badawczego został dobrze wykorzystany; ponadto, przy okazji można wziąć pod uwagę potencjalne terminy wyszukiwania i typy badań.

W dziedzinie badań edukacyjnych pytania mogą dotyczyć: skuteczności interwencji, wpływu niektórych czynników ryzyka, określenia czynników ryzyka, powiązań między zjawiskami socjologicznymi a wynikami, prognoz, kwestii ekonomicznych i diagnostyki pewnych uwarunkowań. Rozpoczynając prace nad przeglądem, powinniśmy mieć rozeznanie, czy znajdziemy odpowiedni materiał dowodowy (dane zastane), umożliwiający analizę wyników. W pytanie przeglądowe (badawcze) wpisane są zarazem kryteria wyszukiwania, które wyznaczają postępowanie badawcze.

Tabela 4. Model PICOSS (oprac. na podstawie: Boland i in., 2014, s. 28).

Jakie są skutki interwencji terapeutycznych: głośnego czytania i czytania w grupach w celu poprawy zdrowia i samopoczucia pacjentów z zaburzeniami neurologicznymi w warunkach klinicznych i opieki długoterminowej?

Kluczowe komponenty pytania postawionego według modelu PICO+SS:

Population (dorośli oraz dzieci z każdym zaburzeniem neurologicznym).

Intervention (czytanie indywidualne; czytanie na głos; czytanie w grupach).

Comparator (porównanie wyznaczonych obu interwencji do siebie, do placebo lub braku interwencji).

Outcomes (jakikolwiek pozytywny dla zdrowia rezultat, subiektywny, mierzony klinicznie; analiza lub metody ilościowe zbioru danych).

Setting (warunki szpitalne, ośrodki rehabilitacji, domy mieszkalne, domy opieki, hospicja).

Study design (wszystkie typy badań).

Modele pytań, pochodzące od **PICO**:

PICOC – (PICO + Context)

PICOS (PICO + Study design – typ badania)

PICOS (PICO + Setting – umiejscowienie)

PICOT (PICO +Time frame – zakres czasowy)

PICo (Population - Intervention- Context) – pomocne w pytaniach jakościowych

PICOTT (Type of Question; Type of Study Design – określenie typów pytania i badania)

W definiowaniu pytania i zarazem zakresu przeglądu można także posłużyć się innymi ramami koncepcyjnymi przedstawionymi poniżej.

PIPOH – opracowano dla wyszukiwania wytycznych w onkologii (druga litera P oznacza lekarzy, a H to określenie opieki zdrowotnej oraz kontekst, w którym te wytyczne będą zastosowane).

PROPHET (Problem, Fenomen zainteresowania, Czas) – model pomocny szczególnie w badaniach jakościowych, podobnie jak SPICE.

SPICE – rama koncepcyjna, określająca metodę formułowania lub precyzowania pytania przeglądowego według następujących elementów: *Setting* (Umiejscowienie, warunki); *Perspective* (Perspektywa, ujęcie); *Phenomenon of Interest* (Fenomen zainteresowania); *Comparison* (Fenomen porównawczy) i *Evaluation* (Ewaluacja).

ECLIPSE – rama koncepcyjna opracowana w celu formułowania pytań w zakresie polityki zdrowotnej i zarządzania: *Expectation* (przyczyna zapotrzebowania na informację); *Client Group* (grupa docelowa); *Location* (lokalizacja usługi); *Impact* (Jaka zmiana usługi jest oceniana? Co stanowiłoby sukces? – *Impact* podobny jest do komponentu *Outcomes* w modelu PICO); *Professionals* (grupa zawodowa); SE – *Service* (rodzaj usług brany pod uwagę) (Wildridge i Bell, 2002).

W dziedzinie zarządzania, odpowiednikiem PICO jest CIMO: *Context* (co będzie badane?); *Intervention* (skutki tego, co jest badane); *Mechanisms* (mechanizmy wyjaśniające relacje pomiędzy Interventions a Outcomes); *Outcomes* (skutki interwencji; sposób pomiaru).

Przykład pytania z użyciem CIMO:

Pod jakimi warunkami (C) styl przywództwa (I) wpływa na wydajność zespołów projektowych (O) i jakie mechanizmy działają pod wpływem stylu przywództwa (M) na wyniki zespołu projektowego (O)? (Booth i in., 2012, s. 57).

Rodzaj pytania badawczego (może składać się z kilku pytań) zależy będzie od celu przeglądu i przedmiotu zainteresowania. Spójrzmy na schematy pytań badawczych w przeglądach systematycznych (Booth i in., 2012):

Pytanie o skuteczność:

Jaki wpływ interwencja X, w porównaniu z interwencją Y, wywiera na wynik Z?

Jaka jest względna analiza kosztów i korzyści x w porównaniu z y?

Pytanie o metodologię:

Jakie metody badawcze były wcześniej zastosowane do badania fenomenu X? Jakie są zalety i wady tych metod?

Pytanie koncepcyjne:

Jak jest identyfikowany i definiowany fenomen lub dana zmienna X? Jakimi teoriami się posłużono, aby wyjaśnić fenomen X? Która teoria najlepiej je tłumaczy w świetle wyników badań empirycznych? Jakie są główne nierozwiązane kontrowersje? Jakie są epistemologiczne i ontologiczne podstawy tej dyscypliny?

Konkretyzowanie pytania badawczego łączy się z postawieniem kilku powiązanych pytań, np. do pytania o skuteczność interwencji dodamy zestaw szczegółowych pytań związanych ze specyfiką jej wdrażania. Tego rodzaju pytanie wymagać będzie przeglądu zróżnicowanych metodologicznie badań o różnym schemacie, np.: eksperymentalne i nieeksperymentalne (RCT, kontrolowane), obserwacyjne (podłużne, quasi-eksperymentalne).

Powołanie zespołu badaczy i ekspertów

Podobnie jak każde badania, przeglądy systematyczne potrzebują planowania, określającego nie tylko zakres tematyczny i metody, lecz również kwalifikacje zespołu, sposoby współpracy czy system zarządzania informacjami.

Zgodnie z wybranymi na ten temat źródłami (np. Gough i in., 2012; Petticrew i Roberts, 2006), przeglądy systematyczne lepiej jest prowadzić w zespołach, np. dwóch badaczy może niezależnie kodować materiał, stanowiąc minimum pozwalające obliczyć współczynnik zgodności. Poza tym przeglądy systematyczne wymagają połączenia konkretnej wiedzy (nieraz z różnych dziedzin) z takimi metodami, jak umiejętności wyszukiwania materiału, analizy i syntezy danych, pisanie tekstów (Gough i in., 2012). Na poszczególnych etapach realizacji przeglądu osoby wchodzące w skład zespołu pełnią różne funkcje i mają różny zakres obowiązków. Współpraca zespołowa ułatwia ukończenie wysokiej jakości przeglądu w terminie.

Wiedza i umiejętności metodologiczne zespołu mogą być uzupełnione przez grupę doradczą (Gough i in., 2012), której członkowie (potencjalni użytkownicy) zapewnią wiedzę wynikającą z ich doświadczenia na różnych polach: badawczym, metodycznym, politycznym lub zawodowym; dzięki grupie doradczej można wziąć pod uwagę różne perspektywy i aspekty, np. na etapie precyzowania tematu, konstruowania pytania badawczego, czy też tworzenia raportu. Organizowanie wsparcia w formie konsultacji z przedstawicielami różnorodnych środowisk jest kluczem do udanego przeglądu, np. do zespołu przeglądu systematycznego interwencji dotyczącej edukacji można włączyć reprezentantów różnych środowisk: nauczyciela, który korzysta z interwencji; menedżera usług edukacyjnych; dyrektora placówki edukacyjnej; badacza, który wcześniej badał lub oceniał daną interwencję, statystyka, ekonomistę, urzędnika edukacyjnego etc. Grupa konsultantów może również mieć wkład w opracowanie protokołu z przeglądu.

Opracowanie protokołu

Po opracowaniu głównego pytania badawczego oraz zarejestrowaniu tematu przeglądu w odpowiedniej bazie rejestrującej tytuły przeglądów systematycznych (międzynarodowe bazy rejestracyjne: C2 i PROSPERO) należy rozpocząć pisanie protokołu, czyli szczegółowego planu przeglądu systematycznego. Protokół przeglądu, opracowany na wstępie projektu, na dalszych etapach ulega weryfikacji i modyfikacji, przy czym każda zmiana musi być udokumentowana i uzasadniona w protokole.

Protokół jest niezbędną częścią pracy nad przeglądem systematycznym, którego zasadnicze komponenty są następujące (Booth i in., 2012):

- tło;
- pytanie przeglądowe;
- strategia wyszukiwania literatury (włącznie z terminami i źródłami);
- kryteria selekcji badań i procedury;
- listy kontrolne do oceny jakości badań i metody;
- strategia wydobycia danych;
- synteza wydobytych wyników badań;
- harmonogram projektu.

W ośrodku badawczym The Campbell Collaboration opracowano godny polecenia standardowy format protokołu dla przeglądów systematycznych i metaanaliz dotyczących skutków interwencji. Najważniejsze elementy składowe Protokołu C2 są następujące:

- strona tytułowa z adresem kontaktowym autorów;
- wstęp (tło i określenie celów);
- metody:
 - kryteria kwalifikowalności badań,
 - metody wyszukiwania (bazy danych, lista słów kluczowych, szara literatura),
 - ekstrakcja danych (arkusze kodowania, ocena jakości, plany kontroli wiarygodności),
 - analiza danych oraz synteza (metaanaliza, analiza podgrupy i moderatora, stroniczość publikacji, analiza czułości etc.);

-
- podziękowania i oświadczenie braku konfliktu interesów;
 - tabele i wykresy odpowiednie dla tła lub metod;
 - bibliografia.

Istotną kwestią na tym etapie jest określenie planowanych analiz, które powinny zostać zaprojektowane z góry i opisane w protokole. Wiąże się to z określeniem, jaki model metaanalizy i jakie metody najprawdopodobniej będą odpowiednie oraz jakie podgrupy i moderatory (*moderators*) zostaną uwzględnione. Procedura oceny środowiskowej (peer-review) może być pomocna w tym zadaniu (Littell i in., 2008).

Ponadto w protokole można ująć informacje: o języku publikacji włączonych do analizy, cezurze czasowej wyszukiwanych źródeł, liczbie specjalistów dla każdego etapu przeglądu oraz sposobach rozstrzygania rozbieżności w niezależnych ocenach autorów przeglądu.

W toku prac nad przeglądem protokołów może ulegać modyfikacjom, które również należy uzasadnić i udokumentować. Dobrą praktyką jest sprawdzenie i ocena protokołu przez osoby, mające wiedzę na dany temat, np. mogą to być eksperci, ale również przewidziani użytkownicy.

<p style="text-align: center;">Strategia wyszukiwania literatury Selekcja źródeł – faza wstępna Selekcja tekstów według kryteriów włączenia i wyłączenia</p>

Strategia wyszukiwania literatury

Po opracowaniu głównego pytania przeglądowego i konsultacji z utworzoną grupą doradczą, wiemy, jakiego rodzaju badania będą adekwatne do treści postawionego pytania. Równie kluczowy dla przeglądu jest następny krok – opracowanie strategii wyszukiwania badań zawartych w doniesieniach naukowych.

Na tym etapie, autorzy przeglądu powinni opracować szczegółowy opis proponowanej strategii wyszukiwania. Najpierw należy przygotować kompletną listę źródeł. Zazwyczaj są to pełnotekstowe i bibliograficzne bazy danych, w tym baza danych Current Contents (regularnie aktualizowana baza firmy Thomson Reuters, zawierająca dostęp do spisów treści ponad 10 000 wiodących światowych periodyków naukowych), biblioteki cyfrowe, strony internetowe odpowiednich instytucji, periodyki w elektronicznej i tradycyjnej postaci, nieopublikowane jeszcze materiały oraz tzw. szara literatura. Oprócz tego należy również uwzględnić wyszukiwanie ręczne i kontakty z ekspertami. Wartościowe i aktualne doniesienia znajdują się w następujących rodzajach źródeł (Niedźwiedzka, 1998):

- czasopisma przeglądowe (zamieszczające *reviews* lub *systematic reviews*, bazujące na wysokiej jakości badaniach naukowych);
- specjalistyczne, recenzowane czasopisma w tradycyjnej i elektronicznej postaci;
- tzw. szara literatura: nieopublikowane lub publikowane dla wąskiego kręgu odbiorców doniesienia, wewnętrzne publikacje akademickie instytucji naukowych i ośrodków badawczych, instytucji państwowych;
- specjalistyczne, pełnotekstowe bazy danych;
- publikacje udostępniane na stronach WWW odpowiednich instytucji i serwisów internetowych.

Realizatorzy przeglądu muszą udokumentować lata objęte wyszukiwaniem i słowa kluczowe, zastosowane w wyszukiwaniu baz danych i bibliografii, a także datę rozpoczęcia i zakończenia procesu wyszukiwania. Informacje na temat źródeł, słów kluczowych i strategii poszukiwań oraz cezurę czasowej dotyczące wyszukiwanej literatury są kluczowymi elementami metod wyszukiwania.

O ile jest to możliwe, w strategii wyszukiwania literatury należy przyjąć perspektywę porównawczą. Dowody z przeglądu systematycznego nie powinny być ograniczone narodowością badaczy i uczestników badania lub językiem, chyba że jest to uzasadnione celami przeglądu.

Najważniejsze źródła wyszukiwania badań dla anglojęzycznych przeglądów systematycznych to:

- Cochrane Central Register of Controlled Trials (via Cochrane Library);
- Cochrane Database of Systematic Reviews (via Cochrane Library);
- Campbell Collaboration Library;
- inne bazy danych: ERIC, Scopus, PsycInfo, CERUK plus, EPPI-Centre Evidence Library;
- czasopisma naukowe online: „Systematic Reviews”:
www.systematicreviewsjournal.com;
- Google Scholar.

Do najważniejszych **polskich** źródeł informujących o badaniach edukacyjnych należą:

- Baza Informacji o Badaniach Edukacyjnych (BIBE), powstała przy Instytucie Badań Edukacyjnych, rejestrująca doniesienia naukowe zazwyczaj z polskich czasopism punktowanych, obejmująca zakres czasowy 1989 – obecnie;
- Repozytorium danych z badań prowadzonych w IBE 2010 – obecnie (archiwum danych zastanych);
- Baza Projektów POKL;
- Centrum Upowszechniania Innowacji w Edukacji (CUIwE);
- Archiwum Danych Społecznych (ADS).

W celu zmniejszenia efektu stroniczości w wynikach, przegląd systematyczny integruje wyniki **wszystkich** istotnych doniesień naukowych. Zatem wyszukiwanie powinno być wszechstronne i kompletne, a jednocześnie zogniskowane na kryteriach kwalifikowalności do danego przeglądu. Kryteria te wpisane są w pytanie badawcze, stanowiące siłę napędową przeglądu. Proces wyszukiwania można podzielić na następujące etapy (Booth i in., 2012):

1. **określenie zakresu wyszukiwania** i wyszukiwanie wstępne: ustalenie baz danych, określenie, które bazy danych będą włączone do pełnego/szczegółowego wyszukiwania, rozpoznanie słów kluczowych odpowiednich dla każdej ustalonej bazy danych, opracowanie i dokumentacja strategii wyszukiwania;
2. **wyszukiwanie w bazach danych**: przeszukanie wszystkich istotnych baz danych według ustalonej na poprzednim etapie listy słów kluczowych i według kluczowych zasad wyszukiwania: za pomocą swobodnych terminów (tytuły i abstrakty) oraz narzędzi, deskryptorów, operatorów Boole'a (zawężenia, rozszerzenia, ograniczenia); wyszukiwanie tzw. szarej literatury, niepublikowanych tekstów; stosowanie filtra metodologicznego (określone rodzaje badań); dokumentacja wszelkich zmian wyszukiwania;
3. **wyszukiwanie poprzez bibliografie**: wyszukiwanie dodatkowych artykułów/badań według list bibliograficznych, kluczowych odniesień; indeksów cytowani (indeks SSCI); wyszukiwanie ręczne/przeoglądanie istotnych czasopism papierowych;
4. **weryfikacja**: konsultacja z ekspertami w celu ustalenia, czy wszystkie istotne dla przeglądu doniesienia zostały pobrane; powtórzenie wyszukiwania, jeśli to konieczne;
5. **dokumentacja** źródeł wyszukiwania, zastosowanej strategii i metod wyszukiwania, liczby znalezionych prac badawczych w każdym ze źródeł wyszukiwania.

Selekcja źródeł – faza wstępna

Przeszukiwanie literatury (zwłaszcza anglojęzycznej) na tym etapie przyniesie prawdopodobnie setki lub tysiące odniesień (często ze streszczeniami), które muszą zostać zredukowane do niezbędnych materiałów, poddanych bardziej szczegółowej selekcji w następnym etapie.

W procesie sprawdzania tytułów i abstraktów oraz (w następnej fazie wyszukiwania) całych tekstów doniesień uczestniczą przynajmniej dwie osoby, zwracające uwagę na:

1. treść pytania przeglądu;
2. schemat badawczy;
3. populację;
4. interwencję;
5. badane wyniki.

Rodzaj poszukiwanych informacji będzie zależeć zarówno od pytania przeglądu, jak i kryteriów włączenia badań do przeglądu.

Strategia hodowli perły (*pearl growing*)

Jedną z technik wykorzystywanych we wstępnym etapie wyszukiwania jest tzw. hodowla perły. Polega ona na identyfikacji najbardziej istotnego (w odniesieniu do pytania badawczego) artykułu naukowego, czyli tzw. perły, skąd pobiera się kluczowe słowa: zarówno terminy swobodne, słowa autorskie (np. z tytułu), jak i deskryptory, na których może później bazować wyszukiwanie. Jeśli nie mamy takiego artykułu, wówczas prowadzimy wyszukiwanie tytułów, aby ustalić kluczowe odniesienia; możemy tę procedurę powtarzać aż do uzyskania pełnej puli kluczowych terminów.

Tabela 5. Przykład techniki „rosnącej perły” (oprac. na podstawie schematu: Booth i in., 2012, s. 73).

Pytanie badawcze: *What is the psychological well-being of individuals with high debt?* (Jakie samopoczucie psychiczne mają osoby z wysokim zadłużeniem?).

PERŁA: Brown, S., Taylor, K., Price S.W. (2005). *Debt and distress: Evaluating the psychological cost of credit*

Właściwości perły: a) terminy swobodne: *debt, credit* b) deskryptory w bazie danych PsycInfo: *distress*.

Hasła swobodne uzyskane z artykułu-perły: *dlug i kredyt*. Ponadto, ten artykuł jest indeksowany w bazie danych PsychInfo pod hasłem *distress* (rozpacz; zmartwienie). Uzyskane terminy można zastosować do przypuszczalnie odpowiednich artykułów powiązanych tematycznie z pytaniem badawczym. Wyszukiwanie z użyciem perły przyniesie jakąś liczbę istotnych artykułów. Procedurę należy powtarzać, stosując dodatkowe nowe słowa kluczowe, dopóki nie otrzymamy nowych, istotnych terminów.

Metody wyszukiwania za pomocą: a) swobodnego tekstu, b) słownika kontrolnego, czyli deskryptorów tezaury, mają swoją specyfikę. Jedną z wad wyszukiwania swobodno-tekstowego są różnice w pisowni anglojęzycznych terminów (brytyjskich i amerykańskich) oraz w terminologii (np. termin „nastolatek”: *young person, adolescent, teenager, youth, child, children* etc.).

W pewnym stopniu pomocne jest korzystanie ze skrótów w wielu bazach danych (* lub \$) lub symboli wieloznaczności (?). Na przykład wpisany do wyszukiwarki termin *teenage** przyniesie rekordy zawierające hasła: *teenage, teenager, teenagers*.

Innym podejściem jest wykorzystanie tezaury (słownika kontrolnego). Wyszukiwanie tezaursowe (według deskryptorów, czyli haseł przedmiotowych) obejmuje listę terminów do indeksowania baz danych, utworzonych przez specjalistów w zakresie informacji naukowej. Niektóre bazy danych zawierające tezaury to: ASSIA, LISA, ERIC, Medline, Cochrane Library (dwie ostatnie zawierają tezaury MeSH; w Polsce baza zawierająca doniesienia naukowe w dziedzinie edukacji – BIBE, zawiera Tezaurus Europejskich Systemów Edukacji (TESE). Dany tezaurus może zawierać poszukiwany przez czytelnika termin; w sytuacji, gdy brak danego terminu, można przeszukać tezaurus i wybrać najbardziej zbliżone hasła.

Kolejnym narzędziem strategii wyszukiwania online jest zastosowanie operatorów Boole'a: AND (zawężenie wyszukiwania), OR (rozszerzenie wyszukiwania), NOT (zawęża wyszukiwanie poprzez wykluczenie nieistotnego terminu). Ponadto, większość baz danych ma przydatne funkcje ograniczania wyszukiwania do określonych dat (cezura czasowa), rodzaju lub języka publikacji.

Do przeglądu systematycznego wskazane jest włączenie także istotnych tekstów z kręgu tzw. szarej literatury (*grey literature*) w celu zredukowania stroniczości publikacji. W dziedzinie nauk społecznych potrzeba włączenia szarej literatury jest nawet większa, ponieważ wiele dowodów jest zamieszczonych w publikacjach organizacji pozarządowych.

Ważną częścią strategii jest sprawdzanie bibliografii i list odniesień oraz wyszukiwanie cytowań. W dziedzinie nauk społecznych powstały bazy danych ułatwiające wyszukiwanie cytowań, np. Social Science Citation Index (SSCI) – technika ta służy znalezieniu badań brakujących lub przeoczonych w wyszukiwaniu bibliograficznym. Można stosować wyszukiwanie według autorów, co jest przydatne, ale może skutkować stroniczością w doborze odniesień i argumentacji. Ponadto, zarówno kontakty bezpośrednie z ekspertami, jak i „listy autorów” zajmujących się danym zagadnieniem – *Listserve* (np. *Child Maltreatment Researchers List*) – wykorzystywane są w strategii wyszukiwania na potrzeby przeglądów systematycznych i metaanaliz (Littell i in., 2008).

Kolejny etap strategii wyszukiwania to weryfikacja, polegająca na konsultacji z ekspertami w celu ustalenia kompletnej listy publikacji i upewnienia się, czy wszystkie istotne badania zostały wyszukane.

Wskazane jest dokumentowanie procesu wyszukiwania na bieżąco, przydatne w napisaniu sprawozdania nt. metodologii wyszukiwania. Powinno ono zawierać: źródła wyszukiwania, zastosowane strategie, liczbę odniesień dla każdego źródła/metody i daty wyszukiwania. Pomocna jest także lista kontrolna PRISMA zawierająca dwa punkty dotyczące wyszukiwania.

Selekcja tekstów według kryteriów włączenia i wyłączenia

Na tym etapie przeprowadzana jest selekcja tekstów doniesień badawczych. Na podstawie sporządzonej listy tytułów, otrzymanych w fazie wstępnej wyszukiwania, autorzy przeglądu wybierają te doniesienia, których pełne teksty będą sprawdzane pod kątem włączenia do analizy i syntezy.

Na każdym etapie selekcji zachowuje się szczegółowy zapis liczby badań włączonych do przeglądu oraz odrzuconych (w tym dubletów). Niektóre z badań można od razu wykluczyć na podstawie abstraktu, ale pozostałe mogą wymagać dotarcia do pełnych tekstów w celu ustalenia, czy spełniają one przeglądowe kryteria włączenia. Wyselekcjonowane teksty naukowe poddane są następnie procesowi wydobycia danych (*data extraction*).

Proces włączenia badań (materiału naukowego) do przeglądu i podejmowanie decyzji dotyczących kodowania odbywa się z udziałem przynajmniej dwóch badaczy – autorów przeglądu, którzy pracują niezależnie, oceniając badania i porównując wyniki. Dodatkowo, jak to się odbywa w C2, przeglądy mogą zostać poddane recenzji oraz redakcji naukowej (*peer-review* oraz *editorial review*).

Ekstrakcja danych z badań włączonych do przeglądu
Ocena krytyczna
Synteza wyników badań

Ekstrakcja danych z badań włączonych do przeglądu

W przeglądach systematycznych stosowane jest podejście metodyczne do wydobycia istotnych informacji z badań podstawowych (*data extraction*). Zazwyczaj wiąże się to z opracowaniem formularza wydobycia danych, wypełnianego przez realizatorów przeglądu dla każdego włączonego badania. Formularz ten zawiera określenie próby (badanej populacji), szczegóły interwencji (o ile występują), wyniki, znaczenie metodologiczne i inne informacje. Metoda ta ma na celu zapewnienie spójności i obiektywności.

Pozyskiwanie danych obejmuje również sporządzanie tabeli opisującej każde badanie zakwalifikowane do przeglądu systematycznego (tylko badania spełniające wszystkie kryteria włączenia), które jest sprawdzane w szczegółach. Przykładowy, znormalizowany formularz dla każdego badania włączonego do przeglądu systematycznego zawiera następujące rubryki:

Tabela 6. Oprac. na podstawie: Simons, M. (2011) *Guideliness for writting systematic reviews*, <http://www.library.mq.edu.au/libguides/Guidelines%20for%20writing%20systematic%20reviews.pdf>.

Źródło – czasopismo, tytuł, autor, nr, strony
Cel badania – podany przez autora/-ów badania
Projekt badania – typ badania
Populacja – dane demograficzne uczestników badania
Interwencja – opis interwencji
Wpływ – opis grupy kontrolnej lub alternatywnej interwencji
Rezultaty – wyniki interwencji i sposób pomiaru statystycznego
Komentarz – szczegóły dotyczące jakości badania

Na tym etapie można wykorzystać specjalistyczne programy, takie jak: EPPI-Reviewer oraz TrialStat SRS, Access czy FoxPro. Podczas ekstrakcji danych należy zawsze pamiętać o pytaniu przeglądowym i odnosić dane do jego treści. Pobierane dane z badań jakościowych różnią się w zależności od przyjętej metody syntezy.

Proces oceny metod i wyników każdego badania jest często określany jako „ocena krytyczna”, a czasem jako „ocena jakości badań”. W przeglądzie systematycznym to zadanie ma na celu ustalenie, czy wyniki badania są adekwatne do odpowiedzi na pytanie (Petticrew i Roberts, 2006).

Ocena jakości badań jest często skrótowym określeniem oznaczającym „zasadność” – czyli stopień, w jakim dane badanie jest wolne od głównych błędów metodologicznych (np. stronniczość doboru badań, reakcji, utraty i uprzedzeń obserwatora; Petticrew i Roberts, 2006).

Jeśli wykonawcy przeglądu zdecydują się na ocenę jakości badań przed wydobyciem danych, będą mogli wykluczyć badania o niskiej jakości metodologicznej. Decyzje są wynikiem konsensusu dwóch recenzentów. Jest wiele sposobów oceny jakości dowodów z badań jakościowych w procesie przeglądu systematycznego. Nie ma tu jednak hierarchii dowodów, jak to jest w przypadku podejścia ilościowego. Zasadnicze pytania kontrolne są następujące:

1. Czy pytanie badawcze było jasno sprecyzowane?
2. Czy badanie przeprowadzono zgodnie z rygorem i etyką badawczą?
3. Czy w sposób przejrzysty podano informacje o metodach zbierania i analizy danych?
4. Czy badanie zawiera informacje wskazujące na błąd systematyczny badacza/y?
5. Czy badanie zawiera informacje o zweryfikowaniu wyników?

Umiejętność krytycznej oceny nie polega na identyfikacji problemów, ale na identyfikacji błędów, które są wystarczająco duże, aby wpływać na interpretację wyników.

Ocena krytyczna

Każde badanie zakwalifikowane do przeglądu należy oceniać na podstawie jego metodologicznej rzetelności. Proces ten przyczynia się do zidentyfikowania istotnych obciążeń stronniczością; pomaga także czytelnikowi zinterpretować dane. Rezultaty oceny krytycznej są wykorzystywane podczas syntezy wyników badań pierwotnych.

Również na tym etapie pomocne jest zastosowanie listy kontrolnej do oceny stronniczości. Różne listy kontrolne i skale jakości, w tym skala Jadada, zawierają następujące, wspólne kryteria oceny (Petticrew i Roberts, 2006):

- **jakość metodologiczna** – stopień, w jakim udało się zapobiec stronniczości na poziomie koncepcji badania;
- **dokładność** – miara prawdopodobieństwa przypadkowych błędów (zwykle przedstawiona w postaci przedziału ufności – CI, miara wiarygodności oszacowania wyniku);
- **trafność zewnętrzną** (*external validity*) – stopień, w jakim wyniki można uogólnić i zastosować w całej lub innej populacji.

Badania eksperymentalne niskiej jakości mogą zostać wyłączone z przeglądu lub syntezy wyników. Orzeczenia na temat jakości badań jakościowych mogą pomóc czytelnikom w wyrobieniu opinii na temat przydatności poszczególnych badań i całego przeglądu.

Pomocnym narzędziem typu GRADE są systemy łączące ocenę dowodów naukowych z rekomendacjami dla decydentów. System GRADE został opracowany do oceny jakości dowodów w przeglądach na temat skuteczności interwencji. Zawiera ocenę jakości badań włączonych do przeglądu oraz rekomendacje na podstawie syntezy wyników tychże badań.

Tabela 7. Stopniowanie jakości danych. Oprac. na podstawie: *Ocena jakości danych i klasyfikacja siły zaleceń*, http://www.gradeworkinggroup.org/_PL/publications/102_GRADE.pdf.

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

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

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

jakość bardzo niska = jakiegokolwiek oszacowanie efektu jest bardzo niepewne.

Istotnym, a nawet najważniejszym standardowym formatem oceny krytycznej przeglądów systematycznych jest lista kontrolna CASP – Critical Appraisal Tools: <http://www.casp-uk.net>.

Listy kontrolne mają ułatwiać ocenę każdego badania włączonego do przeglądu pod względem jakości metodologicznej, na podstawie informacji na temat doboru próby, analizy danych etc.

Narzędzia do oceny badań CASP są dostępne dla następujących typów badań: przeglądów systematycznych; kontrolowanych badań randomizowanych; badań jakościowych; badań ewaluacji ekonomicznej; badań kohortowych; badań kontrolnych przypadku; diagnostycznych badań testowych i badań zasad przewidywania klinicznego.

Warto również wspomnieć w tym miejscu o wytycznych do oceny jakościowej badań ewaluacyjnych opublikowanych przez brytyjski Urząd Rady Ministrów: *Quality in Qualitative Evaluation: A framework for assessing research evidence*, które dotyczą ewaluacji związanych z opracowaniem i wdrażaniem polityk społecznych i programów (Spencer i in., 2003).

Ocena krytyczna określa, jak dużą wagę przywiązuje się do dowodów z poszczególnych badań pierwotnych zakwalifikowanych do analizy i syntezy, w zależności od następujących kryteriów: adekwatność badania do treści tematu/pytania przeglądu; zastosowane metody każdego włączonego badania w kontekście danego przeglądu oraz jakość egzekwowania tych metod (Gough i in., 2012).

Ocena krytyczna często bywa łączona z ekstrakcją danych (Petticrew i Roberts, 2006). Na tym etapie wyodrębnia się kluczowe szczegóły i informacje z każdego badania zakwalifikowanego do przeglądu, wpisując je do szablonu bądź bezpośrednio do bazy danych lub tabeli.

Przykładowe formularze ekstrakcji danych znajdują się na stronach internetowych EPOC i CRD, odpowiednio: <http://www.epoc.uottawa.ca/resources.htm>; <http://www.york.ac.uk/inst/crd>.

Przeglądy systematyczne są badaniami wtórnymi, obserwacyjnymi – dlatego obciążenia błędami systematycznymi występujące w poszczególnych analizowanych badaniach pierwotnych również wpływają na rezultaty danego przeglądu. Zatem, przeglądy systematyczne, jak każde badanie, również powinny podlegać ocenie krytycznej.

Samo określenie „przegląd systematyczny” w tytule badania nie przesądza o jego znaczeniu i rzetelności (Petticrew i Roberts, 2006). Podobnie jak badania podstawowe, przeglądy systematyczne różnią się jakością. Wytyczne dla autorów w formie opracowanych szablonów (*checklists*) sprawdzania jakości PS, zwłaszcza najnowszego szablonu PRISMA (dawniej QUOROM), niewątpliwie podnoszą ich jakość. Opracowano również siedem wskazówek oceny jakości przeglądu systematycznego, przedstawionych w Tabeli 7, o których warto pamiętać podczas lektury przeglądu lub metaanalizy.

Tabela 8. Wskazówki pomocne w ocenie jakości przeglądu systematycznego (oprac. na podstawie: Petticrew i Roberts, 2006, s. 296).

1. Czy określono kryteria włączenia i wyłączenia i czy są one adekwatne?
2. Czy wyszukiwanie literatury wyłoniło przypuszczalnie wszystkie istotne badania?
3. Czy recenzenci ocenili jakość włączonych badań?
4. Czy recenzenci wzięli pod uwagę jakość badań przy syntezie wyników?
5. Jeśli przeprowadzono podsumowanie statystyczne (metaanalizę), czy była ona prawidłowa (tzn. czy badania były wystarczająco podobne, aby mogły zostać poddane syntezie statystycznej)?
6. Czy oceniono heterogeniczność (różnorodność) badań?
7. Czy wnioski recenzentów zostały opracowane na podstawie analizowanych badań?

Synteza wyników badań

Synteza, czyli proces integracji wyników z zestawu włączonych badań w celu uzyskania odpowiedzi na pytanie przeglądowe, stanowi sedno przeglądu systematycznego. Materiał dowodowy, uzyskany w wyniku poprzedzających działań, należy podsumować, aby otrzymać rezultaty (dowody) z przeglądu. Czasami jednak, może brakować dobrej jakości badań odpowiadających na pytanie przeglądowe. Wówczas, w wyniku przeglądu otrzymamy informację, czego nie wiemy danym zakresie i jakie badania należy przeprowadzić w przyszłości w kwestiach istotnych dla konkretnych odbiorców – np. polityków i urzędników edukacyjnych lub nauczycieli.

Synteza nie jest procesem linearnym, gdyż z przeglądem łączy się różna liczba powtórzeń. Przeważnie zawiera siedem głównych etapów odpowiadających kwestiom: jakie jest pytanie i koncepcja (teorie i założenia), jakie dane są dostępne (w odniesieniu do pytania przeglądu i koncepcji), wzory danych (badanie, interwencja, wyniki, charakterystyka próby), w jaki sposób integrowanie danych stanowi odpowiedź na pytanie przeglądu, jak solidna jest synteza pod względem wrażliwości, jakości, spójności i znaczenia, jaki jest rezultat, co on oznacza, co implikuje i jakie wyłoniły się nowe kwestie badawcze? (Gough i in., 2012).

Zarówno zwolennicy, jak i krytycy przeglądów systematycznych przyczyniają się do rozwoju metodologii integrowania wyników, co odzwierciedlają wielorakie rodzaje syntez, jak na przykład: synteza realistyczna (*realist synthesis*); metaprzeгляд (*review of reviews*), czyli przegląd systematyczny integrujący dane z innych przeglądów; metaewaluacja; przegląd systematyczny ewaluacji (jeden z rodzajów metaewaluacji); synteza metod łączonych (*mixed methods synthesis*) zawierająca różnorodne formy dowodów, np. badania ilościowe i jakościowe, bądź różne metody syntezy; metanarracyjna (*meta-narrative*), czyli krytyczna analiza pod kątem związku badań z ich paradygmatami badawczymi; metaetnografia (*meta-ethnography*) – popularna synteza danych jakościowych (Gough i in., 2012).

W nawiązaniu do dyskursu na temat użyteczności metod syntezy dowodów na rzecz polityk publicznych, w tym edukacyjnych, warto wspomnieć o metodzie syntezy

realistycznej¹ Pawsona, integrującej wiele dowodów analizowanych w kontekstach, mającej na celu określenie podstawowych mechanizmów przyczynowych oraz zbadanie, jak i w jakich warunkach te dowody się sprawdzają, odpowiadając na pytanie: *co działa na kogo, w jakich okolicznościach i pod jakim względem*, zamiast jedynie: *co działa*. Synteza realistyczna dąży do wyjaśnienia, dlaczego oceniany program sprawdza się w danym kontekście lepiej, niż w innym (Pawson, 2006).

Autor metody *realist synthesis* krytykuje możliwość stosowania metaanalizy w politykach publicznych i proponuje podejście mniej rygorystyczne metodologicznie. Metoda krytycznej oceny realistycznej zakłada, że wiedza jest wytworem społecznym i historycznym, a więc w analizie programu lub polityki trzeba wziąć pod uwagę kontekst społeczno-historyczny. Skuteczność programu nie zależy od samego wyniku, lecz ocena powinna przebiegać według schematu: kontekst – mechanizm teoretyczny – wynik (produkowany efekt). Kluczowym produktem jest tu informacja. Według Pawsona synteza realistyczna, analogicznie do kodeksu drogowego dla kierowców, informuje polityków o sprawach, które powinny być brane pod uwagę. Zamierzona funkcja przeglądu realistycznego pozostaje taka sama, jak konwencjonalnego przeglądu systematycznego, czyli wykorzystanie w podejmowaniu decyzji dotyczy kwestii, czy i jak wdrożyć przyszłe interwencje.

Synteza dowodów jakościowych (jakościowa)

Synteza danych jakościowych w przeglądzie systematycznym ma na celu zebranie wyników z doniesień o badaniach jakościowych oraz zidentyfikowania zgodności i różnic w obrębie tych wyników. O ile synteza ilościowa zasadniczo pozwala określić efekt (np. czy dana interwencja sprawdza się w praktyce, czy nie), o tyle synteza dowodów jakościowych umożliwia pogłębioną analizę tej samej kwestii – np. dla kogo i dlaczego coś się sprawdza (Boland i in., 2014).

¹ Przykładowy protokół syntezy realistycznej jest dostępny na stronie internetowej EPPI-Centre: Westhorp G., Walker B. i Rogers P. (2012). *Under what circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor?*. Protokół – *W jakich okolicznościach zwiększenie odpowiedzialności i umacnianie społeczności poprawia wyniki kształcenia, szczególnie wśród biednych?*

Synteza dowodów jakościowych pozwala podsumować punkt widzenia uczestników badań i jest bogatym źródłem informacji na temat wpływu warunków, interwencji lub polityki na doświadczenia i przeżycia uczestników. Łączenie rezultatów badań jakościowych może prowadzić do lepszego zrozumienia drażliwych kwestii, często poruszanych w tego typu badaniach.

Zaletą przeglądów badań jakościowych jest wolny wybór metody, który wiąże się z koniecznością obrony wybranej techniki analizy danych oraz uzasadnienia na piśmie wyboru konkretnej metody.

O ile w analizie dowodów ilościowych np. pięciu badaczy wykonujących niezależnie tę samą metaanalizę otrzyma takie same rezultaty, o tyle – stosując to samo pytanie przeglądowe do analizy danych jakościowych – możemy otrzymać pięć kompletnie różnych wniosków, w zależności od wyboru metody analizy, metody syntezy wyników oraz punktu widzenia poszczególnych badaczy.

Kluczowe różnice między przeglądem systematycznym badań ilościowych i jakościowych polegają na: ekstrakcji danych, krytycznej ocenie jakości badań włączonych (metodologii) oraz syntezie wyników.

Tabela poniżej przedstawia najważniejsze podobieństwa i różnice w syntezie dowodów jakościowych i ilościowych.

Tabela 9. Podobieństwa i różnice syntez (oprac. na podstawie: Boland i in., 2014, s. 145).

Etap	Synteza dowodów jakościowych	Synteza dowodów ilościowych
1.	Pytanie przeglądowe, określanie strategii wyszukiwania i protokół	
2.	Wyszukiwanie literatury naukowej	
3.	Sprawdzanie tytułów i abstraktów	
4.	Uzyskanie artykułów naukowych	
5.	Selekcja pełnych tekstów artykułów naukowych	
6.	Przyjęta teoria i plan analizy	Ocena jakości
7.	Ekstrakcja danych i ocena jakości	Ekstrakcja danych
8.	Analiza i synteza jakościowa	Analiza i synteza ilościowa
9.	Napisanie sprawozdania i edycja	

Na etapie tworzenia **pytania badawczego**, któremu podporządkowany jest cały przegląd, ważne jest skonkretyzowanie go w taki sposób, aby treść pytania odnosiła się do **dostępnych danych**; ważne jest również określenie kryteriów włączenia dających się zastosować do syntezy jakościowej lub ilościowej. W przypadku przeglądu badań jakościowych pytanie o doświadczenie matek wracających na rynek pracy, dla których jedynym rozwiązaniem jest umieszczenie dziecka w żłobku, jest jasno sformułowane, założmy jednak, że niewiele badań jakościowych będzie zajmować się tą kwestią. Bardziej prawdopodobne jest, że będzie to komponent badań o szerszym zakresie, w dziedzinie pielęgniarstwa i opieki podstawowej, psychologii lub szerszy temat w obszarze ekonomii rynku pracy. Fakt ten może stanowić asumpt do oceny stanu badań pod kątem pytania przeglądowego, przyczynić się do odkrycia innych ważniejszych aspektów danego fenomenu i w rezultacie – podążając w kierunku konkretnych badań – do zmodyfikowania pytania przeglądowego. Jest to często stosowane, chociaż czasochłonne podejście w wyznaczaniu zakresu przeglądu badań jakościowych.

Kolejną ważną różnicę stanowi opracowanie kryteriów włączenia. Istotne jest, aby opisać kryteria włączenia na początku przeglądu i zapisywać każde zmiany pojawiające się w trakcie przeprowadzania PS. Należy dokładnie opisać, jakie projekty badań, populacje, interwencje i wyniki są włączone/wyłączone w przeglądzie. Kryteria te powinny być opisane w raporcie końcowym lub artykule (Petticrew i Roberts, 2006).

W przeglądach badań ilościowych często stosowany jest model PICO lub tabela z rubrykami: kto, co, jak, gdzie.

Kryteria włączenia w odniesieniu do przeglądu badań jakościowych są mniej uszczegółowione i rekomenduje się raczej inny model – **PICo**: próba (*population*), dane zjawiska jako przedmiot zainteresowania (*phenomena of interest*), może to być stan (*condition*) lub interwencja oraz kontekst (*context*) (Joanna Briggs Institute, 2011 za: Boland i in., 2014, s. 146). Autorzy przeglądów systematycznych muszą uzasadnić i udokumentować wybór metody analizowania i łączenia wyników.

Tabela 10. Metody syntezy badań jakościowych (oprac. na podstawie: Boland i in. 2014).

Nazwa metody	Definicja
Metanarracja	– sposób identyfikacji „wątków” różnych badań jakościowych. Zazwyczaj wykorzystywana do informowania decydentów.
Metaetnografia	– sposób łączenia badań, zwłaszcza danych etnograficznych, w celu uzyskania nowego znaczenia lub teoretycznego zrozumienia.
Metaagregacja	– sposób zbierania wniosków z badań podstawowych i łączenie na podstawie podobieństwa znaczenia.
Metastreszczenie	– sposób mapowania treści badań jakościowych.
Metaanaliza ugruntowana	– analiza uzasadniająca zawartość tekstu w kontekście miejsca jego powstania.
Jakościowa synteza dowodów	– przeglądy integrujące i analizujące wyniki empirycznych badań jakościowych.

Synteza dowodów ilościowych (ilościowa)

Istotą syntezy dowodów ilościowych jest łączenie informacji o sile efektów z badań, odpowiadających na określone pytanie badawcze. Proces syntezy dotyczy analizy uzyskanych danych, szukania wzorów i interpretacji. Metody integracji i interpretacji danych zależą od treści pytania oraz metod zastosowanych w badaniach włączonych.

Synteza zawiera również element kreatywny – nie tylko integruje wyniki z badań włączonych, lecz próbuje podsumować informacje, aby uzyskać autorytatywną, odpowiedź, lepszą niż pojedyncze badania podstawowe włączone do przeglądu.

Jednostką analizy są miary efektów (*effect size*), które najprościej można zdefiniować jako wskaźniki siły i kierunku zależności między dwoma zmiennymi. Miary efektów (ES) są obliczane dla każdego badania uwzględnianego w analizie, a następnie oszacowywany jest rzeczywisty efekt. Zazwyczaj jest to średnia ważona, w której poszczególnym badaniom przypisane są różne wagi. Drugim ważnym celem metaanalizy jest oszacowanie różnicowania efektów. Skupienie uwagi na ES odróżnia metaanalizę od tradycyjnego, i często krytykowanego w ostatnich latach, podejścia do analizy danych, w którym zbyt

duże znaczenie przypisuje się istotności statystycznej efektów (np. Cumming, 2012; Kline, 2013). Istotą metaanalizy jest skupienie uwagi na praktycznej istotności wyników, co sprzyja popularności tej metody syntezy wyników, zwłaszcza wśród praktyków, dla których przebicie się przez gąszcz doniesień naukowych i wyciągnięcie z nich wniosków jest trudne i czasochłonne. Dodatkową zaletą metaanalizy jest możliwość rozszerzenia zakresu wnioskowania: np. przez uwzględnienie badań przeprowadzonych w różnym czasie, w odmiennych populacjach, różnych lokalizacjach czy różniących się schematem badania.

Proces przygotowania metaanalizy wiąże się z wyzwaniami podobnymi do procesu przygotowania przeglądu systematycznego i obejmuje analogiczne kroki postępowania. Z tych powodów wielu autorów przeglądu traktuje metaanalizę jako praktyczny i systematyczny sposób wyprowadzania wniosków z przeglądu systematycznego. Zarówno w przeglądach systematycznych, jak i syntezie wyników ilościowych, szczególne znaczenie ma decyzja, czy łączenie badań przeprowadzonych w różnych populacjach, obejmujących różne interwencje czy stosujących różne narzędzia pomiarowe, jest uzasadnione (PICO). Ważne jest też zapewnienie rzetelności we wprowadzaniu danych i kodowaniu badań czy ocena wiarygodności uwzględnianych badań. Ogólną zasadą jest przejrzystość kolejnych kroków oraz możliwość replikacji uzyskanych wyników. Zasadniczą rolę w przydatności wyników metaanalizy ma także sposób postawienia głównego pytania, na które analizy mają dać odpowiedź. Jak zauważają Jaeschke i in. (2008) najbardziej pracochłonnym etapem metaanalizy jest czytanie artykułów wraz z jakościową i ilościową ekstrakcją informacji (oraz ewentualne kontakty z autorami). Od czasu poświęconego na ten etap i jakości tego procesu zależy wiarygodność wyników. Przeprowadzenie analiz statystycznych stanowi więc jeden z prostszych i najmniej pracochłonnych części procesu. Ważnym elementem jest też zapewnienie rzetelności wprowadzania informacji z badań, w tym kodowania i oceny jakości dowodów. Do oceny rzetelności można wykorzystać techniki statystyczne służące ocenie zgodności oceniających.

Wyniki metaanalizy są przedstawiane w postaci zestawienia badań i uzyskanych w nich efektów statystycznych oraz wyników analiz statystycznych na tych danych. Najczęściej wykorzystuje się do tego tzw. wykres drzewkowy (*forest plot* – zob. *Suplement*).

Wszystkie zrealizowane metaanalizy, jako osobno wydane badania, są zindeksowane i łatwe do zidentyfikowania dla czytelników, gdyż autorzy zobowiązani są umieścić termin *metaanaliza* w tytule raportu końcowego. Najbardziej polecany format raportowania metaanalizy, MARS (Meta-Analysis Reporting Standards), jest jednocześnie praktycznym przewodnikiem po jej niezbędnych etapach (Cooper, 2010).

Zalecenia dotyczące opisu metodologii metaanalizy

Tabela 11. Wybrane etapy raportowania metaanalizy i opracowanie na podstawie standardu MARS (za: Cooper, 2010, s. 222–224).

Kryteria włączenia i wyłączenia
<p>Sposób operacjonalizacji zmiennej(-ych) niezależnych (wyjaśniających) i zależnych (wyjaśnianych).</p> <p>Badane populacje.</p> <p>Uwzględniane schematy badań (np. tylko badania z losowym przydziałem uczestników, minimalna wielkość próby).</p> <p>Cezura czasowa badań.</p> <p>Obwarowania geograficzne lub kulturowe.</p>
Analizy zmiennych niezależnych
<p>Definicje wszystkich kategorii kodowania zmiennych niezależnych, które wpływają na siłę analizowanej zależności, czyli moderatorów (<i>moderators</i>), lub pośredniczą między zmiennymi wyjaśniającymi a wyjaśnianymi (<i>mediators</i>).</p>
Strategie wyszukiwania badań do metaanalizy
<p>Przeszukiwanie baz danych: bibliograficznych i cytowań.</p> <p>Przeszukiwanie rejestrów (włącznie z rejestrami prospektywnymi (zapowiadającymi metaanalizę):</p> <ul style="list-style-type: none"> • słowa kluczowe zastosowane w bazach danych i rejestrach; • oprogramowanie wyszukiwawcze i wersja. <p>Cezura czasowa dla badań, jeśli dostępna</p> <p>Inne starania w celu znalezienia wszystkich dostępnych badań, np.:</p> <ul style="list-style-type: none"> • listy autorów indeksowane wg zagadnień (<i>listservs queried</i>); • kontakty bezpośrednie z autorami (sposób wyboru autorów); • sprawdzanie według list odniesień raportów.

Strategie wyszukiwania badań do metaanalizy
<p>Sposób adresowania raportów w językach innych niż angielski.</p> <p>Proces określania kwalifikowalności badań do metaanalizy.</p> <p>Aspekty sprawdzania sprawozdań (np. tytuł, abstrakt lub cały tekst)</p> <ul style="list-style-type: none"> • liczba i kwalifikacje trafności oceniających; • podanie umowy [zasad oceniania]; • rozwiązywanie kwestii spornych [ocena kompromisowa, np. kalkulator Kappa]. <p>Traktowanie badań niepublikowanych</p>
Procedury kodowania
<p>Liczba i kwalifikacje kodujących (np. poziom doświadczenia w danej dziedzinie, wykształcenie).</p> <p>Wiarygodność/niezawodność.</p> <p>Jeśli każdy raport był kodowany przez więcej niż jedną osobę, jak rozwiązywano kwestie sporne.</p> <p>Ocena rzetelności metodologicznej badań:</p> <ul style="list-style-type: none"> • opis kryteriów i procedur zastosowania skali jakości, jeśli występuje; • podać cechy projektu/schematu badania, jeśli były kodowane. <p>Sposób rozwiązania problemu braku danych.</p>
Metody statystyczne
<p>Miary efektów</p> <ul style="list-style-type: none"> • sposób obliczania miar efektów i ich transformacji; • poprawianie siły efektu (np. obciążenie (bias) wynikające z małej próby, korekta nierównych wielkości próby, etc.). <p>Metoda/y uśredniania lub ważenia efektu.</p> <p>Jak obliczano przedziały ufności efektu (lub błędy standardowe).</p> <p>Jak obliczano przedziały wiarygodności efektu (jeśli występuje).</p> <p>Jak rozwiązano kwestie badań mających więcej niż jeden efekt. lub</p> <p>Jakiego modelu analizy użyto i czy uzasadniono wybór modelu.</p> <p>Jak oceniano lub analizowano niejednorodność efektów.</p> <p>Średnie i odchylenia standardowe pomiaru artefaktów, jeśli przedmiotem zainteresowania były związki na poziomie mierzonych konstruktywów.</p> <p>Sposób analizy obciążeń danych wynikających z problemu selekcji (np. dotyczące stronniczości publikacji, wybiórczej sprawozdawczości).</p> <p>Testy statystycznych wartości obserwacji skrajnych (<i>outliers</i>).</p> <p>Moc statystyczna metaanalizy.</p> <p>Programy i pakiety oprogramowania użyte do prowadzenia analizy statystycznej.</p>

<p style="text-align: center;">Ocena stronniczości publikacji Opracowanie raportu Upowszechnianie wyników</p>
--

Ocena stronniczości publikacji

Należy wziąć pod uwagę wpływ stronniczości publikacji oraz innych wewnętrznych i zewnętrznych uwarunkowań obciążających wyniki (*biases*). Wiadomo, że takie kwestie, jak zakres i jakość badania, źródła finansowania oraz stronniczość publikacji mogą wpływać na wyniki badań podstawowych. Mogą także mieć istotny wpływ na wnioski z przeglądu systematycznego badań ilościowych, a w najgorszym razie – wyolbrzymiać prawdziwą wielkość efektu, o którym mowa w pytaniu. W przypadku badań ilościowych skutki obciążeń stronniczością mogą być przedstawione opisowo lub za pomocą wykresu lejkowego (np. *funnel plot*).

Do najbardziej znanych czynników obciążających wyniki należą: *selection bias*, czyli stronniczość wyboru badań (np. tylko badania o pozytywnych wynikach), i *publication bias*, czyli stronniczość publikacji (np. brak uwzględnienia tzw. szarej literatury, tylko czasopisma recenzowane).

Niestety, odchylenia obciążające wyniki możliwe są na każdym etapie, stąd przyjęte rygory badawcze i opracowane listy wymaganych kryteriów, aby możliwie jak najbardziej zredukować ewentualne zniekształcenia wyników.

Opracowanie raportu lub artykułu

Etapem końcowym procesu przeglądu systematycznego jest publikacja raportu lub artykułu naukowego. Wersja końcowa opracowania przeglądu systematycznego musi zawierać szczegóły dotyczące całościowego wyszukiwania i selekcji badań w procesie

przeglądu, włącznie z tym, jakie i jak wiele badań zostało odrzuconych na każdym etapie i dlaczego, oraz daty rozpoczęcia i zakończenia procesu wyszukiwania. Podanie tej informacji (np. w schemacie blokowym) jest warunkiem wstępnym publikacji w niektórych czasopismach naukowych o tematyce medycznej.

Częstym warunkiem publikowania jest postępowanie autorów przeglądów systematycznych zgodnie z wytycznymi, np. dawniej QUOROM (Quality Of Reporting Of Meta-analyses), obecnie PRISMA 2009 – zapewniającymi wysoką jakość sprawozdawczości: Moher D., Liberati A., Tetzlaff J. i Altman D.G., The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*. PLoS Med 6(6). *PRISMA 2009 checklist*. <http://www.prisma-statement.org>.

Przeglądy są regularnie aktualizowane (w C2 co pięć lat), zapewniając interesariuszom bazowanie w procesie decyzyjnym na najbardziej aktualnych i wiarygodnych dowodach.

Upowszechnianie wyników

W początkowej fazie przeglądu trzeba jasno określić odbiorców, a nawet z góry opracować plan upowszechniania wyników przeglądu, ich interpretacji i wykorzystywania (uwzględnić w protokole).

W końcowym etapie upowszechnianie może obejmować opracowanie streszczenia lub innych uproszczonych i skróconych wersji dla decydentów, urzędników i innych odbiorców. Upowszechnianie może także obejmować współpracę z odbiorcami w celu wdrożenia wyników oraz pomoc potencjalnym użytkownikom w zrozumieniu znaczenia przeglądu i ustaleń dotyczących polityki, praktyki i przyszłości. W tym celu można zastosować takie środki, jak: konferencje, spotkania informacyjne, seminaria, spotkania publiczne, media, sieci organizacji (*networking*) etc.

Kilka pojęć stosuje się do opisu procesu i form upowszechniania i wykorzystywania dowodów, np. transfer wiedzy, tłumaczenie wiedzy, pośrednictwo w wymianie wiedzy i uruchamianie wiedzy (odpowiednio: *knowledge transfer*, *knowledge translation*, *knowledge mediation*, *knowledge mobilisation*).

Zakończenie

W świetle polityki finansowej Komisji Europejskiej projekty dotyczące wdrażania metodologii przeglądów systematycznych i związanych z nimi baz danych w Polsce stanowią jeden z priorytetów unijnego finansowania. Przeglądy te umożliwiają spożytkowanie danych z badań podstawowych, zgodnie ze wskazaniem KE do wykorzystywania w aktywności badawczej danych zastanych ze względów ekonomicznych i praktycznych, co zostało podkreślone przez Jana Pakulskiego – kierownika Działu ds. Statystyk, Badań i Analiz, Dyrekcji Generalnej ds. Edukacji i Kultury w Komisji Europejskiej – na trzeciej międzynarodowej konferencji EIPPEE o wykorzystaniu dowodów naukowych w Oslo: *Finding a way... to facilitate the better use of research in education* (14-15 maja, 2014).

Przeglądy systematyczne są podstawową metodą zarządzania wiedzą w kontekście tzw. ruchu na rzecz wykorzystywania dowodów w procesie decyzyjnym. Dowodem uzyskanym z przeglądu systematycznego jest rezultat syntezy wyników wszystkich istotnych badań naukowych, przeprowadzonej w sposób jawny, obiektywny, przejrzysty, powtarzalny, odpowiedzialny i podlegający aktualizacji. W tych aspektach przeglądy systematyczne różnią się od tradycyjnego, subiektywnego i wybiórczego sposobu przedstawiania stanu badań w tradycyjnych, przeglądowych pracach naukowych.

Cele przeglądu systematycznego mogą być następujące: wykazanie skutecznych i pomocnych rozwiązań w edukacji oraz wyeliminowanie fałszywych teorii, pokazanie mocnych i słabych stron zastanej wiedzy w danej kwestii, wskazanie na luki w stanie badań, ocena potrzeby przeprowadzenia szeroko zakrojonych badań, zintegrowanie wyników z wielu badań, poprawienie wyników interwencji i unikanie nieskutecznych programów. Metoda przeglądu systematycznego może także służyć doradztwu w zakresie polityk prowadzonych na podstawie rzetelnych i konkretnych dowodów.

Korzyścią, którą warto zaznaczyć, jest oszczędność czasu, jaki inaczej musieliby poświęcić badacze, politycy i interesariusze na opanowanie informacyjnego „tsunami”, aby otrzymać obiektywne i solidne dowody w danej kwestii. Przeglądy systematyczne są praktycznym narzędziem zarządzania ogromnym zasobem informacji, niemożliwych do przetworzenia w inny sposób.

W naukach społecznych na Zachodzie przeglądy systematyczne stały się standardem badawczym (pozwalającym dostrzec różnicę pomiędzy faktyczną a domniemaną wiedzą w danej kwestii) oraz narzędziem pomocnym w ocenie, podsumowaniu i przekazywaniu wyników danych naukowych. Przeglądy systematyczne są tam szeroko stosowane nie tylko w dziedzinie medycyny i ochrony zdrowia, lecz także w naukach społecznych i humanistycznych; mają bogatą bibliografię oraz towarzyszącą im infrastrukturę w formie ośrodków naukowych i elektronicznych baz danych.

Instytucje takie jak The Campbell Collaboration, What Works Clearinghouse, EPPI-Centre oraz Danish Clearinghouse for Education przyczyniają się do produkcji, archiwizowania i upowszechniania przeglądów systematycznych w naukach społecznych, w tym w dziedzinie edukacji. Doceniając ich szersze zastosowanie w polityce i praktyce edukacyjnej, warto podkreślić użyteczną funkcję przeglądów w określaniu wpływu i skuteczności rządowych programów. W Polsce brakuje podobnych baz danych archiwizujących dowody i jest to zadanie na perspektywę 2014–2020.

Prawidłowo przeprowadzony przegląd systematyczny, w połączeniu z metaanalizą, to dwa bardzo kompatybilne podejścia do syntezy wyników, które tworzą potężne, naukowe narzędzie do uzyskiwania najbardziej rzetelnych dowodów z zastanych badań. Jednak, warunkiem wydobycia pełnego potencjału z badania wtórnego, jakim jest przegląd systematyczny, jest biegła znajomość jego fundamentalnych etapów.

Skutecznie rozpowszechnione wyniki przeglądów systematycznych mogą wywierać istotny wpływ na decydentów, jak na przykład zatrzymanie i wycofanie wadliwego programu lub (przy wysokim poziomie jakości dowodów) wprowadzenie i udoskonalenie korzystnych interwencji.

Na użytek polityki i praktyki edukacyjnej opartej na dowodach potrzebne są wiarygodne wnioski z badań, zwłaszcza w celu podejmowania uzasadnionych decyzji dotyczących finansowania edukacji. Przeglądy systematyczne odpowiadają na tę potrzebę, pomagając w podejmowaniu decyzji korzystnych dla społeczeństwa. Dostarczając solidnych i stale aktualizowanych dowodów naukowych, zapewniają wiedzę, która może być wykorzystywana przez długi czas.

W Polsce obecnie jesteśmy na początku drogi pod względem realizacji przeglądów systematycznych na potrzeby edukacji – mamy nadzieję, że niniejsza publikacja okaże się dla przyszłych realizatorów tych badań przydatnym drogowskazem.

Suplement

Podstawy statystyczne metaanalizy

Techniki statystyczne wspierające metaanalizę są wyspecjalizowaną, dynamicznie rozwijającą się dziedziną i nie sposób wyczerpująco omówić tego obszaru w ramach niniejszej publikacji. Dlatego poniżej omówiono tylko główne kwestie związane z metaanalizą, unikając technicznych szczegółów. Zainteresowanych czytelników odsyłamy do bogatej literatury przedmiotu, w tym licznych przewodników metodologicznych (np. Ellis, 2010; Lipsey i in., 2001)¹.

Metaanaliza w pakietach statystycznych

Procedury wykorzystywane w metaanalizie nie są integralną częścią popularnych pakietów statystycznych. Są oferowane jako dodatkowe moduły (np. moduł „Metaanaliza” w „Statistica. Zestaw Medyczny”) bądź są dostępne jako procedury napisane przez użytkowników (np. w programie Stata czy SPSS). Istnieją dedykowane programy wspomagające metaanalizę: zarówno komercyjne (np. MetaWin czy Comprehensive Meta-Analysis), jak i bezpłatne (np. RevMan). Duże możliwości dają procedury dostępne w darmowym pakiecie R (zob. <http://cran.r-project.org/web/views/MetaAnalysis.html>). Dostępne są także dodatki umożliwiające przeprowadzanie podstawowych obliczeń w arkuszu kalkulacyjnym (np. Excel).

Miary efektów

W metaanalizie posługujemy się dostępnymi wynikami badań, w których często wykorzystywano różne narzędzia pomiarowe, np. mierzono umiejętności uczniów, używając różnych rodzajów testów umiejętności matematycznych. Porównanie wyników poszczególnych badań byłoby niemożliwe bez uzgodnienia wspólnej miary siły efektów. Miara taka powinna umożliwiać porównanie różnych badań, czyli dawać możliwość obliczenia efektu i jego błędu standardowego, i być możliwa do prostego uzyskania lub wyliczenia na podstawie opublikowanych wyników badań (bez przeprowadzania

¹ W języku polskim opis podstawowych zasad metaanalizy jest dostępny w podręcznikach medycznych (np. Leśniak i in. 2008).

wtórnych analiz danych jednostkowych). Miara efektów powinna też być łatwo interpretowalna i czytelna dla odbiorców.

Wykorzystywane w metaanalizach miary efektów są pochodną głównego pytania badawczego metaanalizy, schematu porównywanych badań i rodzaju analizowanych zmiennych. Aby poradzić sobie z problemem porównywalności wyników, wypracowano różne rodzaje miar efektów. Można je podzielić na trzy kategorie: miary wykorzystujące współczynniki korelacji, miary opierające się na wartościach średnich oraz miary dla zmiennych dychotomicznych (w których zmienna wyjaśniana przyjmuje dwie wartości 0 lub 1, zwykle odzwierciedlające wystąpienie lub nie określonego wydarzenia).

Tabela S1. Wybrane miary efektów wykorzystywane w metaanalizach (oprac. na podstawie: Ellis, 2010).

Miary oparte na korelacjach lub wyjaśnianej wariancji	
r	Współczynnik korelacji liniowej Pearsona wyraża liniową zależność między dwiema zmiennymi (zmiernymi co najmniej na skali interwałowej). Korelacja jest silna, jeśli może być opisana przy pomocy linii prostej.
r^2	Współczynnik determinacji – r podniesione do kwadratu. Wyraża siłę związku – proporcję wspólnej zmienności dwóch zmiennych.
R^2	Współczynnik wielorakiej determinacji – stopień, w jakim dwie lub więcej zmiennych objaśniających (niezależnych lub X) jest powiązanych ze zmienną objaśnianą (zmienna zależna Y).
η^2	Eta-kwadrat – <i>wskaźnik siły efektu</i> : pokazuje, jaki procent zmienności w zakresie zmiennej objaśnianej (zmiennej zależnej) jest wyjaśniany przez zmienną objaśniającą (zmienną niezależną).
ω^2	Omega-kwadrat – względnie nieobciążony wariant eta-kwadrat.
ρ (lub r_s)	Rho Spearmana – współczynnik korelacji dla zmiennych porządkowych. Oblicza się go na podstawie rang, a nie wartości.
r_{pb}	Współczynnik korelacji dwuseryjnej – do obliczania siły związku między zmienną ciągłą a dychotomiczną.
V Cramera	Wykorzystywany do siły związku zmiennych nominalnych.

Porównanie grup na ciągłej zmiennej wyjaśnianej	
d	d Cohena: Standaryzowana różnica średnich: różnica między średnimi wyrażona w jednostkach odchylenia standardowego.
g	g Hedgesa: Skorygowana różnica średnich.
Δ	Delta Glassa: Standaryzowana różnica średnich, w których różnica jest wyrażona w jednostkach odchylenia standardowego w grupie kontrolnej.
Porównanie grup na dychotomicznej zmiennej wyjaśnianej	
OR	Iloraz szans: porównuje szansę wystąpienia zdarzenia w jednej grupie z szansą wystąpienia tego zdarzenia w drugiej grupie.
RR	Względne ryzyko: porównanie prawdopodobieństwa wystąpienia zdarzenia w jednej grupie z prawdopodobieństwem wystąpienia tego wydarzenia w drugiej grupie.
RD	Różnica ryzyk: różnica między prawdopodobieństwem wystąpienia zdarzenia w dwóch grupach.

Najpopularniejszą i najbardziej intuicyjną miarą wykorzystywaną w porównaniach wyników badań jest standaryzowana różnica średnich (d Cohena). Można ją wykorzystywać w sytuacji, gdy zmienna zależna jest zmienną ciągłą (którą można wyrazić na skali numerycznej), jest lub może być wyrażona we wszystkich badaniach na tej samej skali i jest łatwo interpretowalna. Przykładowo, różnica między wynikami chłopców i dziewcząt w części egzaminu gimnazjalnego mierzącej umiejętność czytania z 2012 r. wyniosła ok. 0,41 wartości odchylenia standardowego. Wartość ta jest nieco niższa w stosunku do różnicy uzyskanej przez polskich piętnastolatków w badaniu OECD PISA (d=0,48). Stwierdzenie tego faktu nie jest łatwe, jeśli porównać wartości, w których wyrażone są wyniki obu testów (w egzaminie gimnazjalnym z 2012 r. średnia dziewcząt wyniosła 69%, chłopców – 61%, a odchylenie standardowe – 19,3%. W badaniu PISA 2012 było to odpowiednio: 539, 497 i 87 pkt.). Słabością standaryzowanej różnicy średnich jest przeszacowywanie efektów w przypadku badań na małych próbach.

Tę słabość niweluje korekta we wzorze na tzw. g Hedges'a. Trzecią z najczęściej wykorzystywanych miar opartych na standaryzowanej różnicy średnich jest g Glassa (różnica między średnimi podzielona przez odchylenie standardowe w grupie kontrolnej, oznaczana jako g_{Glass} , g' lub D).

Drugim, często używanym rodzajem miar efektów, są miary oparte na współczynniku korelacji, stosowane w porównaniach wyników badań, w których zestawiane są dwie zmienne ciągłe. Ponieważ współczynników korelacji nie można uśredniać, w analizach zwykle przekształca się je na wartości z Fishera, a uzyskany wynik przekształca się w bardziej intuicyjny i łatwiej interpretowalny współczynnik korelacji.

W badaniach, w których analizowane zmienne zależne są zmiennymi dychotomicznymi (najczęściej w postaci tabeli 2×2), efekty mogą być wyrażone jako ryzyko wystąpienia danego zdarzenia, współczynnika ryzyka (RR) czy różnicy ryzyk (RD). Często stosowaną miarą efektu jest iloraz szans (OR). Analizę miar efektu z poszczególnych wyrażonych w postaci ilorazu efektów przeprowadza się na skali logarytmicznej (która ścieśnia wartości na końcach przedziału). Mimo że dla pojedynczej jednostki zmienna przyjmuje tylko dwie wartości, to w przypadku większej liczby badanych można wyliczyć proporcje występowania danego zdarzenia. Tego rodzaju miary są stosowane w badaniach o schemacie eksperymentalnym i badaniach szacujących efekty interwencji.

W literaturze dotyczącej metaanaliz dostępne są wzory umożliwiające dokonywanie przekształceń między różnymi rodzajami efektów. Jednak, mimo że jest np. możliwe przekształcenie korelacji na d Cohena, to w przypadku uwzględnienia bardzo różnych sposobów pomiaru zmiennych czy różnych schematów badawczych zalecana jest ostrożność, bo tego rodzaju przekształcenia mogą wprowadzać obciążenia i utrudniać interpretację uzyskanych wyników. Warto zauważyć, że prowadzenie metaanaliz jest łatwiejsze w przypadku nowszych doniesień, bo w ostatnich latach wiele czasopism i organizacji profesjonalnych zaleca w publikacjach naukowych wyrażanie wyników w postaci miar efektów.

W naukach społecznych określone wartości progowe miar efektów są uznawane za umowne granice efektów słabych, przeciętnych i silnych. W praktyce jednak siła efektów powinna być interpretowana w kontekście badań danego zjawiska, gdyż może ona różnić się w badaniu poszczególnych zjawisk i różne może być ich praktyczne znaczenie. Przykładowo, w medycynie nawet niewielkie wartości miary efektów interwencji

mogą przemawiać za zaleceniem jej stosowania. Z kolei w badaniu niektórych zjawisk społecznych uzyskanie silnych efektów jest czasami niemożliwe.

Tabela S2. Interpretacja wartości miar wybranych efektów zaproponowana przez Cohena (1988)

	Efekt słaby	Efekt umiarkowany	Efekt silny
Iloraz szans (OR)	1,5	2,5	4,3
Standaryzowana różnica średnich (SMD)	0,2	0,5	0,8
Współczynnik wielorakiej determinacji (R ²)	0,02	0,13	0,26

Precyzję wyników przedstawia się w postaci błędów standardowych lub wyliczonych na podstawie błędów standardowych przedziałów ufności. Błąd standardowy jest funkcją wielkości próby i zróżnicowania danego parametru w próbie. Im większy błąd standardowy, tym mniejsza precyzja wyniku. Na precyzję oszacowań wpływa schemat losowania oraz schemat badania. Schemat losowania próby może zwiększać lub zmniejszać efektywność prób. Przykładowo, próba uczniów w badaniu, w którym zastosowano losowanie dwustopniowe: najpierw wylosowano szkoły, a następnie uczniów w tych szkołach, będzie mniej efektywna od prostej próby losowej, w której losujemy pojedynczych uczniów do badania spośród wszystkich uczniów w Polsce. Na precyzję oszacowań ma także wpływ schemat badania. Jeśli całe szkoły lub oddziały w szkołach poddane są interwencji, a grupę kontrolną stanowią inne szkoły lub oddziały, to będzie to schemat mniej efektywny od zwykłego badania eksperymentalnego, w którym do interwencji są przydzielane losowo pojedyncze osoby. W metaanalizach zazwyczaj przypisuje się większą wagę badaniom, w których oszacowania wyników są bardziej precyzyjne.

Drugim źródłem błędów jest rzetelność pomiaru. Niska rzetelność wykorzystanych narzędzi pomiarowych (np. testów psychologicznych) wpływa na oszacowywane

efekty. Z tego względu w niektórych metaanalizach uwzględnia się informację o rzetelności wykorzystywanych miar w oszacowaniu efektów. W przypadku braku informacji o rzetelności w niektórych badaniach przyjmuje się zazwyczaj średnią wartość rzetelności.

Obliczanie i prezentowanie łącznej siły efektu

Metaanaliza opiera się na założeniu, że efekty uzyskane w poszczególnych badaniach mają określony rozkład i możliwe jest analizowanie zmienności efektów za pomocą technik statystycznych. Ważne jest przy tym założenie, że poszczególne badania są od siebie niezależne. W praktyce często spotykanym problemem jest występowanie kilku publikacji opartych na tych samych danych. Zazwyczaj są one uśredniane, ale zdarza się, że uzyskane różnice mogą wynikać z innych sposobów analizy wyników, które mogą być interesujące z punktu widzenia celów metaanalizy. Wówczas można te różnice uwzględnić w metaregresji.

W metaanalizach wykorzystywane są najczęściej dwa modele regresji wielopoziomowej (hierarchicznej): modele efektów stałych i modele efektów losowych (zob. np. Hedges i Vevea 1998). Podstawowym założeniem tych modeli jest rozróżnienie między efektem prawdziwym (*true effect*) i zaobserwowanym (*observed effect*). Istotą ogólnej klasy modeli wielopoziomowych jest wprowadzenie możliwości analizy wariancji i ich zróżnicowania, w tym dzielenia wariancji na różne komponenty. Ogólne różnice między tymi modelami przedstawia tabela S2. W modelu efektów stałych zakładamy, że istnieje jedna prawdziwa siła efektów (stąd nazwa), wspólna dla wszystkich badań uwzględnionych w analizie. Opiera się ona na założeniu, że różnice między wynikami uzyskanymi w poszczególnych badaniach traktowane są jako błąd próbkowania (*sampling error*). Przykładowo, przeprowadzając badanie tą samą metodą na różnych próbach z tej samej populacji, możemy uzyskać różną wartość efektów, ponieważ skład wylosowanych prób będzie się nieznacznie różnił. W modelu efektów losowych dopuszczamy możliwość, że prawdziwe efekty mogą się różnić, w związku z czym możemy wyróżnić część zróżnicowania, za którą odpowiada błąd próbkowania i część zróżnicowania niewyjaśnioną przez ten błąd. Niewyjaśnioną część zróżnicowania mogą powodować np. różnice w intensywności interwencji czy sposobie jej wdrożenia.

Jeśli byłoby możliwe przeprowadzenie bardzo dużej ilości badań, prawdziwe efekty przyjąłby postać rozkładu normalnego i byłyby skoncentrowane wokół średniej tego rozkładu. Popularniejsze w metaanalizach są modele efektów losowych, ponieważ większość metaanaliz obejmuje badania zróżnicowane ze względu na schemat badania czy badane populacje.

Tabela S3. Różnice między modelem efektów stałych a modelem efektów losowych (oprac. własne).

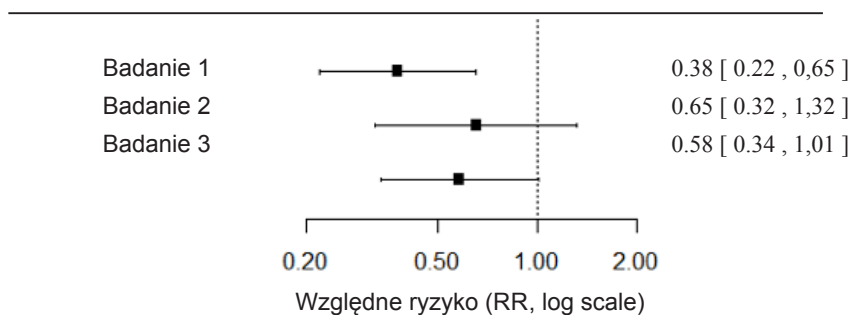
	Model efektów stałych	Modele efektów losowych
Założenie dotyczące efektów	Istnieje wspólny prawdziwy efekt dla wszystkich badań uwzględnionych w analizie.	Prawdziwy efekt w uwzględnionych badaniach jest próbą rozkładu prawdziwych efektów.
Założenia dotyczące zróżnicowania efektów	Różnice między badaniami odzwierciedlają tylko błędy próbkowania (<i>sampling error</i>). Wagi przypisane badaniom minimalizują ten błąd (wariancja wewnątrz badań).	Ponieważ celem jest oszacowanie średniej w rozkładzie, to uwzględniane są dwa źródła wariacji (wariancja wewnątrz badań i wariancja między badaniami). Wagi przypisane badaniom minimalizują oba rodzaje wariacji.
Wynik	Łączny efekt jest oszacowaniem wspólnego efektu dla uwzględnionych badań (efektu stałego).	Łączny efekt jest oszacowaniem wspólnego efektu dla uwzględnionych badań oraz błędu wynikającego z tego, że efekty są próbą efektów z wszystkich możliwych badań. Celem jest oszacowanie średniego efektu i jego zróżnicowania w populacji. Wiąże się to z większym błędem standardowym oszacowania.

Wyniki metaanalizy przedstawia się w postaci miar efektów, najczęściej tych samych, które analizowano w uwzględnianych badaniach. W obliczaniu łącznego efektu nigdy

nie wylicza się prostej średniej arytmetycznej, ale uwzględnia różną wagę badań, przypisując większe znaczenie tym, w których oszacowanie efektów jest precyzyjniejsze. Najczęściej przyjętym rozwiązaniem jest ważenie badań na podstawie wielkości próby badawczej. Przypisana waga może też zależeć od przyjętego modelu analizy (np. może różnić się w przypadku zastosowania modelu efektów stałych i modelu efektów losowych).

Powszechnie przyjętym sposobem komunikowania wyników metaanalizy jest przedstawienie ich w formie graficznej w postaci tzw. wykresu drzewkowego (w języku polskim zwanym też leśnym – od ang. *forest plot*). Wykres ten obrazuje efekt łączny w kontekście wyników poszczególnych badań uwzględnionych w metaanalizie. Na wykresie przedstawia się wyniki poszczególnych badań, wraz z przedziałami ufności oszacowanych efektów oraz wynik łączny, umieszczony na dole wykresu. Na wykresie można także uwzględnić inne informacje (np. wagę badania zilustrowaną wielkością punktu prezentującego oszacowanie efektu).

Przykład wykresu drzewkowego (*forest plot*)



Łączny efekt metaanalizy przedstawia się wraz z przedziałem ufności oraz z podaniem informacji o poziomie istotności efektu. W metaanalizie wykorzystuje się też pojęcie przedziału wiarygodności. Różnica między tymi przedziałami wiąże się z rozróżnieniem dwóch rozkładów: rozkładu efektów w poszczególnych badaniach i rozkładu efektów populacji. Przedział ufności odzwierciedla rozkład wokół średniej z poszczególnych badań, a przedział wiarygodności rozkład wokół oszacowanej średniej prawdziwego efektu (efektu na poziomie populacji badań).

Analiza w grupach, metaregresja i analiza wrażliwości

Ponieważ celem syntezy ilościowej jest nie tylko oszacowanie jednego wskaźnika, podsumowującego siłę efektu, ale też analizę jego zróżnicowania, sporo miejsca w literaturze dotyczącej metaanalizy poświęca się problemowi niejednorodności efektów (*effect heterogeneity*). Analiza wariancji umożliwia wyliczenie wielu statystyk obrazujących poszczególne aspekty zróżnicowania czy testowania założeń o jednorodności zróżnicowania między badaniami (np. T^2 , Q , I^2). Przykładowo, parametr I^2 pokazuje, w jakiej części różnice w oszacowaniu efektów wynikają z niejednorodności uwzględnionych badań. Możliwe jest także przeprowadzenie analizy wrażliwości, np. badając wpływ poszczególnych badań na wynik łączny. Dodatkową komplikacją w syntezie wyników jest możliwość zróżnicowania efektów w różnych podgrupach. Z tego względu wyniki metaanaliz podaje się osobno w podgrupach badań bądź do analiz statystycznych wprowadza się dodatkowe zmienne, dzielące badania na podgrupy. Możliwe jest także przypisanie badaniom określonych zmiennych (na podstawie wcześniej ustalonych kodów, wykorzystywanych w kodowaniu badań) i uwzględnienie ich w analizie (tzw. metaregresja). Analizy w podgrupach i włączanie zmiennych na poziomie badań mają jednak dwa poważne ograniczenia: po pierwsze, zwykle w metaanalizie liczba badań jest relatywnie niewielka, co ogranicza moc statystyczną (*statistical power*) wykorzystywanych testów statystycznych. Po drugie, i ważniejsze, uzyskane wnioski mają wyłącznie charakter korelacyjny i nie mogą służyć do wnioskowania przyczynowo-skutkowego.

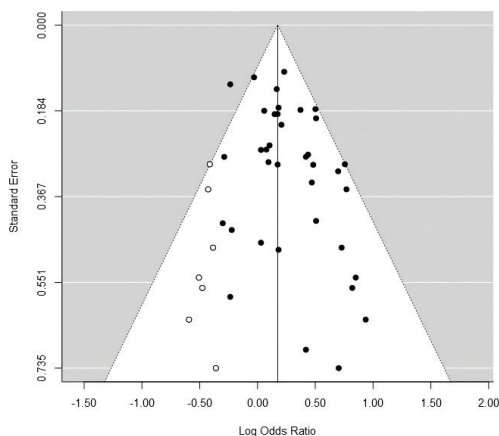
Ocena wybiórczego publikowania wyników (*publication bias*)

Ważnym elementem metaanalizy jest ocena obciążeń wyników wynikająca z błędu publikacji. W badaniach społecznych dąży się zwykle do ograniczenia tzw. wyników fałszywie pozytywnych, czyli błędnego potwierdzenia tezy badacza (błąd I rodzaju). Często mamy jednak do czynienia z sytuacją, gdy słuszne przewidywania nie zostały potwierdzone (błąd II rodzaju). Błąd ten może wynikać ze zbyt małej wielkości próby (zbyt słabej mocy testu statystycznego) bądź, co jest bardziej prawdopodobne,

z mniej lub bardziej uświadomianego unikania publikowania tego rodzaju wyników. Brak uwzględnienia tego rodzaju badań może być źródłem poważnego obciążenia wyników metaanalizy.

Ryzyko pominięcia badania maleje wraz ze wzrostem siły obserwowanego efektu i wzrostem precyzji wyników (na którą wpływa np. wielkość próby czy rzetelność narzędzia pomiarowego). Zależności te wykorzystuje tzw. wykres lejkowy. Opiera się on na założeniu, że zależność między uzyskanymi efektami w poszczególnych badaniach a wielkością prób badawczych (lub wartości błędu standardowego) ma kształt odwróconego lejka. Przykład takiego wykresu przedstawiono poniżej: oś y obrazuje wielkość błędu standardowego, a oś x – zaobserwowaną siłę efektu. W przypadku występowania obciążeń związanych z publikowaniem wyników można się spodziewać w miarę symetrycznego rozkładu w górnej części wykresu i mniej symetrycznego rozkładu dolnej części wykresu, w której znajdują się badania przeprowadzone na mniejszych próbach, gdzie powinien występować większy rozrzut zaobserwowanych wyników. Z takim przypadkiem mamy do czynienia na poniższym wykresie, na którym czarnym kolorem zaznaczono badania opublikowane i uwzględnione w przeglądzie – a kolorem białym badania, które potencjalnie mogłyby zostać opublikowane, ale niedostępne dla autorów przeglądu.

Przykład wykresu lejkowego (*funnel plot*)



Źródło: http://www.metafor-project.org/doku.php/plots:funnel_plot_with_trim_and_fill.

Wykres lejkowy umożliwia jedynie wzrokową ocenę możliwych obciążeń. Do oceny ilościowej można wykorzystać testy dotyczące zależności między błędem standardowym efektu a jego wartością (np. test Begga i Mazumdar, test Eggera) czy wartość wskaźnika *Fail-safe N*. Opracowano także procedury umożliwiające ocenę, na ile błąd publikacji wpłynął na łączny wynik uzyskany w metaanalizie (umożliwia to np. technika *trim and fill*, polegająca na odcinaniu najbardziej skrajnych przypadków i symetrycznym uzupełnianiu brakujących przypadków – jak to zrobiono na powyższym wykresie; zob. Duval, 2005). Techniki tego rodzaju nie umożliwiają oszacowania prawdziwej siły efektu, gdyż opierają się zbyt wielu założeniach. Należy je więc traktować jako jeden ze sposobów analizy wrażliwości uzyskanych wyników. Analizy tego rodzaju są standardowym elementem raportu z metaanalizy, gdyż umożliwiają czytelnikowi samodzielną ocenę wiarygodności uzyskanych wyników.

Bibliografia

Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Taylor & Francis.

Duval, S. J. (2005). The trim and fill method. W: H. R. Rothstein, A. J. Sutton i M. Borenstein (red.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (s. 127–144). Chichester, England: Wiley.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.

Hedges, L. V. i Vevea, J. L. (1998). Fixed-and Random-Effects Models in Meta-Analysis. *Psychological Methods*, 3 (4), 486–504.

Hedges, L.V. i Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6, 203–217.

Jaeschke, R., Cook, D., Guyatt, G. i Bochenek, T. (1998). Ocena artykułów przeglądowych i metaanaliz. *Medycyna Praktyczna*, 1998/09.

Kline, R. B. (2013). *Beyond Significance Testing: Statistics Reform in the Behavioral Sciences*. 2nd ed. Washington, DC: American Psychological Association.

Kusy, M. (2010). *Ilościowa synteza wyników badań pierwotnych – metaanaliza w Statistica w zestawie medycznym*. Pobrane z http://www.statsoft.pl/Portals/0/Downloads/Ilosciowa_synteza_wynikow.pdf

Leśniak, W., Bała, M., Mrukowicz, J., Brożek, J., Jaeschke, R., Gajewski, P. (2008). Przegląd systematyczny i metaanaliza. W: P. Gajewski, R. Jaeschke i J. Brożek, *Podstawy EBM czyli medycyny opartej na danych naukowych dla lekarzy i studentów*. Kraków: Medycyna Praktyczna.

Lipsey, M. W. i Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: SAGE.

Slavin, R. E. i Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31 (4), 500–506.

Glosariusz

Attrition – utrata uczestników badania po ich zarejestrowaniu/zakwalifikowaniu (lub losowym przydziale) do badania. W badaniach powtarzanych (np. panelowych) oznacza „wykruszanie się” uczestników z badania w kolejnych jego falach.

Attrition rate – wskaźnik utraty.

Bias – obciążenie, błąd systematyczny w oszacowaniu lub wnioskowaniu; powoduje przeszacowanie lub niedoszacowanie skutków; obciążenie błędem; stronniczość, tendencyjność.

Blinding – maskowanie (tzw. zaślepienie), zatajenie informacji o rodzaju terapii, przewidziane dla konkretnych uczestników; może być zastosowane do uczestniczących w badaniu lub pracowników zbierających dane (jeśli zastosowane wobec tych obu grup, wtedy badanie jest podwójnie zaślepienie/zamaskowane). Taki zabieg zmniejsza zniekształcenie wyników.

Comparison group – grupa nieobjęta interwencją, która jest porównywana z grupą poddaną interwencji.

Confidence interval (CI) – przedział ufności; zakres wartości, w którym, z zadany­m prawdopodobieństwem i przy założeniu rozkładu normalnego zmiennej, leży prawdziwa wartość określonej statystyki. Przy 95% przedziale ufności 95% niezależnych próbek losowych z populacji przyniosłoby oszacowania, które znajdują się w tym zakresie.

Confirmation bias – tendencja do podkreślania dowodów na poparcie danej hipotezy oraz pomijania dowodów, które ją podważają.

Confounding variable – zmienna zakłócająca (uwikłana) związana z inną zmienną niezależną. Zmienna zakłócająca zniekształca wpływ zmiennej niezależnej na zmienną zależną.

Control group – grupa kontrolna (porównawcza), która w schemacie eksperymentalnym nie jest poddana interwencji.

Controlled clinical trial (CCT) – badanie porównujące dwie lub więcej grup objętych interwencją za pomocą quasi-losowej metody alokacji (daty urodzin lub numery

rekordów) lub metodą alokacji prawdopodobnie losowej (lub prawdopodobnie quasi-losowej), lecz niejednoznacznie określona.

Continuous variable – zmienna, która może przybrać zakres wartości, które mogą być wyrażone w skali numerycznej.

Counterfactual – kontrfaktyczny/alternatywny: hipotetyczny stan sytuacji nieobjęcia interwencją osób, które zostały poddane interwencji (lub odwrotnie).

Detection bias – błąd/tendycyjność wykrywania: systematyczne różnice pomiędzy grupami w zakresie zbierania danych dotyczących wyników/rezultatów.

Dichotomous variable – zmienna dychotomicznie przyjmująca tylko dwie wartości, np. zerojedynekowa.

Effect size (ES) – wielkość efektu; miara siły i kierunku relacji między zmiennymi.

Efficacy – skuteczność; wpływ interwencji w idealnych warunkach (idealnej sytuacji).

Evidence-based policy – polityka prowadzona na podstawie najlepszych, dostępnych dowodów naukowych i ekspertyz w celu podejmowania świadomych, uzasadnionych rzetelną wiedzą decyzji politycznych.

Evidence-based practice – praktyka wykorzystująca najlepsze, dostępne dowody naukowe w połączeniu z praktyczną wiedzą specjalistyczną i ekspercką w celu podejmowania uzasadnionych, świadomych decyzji w odniesieniu do różnych grup zawodowych (np. nauczycieli).

External validity – trafność zewnętrzna, stopień, w jakim wyniki badań można uogólniać na inne populacje.

Forest plot – wykres drzewkowy przedstawiający rozmiary efektu i przedziały ufności z jednego lub większej liczby badań.

Funnel plot – wykres lejkowy; graficzny wykaz wykrywania stronniczości publikacji, innych źródeł obciążeń oraz efektów z badań na małych próbach; bazuje na założeniu, że rozmiary efektu badań są zwykle rozłożone wokół średniej efektu (*mean effect*). Gdy nie ma błędu systematycznego (*bias*), rozkład rozmiaru efektu będzie niewielki na szczycie wykresu (gdzie pokazane są dokładniejsze oszacowania) a szerszy na dole wykresu.

Heterogeneity – niejednorodność, (heterogeniczność); stopień zróżnicowania efektów. Dotyczy różnic pomiędzy badaniami pod względem wyników (statystyczna niejednorodność), populacji i metod (zróżnicowanie metodologiczne).

Heterogeneous population – zbiorowość niejednorodna, niejednorodna.

Homogeneous population – populacja jednorodna.

Homogeneity – jednorodność; podobieństwo wyników danego zbioru badań; założenie (w modelu ze stałym efektem), że wszystkie badania są z tej samej populacji i rozmiar efektu każdego poziomu badania jest oszacowaniem parametru poszczególnej populacji.

Interrater agreement – w statystyce: zgodność oceniania.

Odds ratio (OR) – iloraz szans.

Outcome reporting bias – tendencja do raportowania informacji raczej o wynikach istotnych statystycznie niż zerowych.

p value – wartość p z testów istotności statystycznej, odnosi się do prawdopodobieństwa błędu typu I (odrzuć hipotezę zerową, gdy hipoteza zerowa jest prawdziwa).

PICOC – akronim pomocny w konkretyzacji pytania badawczego poprzez określenie pięciu elementów: Population-Intervention-Comparison-Outcome-Context.

Point estimate – oszacowanie punktowe (estymator punktowy) wartość użyta do oszacowania realnej wartości parametru (np. efektu fenomenu zainteresowania) w większej populacji; oszacowaniom punktowym towarzyszą przedziały ufności.

Protocol – protokół, [tu:] z góry opracowany plan badawczy. Protokół przeglądu systematycznego lub metaanalizy opisuje metody i procedury, które będą zastosowane.

Publication bias – tendencyjna selekcja badań wykazująca systematyczne różnice między badaniami publikowanymi i niepublikowanymi; zazwyczaj odnosi się do większego prawdopodobieństwa publikacji o wynikach statystycznie istotnych i nieobecności zerowych lub negatywnych wyników w publikacjach.

Randomized controlled trial (RCT) – badanie porównawcze, w którym uczestników przydziela się do grup eksperymentalnej i porównawczej lub kontrolnej wyłącznie losowo.

Review – analiza i ocena krytyczna literatury naukowej w danym zakresie (Gough i in., 2012).

Review of reviews – przegląd systematyczny analizujący dane pochodzące z innych przeglądów (Gough i in., 2012).

Traditional literature review – przegląd literatury naukowej niewymagający zastosowania szczegółowych, jawnych i sprawdzalnych metod (w odróżnieniu od przeglądu systematycznego) (Gough i in., 2012).

Bibliografia

Bienias, S., Strzęboszewski, P. i Opalka, E. (red.). (2012). *Ewaluacja. Poradnik dla pracowników administracji publicznej*. Warszawa: Ministerstwo Rozwoju Regionalnego.

Boland, A., Cherry, M. G. i Dickson, R. (2014). *Doing a Systematic Review: A Student's Guide*. London: Sage.

Booth, A. (2001). *Systematyczny Przegląd Badań Naukowych*. W: P. Franaszek (red.). *Informacja naukowa w zdrowiu publicznym* (s. 25–44). Przeł. Barbara Niedźwiedzka. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Booth, A., Papaioannou, D. i Sutton, A. (2012). *Systematic Approaches to a Successful Literature Review*. London: Sage.

Bukowska, E. i Kołasiński, A. (2010). *Systematyczny Przegląd Literatury – szczegółowy opis realizacji*. Pobrane z <http://madeyski.e-informatyka.pl/download/presentations/LaTeXandSLR/SLR.pdf>

Czakon, W. (2011). *Metodyka systematycznego przeglądu literatury*. *Przegląd Organizacji*, 3, 57–61.

Cooper, H. (2010). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. 4th Edition. Thousand Oaks, CA: Sage.

Davies, P. (2000). *The relevance of systematic reviews to educational policy and practice*. *Oxford Review of Education*, 26, 3/4, 365–378.

DIPF (2007). *Knowledge for action – Research Strategies for an Evidence-Based Education Policy*. Symposium during Germany's EU Presidency 28–30 March 2007 in Frankfurt / Main Conference Volume. Pobrane z <http://ice.dipf.de/de/pdf/tagungsdokumentation>

European Commission (2007). *Towards more knowledge-based policy and practice in education and training*. Commission Staff Working Document, 28.08.2007, SEC(2007) 1098, Brussels. Pobrane z <http://www.oecd.org/edu/cei/39442860.pdf>

European Commission (2010). *Communicating research for evidence-based policymaking. A practical guide for researchers in socio-economic sciences and humanities*. Pobrane z http://ec.europa.eu/research/social-sciences/pdf/guide-communicating-research_en.pdf

Federowicz, M. i Sitek, M. (2012). Zmiany i wyzwania w systemie edukacji. *Polityka Społeczna, I*. Warszawa: Instytut badań Edukacyjnych. Pobrane z file:///C:/Users/Uzytkownik/Downloads/federowicz_sitek-badania_polityka%20spoleczna_polityka%20edukacyjna_%20potrzeby%20praktyki.pdf

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8.

Goldacre, B. (2012). *Bad Pharma: How drug companies mislead doctors and harm patients*. Harper Collins Publishers. Kindle Edition.

Gough, D., Tripney, J., Kenny, C. i Buk-Berge, E. (2011). *Evidence Informed Policy in Education in Europe: EIPEE final project report*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. Pobrane z <http://www.eippe.eu/cms/LinkClick.aspx?fileticket=rN52NrA0dbQ%3d&tabid=3212>

Gough, D., Olivier, S. i Thomas, J. (red.). (2012). *An Introduction to Systematic Reviews*. London: Sage.

Gough, D., Olivier, S. i Thomas, J. (2013). *Learning from Research: Systematic Reviews for Informing Policy Decisions. A Quick Guide*. A paper for the Alliance for Useful Evidence. London: Nesta.

Górniak, J. (2008). Sprawne państwo – cykle tworzenia i oceniania polityk/programów publicznych. *Dialog*, 4, s. 61–68.

Górniak, J. (2010). *Ewaluacja oparta na badaniu użyteczności – analiza w kontekście realizacji zasady good governance w wybranych programach operacyjnych*. Pobrane z <http://www.portal.uj.edu.pl/documents/4628317/5fd8f097-8ce8-4ad7-af64-6c386fa6a51e>

Górniak, J. i Mazur, S. (2011). Polityki publiczne oparte na dowodach i ich zastosowanie do rynku pracy. W: J. Górniak i S. Mazur (red.), *W kierunku polityki rynku pracy opartej na*

dowodach (s. 9–34). Warszawa: Obserwatorium Regionalnych Rynków Pracy Pracodawców Rzeczypospolitej Polskiej. Pobrane z http://www.obserwatorium.pracodawcyrp.pl/images/W_kierunku_polityki_opartej_na_dowodach_ost.pdf

Komisja Europejska (2010). *Europa 2020*. Bruksela. Pobrane z http://ec.europa.eu/europe2020/index_pl.htm

Littell, J. H., Corcoran, J. i Pillai, V. (2008). *Systematic Reviews and Meta-Analysis*. Series Title: Pocket Guides to Social Work Research Methods. Oxford: Oxford University Press.

Niedźwiedzka, B. (1998). Praktyka lekarska oparta na wynikach badań naukowych II: Źródła informacji w nurcie Evidence Based Practice. *Medycyna po Dyplomie*, t. 7, nr 5, 11.

Mauer-Górska, B. (bdw). *Strategia Evidence –Based Librarianship (EBL)*. Pobrane z <http://skryba.inib.uj.edu.pl/wydawnictwa/e04/mauer.pdf>

Ministerstwo Infrastruktury i Rozwoju (2014). *Programowanie perspektywy finansowej na lata 2014–2020*. Warszawa: Ministerstwo Infrastruktury i Rozwoju. Pobrane z https://www.mir.gov.pl/aktualnosc/fundusze_europejskie/Documents/Umowa_Partnerstwa_21_05_2014.pdf

Ministerstwo Rozwoju Regionalnego (2012). *Strategia rozwoju kraju 2020*. Warszawa: Ministerstwo Infrastruktury i Rozwoju. Pobrane z http://www.mir.gov.pl/rozwoj_regionalny/Polityka_rozwoju/SRK_2020/Documents/Strategia_Rozwoju_Kraju_2020.pdf

Ministerstwo Rozwoju Regionalnego. Krajowa Jednostka Oceny (2007). *Wytyczne w zakresie ewaluacji programów operacyjnych na lata 2007–2013*. Warszawa.

Moher, D., Liberati, A., Tetzlaff, J. i Altman, D.G. The PRISMA Group (2009). Pobrane z <http://www.bmj.com/content/339/bmj.b2535>

Noblit, G. i Hare, R.D. (1988). *Meta-ethnography: Synthesizing Qualitative Studies*. Newbury Park, CA: Sage.

OECD (2007). *Evidence in Education: Linking research and policy*. Pobrane z <http://www.oecd.org/edu/cei/evidenceineducationlinkingresearchandpolicy.htm>

Olejniczak, K. (2008). *Mechanizmy wykorzystania ewaluacji*. Uniwersytet Warszawski. Centrum Europejskich Studiów Regionalnych i lokalnych. Studium ewaluacji średniookresowych

INTERREG III. Warszawa: Wydawnictwo Naukowe „Scholar”. Pobrane z http://www.euroreg.uw.edu.pl/dane/web_euroreg_publications_files/1251/mechanizmy_wykorzystania_ewaluacji.pdf

Pawson, R. (2006). *Evidence-based Policy: A Realist Perspective*. London: Sage.

Petticrew, M. (2001). Systematic reviews from astronomy to zoology: myths and misconceptions. *BMJ*. Jan 13, 322(7278), s. 98–101. Pobrane z <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1119390/>

Petticrew, M. i Roberts, H. (2006). *Systematic Reviews in the Social Sciences*. Oxford: Blackwell.

Westhorp G., Walker B. i Rogers P. (2012). *Under what circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor? A Realist Synthesis. Protocol*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. Pobrane z <http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=1w2U0N19vac%3D&tabid=3174>

Wildridge, V. i Bell, L. (2002). *How CLIP became ECLIPSE: A mnemonic to assist in searching for health policy/management information*. *Health Information and Libraries Journal*, 19(2), 113–115. Pobrane z <https://ejournals.library.ualberta.ca/index.php/EBLIP/article/viewFile/9741/8144>

Rada Unii Europejskiej (2006). Rozporządzenie Rady (WE) nr 1083/2006 z dnia 11 lipca 2006 r. ustanawiające przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego oraz Funduszu Spójności i uchylające rozporządzenie (WE) nr 1260/1999. W: *Dziennik Urzędowy Unii Europejskiej L 210 z 31.7.2006*, s. 25–30. Pobrane z <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:210:0025:0078:PL:PDF>

Simons, M. (2011). *Guidelines for writing systematic reviews*. Pobrane z <http://www.library.mq.edu.au/libguides/Guidelines%20for%20writing%20systematic%20reviews.pdf>

Spencer, L., Ritchie, J., Lewis, J. i Dillon, L. (2003). *Quality in Qualitative Evaluation: A framework for assessing research evidence*. Government Chief Social Researcher’s Office, London: Cabinet Office.

Szarfenberg, R. (2011). Dowody naukowe jako podstawa polityki społecznej, zarządzania społecznego i pracy socjalnej. *Problemy Polityki Społecznej*, 15, s. 13–28. Pobrane z <http://rszarf.ips.uw.edu.pl/pdf/EBP.pdf>

Walczak, T. (2011). *Słownik terminów statystycznych angielsko-polski, polsko-angielski*. Wydanie 1. Warszawa: C.H. Beck.

Zybała, A. (2012a). *O lepszą jakość polityk publicznych*. Seria: Analizy i Opinie. Nr 127. Warszawa: Instytut Spraw Publicznych. Pobrane z <http://www.isp.org.pl/uploads/analyses/2048186281.pdf>

Zybała, A. (2012b). *Polityki publiczne. Doświadczenia w tworzeniu i wykonywaniu programów publicznych w Polsce i w innych krajach. Jak działa państwo, gdy zamierza/chce/musi rozwiązać zbiorowe problemy swoich obywateli?* Warszawa: KSAP. Pobrane z http://www.ksap.gov.pl/ksap/file/publikacje/polityka_publiczna.pdf

Netografia

The Campbell Collaboration: <http://www.campbellcollaboration.org/>

EPPI-Centre: <http://eppi.ioe.ac.uk/cms/>

CASP (2013) Critical Appraisal Skills Program: Making sense of evidence. <http://www.casp-uk.net/>

Centre for Reviews and Dissemination (2009) Systematic reviews: CRD guidance for undertaking reviews in healthcare: <http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm>

Joanna Briggs Institute Reviewers' Manual (2011): <http://joannabriggs.org/assets/docs/sumari/ReviewersManual-2011.pdf>

JBS Approach: <http://joannabriggs.org/jbi-approach.html#tabbed-nav=JBI-approach>

Social Care Institute for Excellence (2010) SCIE systematic research reviews: Guidelines: <http://www.scie.org.uk/publications/researchresources/rr01.pdf>

PRISMA: <http://www.prisma-statement.org/usage.htm>

PRISMA statement – Checklist 2009: <http://www.prisma-statement.org/2.1.2%20-%20PRISMA%202009%20Checklist.pdf>

Weight of Evidence: <http://www.strath.ac.uk/aer/materials/8systematicreview/unit10/evidence/>

Oprogramowanie dla przeglądów systematycznych i metaanaliz:

- Comprehensive Meta-analysis:
<http://www.meta-analysis.com/index.php>
- EPPI-Reviewer:
<http://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4>
- Meta-analysis made easy:
<http://www.meta-analysis-made-easy.com/download/index.html>
- Practical Meta-Analysis Effect Size Calculator:
http://www.campbellcollaboration.org/resources/effect_size_input.php
- RevMan: <http://tech.cochrane.org/revman>
- TrialStat SRS:
http://www.clinpage.com/article/trialstats_srs_for_systematic_reviews/C12

ISBN 978-83-61693-58-1