



Agencja Oceny Technologii Medycznych i Taryfikacji



Fundusze Europejskie
Wiedza Edukacja Rozwój



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz Społeczny



Opis metodyki przeglądów systematycznych w zakresie interwencji zdrowia publicznego oraz tworzenia programów polityki zdrowotnej

Sylwia Chylak¹, Patrycja Głowik¹, Michał Sawicki¹

¹Dział Programów Polityki Zdrowotnej, Agencja Oceny Technologii Medycznych i Taryfikacji

Warszawa, grudzień 2018

Spis treści

1. Abstrakt.....	4
2. Informacje wstępne	5
2.1. Interwencje oraz działania z zakresu zdrowia publicznego.....	5
3. Metodologia.....	6
3.1. Formułowanie pytania klinicznego.....	6
3.2. Definiowanie problemu zdrowotnego oraz określenie wskaźników epidemiologicznych.....	6
3.3. Scoping	7
3.4. Definiowanie PICOS	7
3.5. Selekcja informacji	8
3.6. Podstawowe źródła danych.....	9
3.7. Dodatkowe, elektroniczne bazy danych	10
3.8. Typy badań	11
3.9. Strategia wyszukiwania	14
3.10. Ocena jakości informacji	15
3.11. Ocena jakości informacji badań pierwotnych.....	15
3.12. Ocena jakości informacji badań wtórnych.....	15
3.13. Dodatkowe elementy brane pod uwagę przy tworzeniu przeglądu systematycznego interwencji z zakresu zdrowia publicznego	16
Zagadnienia etyczne i nierówności	16
Kontekst.....	17
Ciągłość, trwałość.....	18
Stosowalność	19
3.14. Przedstawienie badań włączonych do przeglądu systematycznego i ekstrakcja danych.....	20
3.15. Synteza jakościowa.....	20
3.16. Synteza ilościowa	21
3.17. Efektywność kosztowa interwencji z zakresu zdrowia publicznego	22
4. Organizacja oraz monitorowanie i ewaluacja programów polityki zdrowotnej.....	26
4.1. Opcjonalne technologie medyczne	26
4.2. Stan finansowania interwencji ze środków publicznych w Polsce.....	26
4.3. Warunki realizacji interwencji w ramach programów.....	26
4.4. Monitorowanie i ewaluacja.....	26
5. Bibliografia	32
Aneks 1. Tworzenie strategii wyszukiwania	35
Aneks 2. Skala Jadad	42

Aneks 3.	Narzędzie Cochrane do oceny ryzyka <i>bias</i> oraz rodzaje <i>bias</i>	43
Aneks 4.	Skala AMSTAR.....	45
Aneks 5.	Narzędzie ROBIS	47
Aneks 6.	Skala GRADE.....	51
Aneks 7.	Instrument AGREE.....	52
Aneks 8.	Narzędzie Health Evidence™ – Quality Assessment Tool – Review Articles (dot. badań wtórnych)	54
Aneks 9.	Przykładowa interwencja zależna od kontekstu	56
Aneks 10.	Różnice pomiędzy <i>applicability</i> oraz <i>transferability</i> w zakresie stosowalności	57
Aneks 11.	Rodzaje heterogeniczności badań w metaanalizie.....	58
Aneks 12.	Potencjalna hierarchia efektów danego programu.....	59
Aneks 13.	Główne kryteria ewaluacji programu	60
Aneks 14.	Przykład wykorzystania matrycy logicznej w ewaluacji programu	61
Aneks 15.	Pytania ewaluacyjne w zakresie kryteriów użyteczności i wykonalności celem określenia kierunku ewaluacji	62

1. Abstrakt

Metodyka przeglądów systematycznych w zakresie zdrowia publicznego

Celem publikacji jest przedstawienie zasad tworzenia przeglądów systematycznych w zakresie zdrowia publicznego w oparciu o zasady *Evidence Based Medicine* (EBM) i *Evidence Based Public Health* (EBPH). Informacje te mają służyć tworzeniu efektywnych programów polityki zdrowotnej (PPZ) ukierunkowanych na dobro pacjenta, zapewnienie bezpieczeństwa zdrowotnego, uzyskanie efektów zdrowotnych o największej wartości oraz optymalne wykorzystanie dostępnych zasobów finansowych, kadrowych i rzeczowych.

Publikacja została przygotowana na podstawie przeglądu literatury w dostępnych bazach informacji naukowej oraz na stronach internetowych organizacji zajmujących się tworzeniem przeglądów systematycznych z zakresu zdrowia publicznego. Praca została opracowana przede wszystkim w oparciu o dokumenty metodologiczne w zakresie tworzenia przeglądów systematycznych w zdrowiu publicznym EUnetHTA, Cochrane i *Centre for Reviews and Dissemination* (CRD).

W publikacji opisano metodykę przeprowadzania przeglądów systematycznych w odniesieniu do interwencji zdrowia publicznego, obejmującą formułowanie pytania klinicznego, selekcję informacji, ocenę jakości informacji, sposoby przedstawiania badań i ekstrakcje danych oraz dodatkowe elementy, które są brane pod uwagę w tworzeniu takich przeglądów. W publikacji skupiono się także na sposobie organizacji oraz monitorowania i ewaluacji programów polityki zdrowotnej.

słowa kluczowe: przeglądy systematyczne w zdrowiu publicznym, programy polityki zdrowotnej, programy zdrowotne, zdrowie publiczne, badania naukowe

Abstract

Methodology of systematic reviews in public health

The aim of the present paper is to provide guidance to creating systematic reviews in the field of public health based on the principles of Evidence Based Medicine (EBM) and Evidence Based Public Health (EBPH). This paper intends to help of effective health policy programmes design (PPZ), which aimed at patient welfare, health safety, obtaining the health effects of the highest value and optimal use of available financial, human and material resources.

This paper based on literature review in available biomedical databases and on the websites of organizations involved in the creation of systematic reviews in the field of public health, especially EUnetHTA, Cochrane and Center for Reviews and Dissemination (CRD).

The publication describes a methodology for conducting systematic reviews in the field of public health interventions, including following elements: defining the review question, selecting evidence, assessing quality of evidence, ways of presenting results and analyzing data, and additional elements that are taken into account in creating systematic reviews in the field of public health. The publication also focuses on the organization and monitoring and evaluation of health policy programmes.

key words: systematic reviews in public health intervention, health policy programmes, health programmes, public health, research

2. Informacje wstępne

Grupę docelową odbiorców opracowania, oprócz pracowników Agencji Oceny Technologii Medycznych i Taryfikacji (AOTMiT), stanowią pracownicy jednostek samorządów terytorialnych (JST) odpowiedzialni za wdrażanie i realizację programów polityki zdrowotnej (PPZ), pracownicy Ministerstwa Zdrowia (MZ) i Narodowego Funduszu Zdrowia (NFZ), Rada Przejrzystości¹ (RP). Wytyczne mogą służyć też świadomym roli HTA (*health technology assessment*) lekarzom i pacjentom.

Zgodnie z obowiązującymi przepisami prawa, tj. Ustawą z dnia 27 sierpnia 2004 r. o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych (Dz.U. 2018, poz. 1510 z późn. zm., zwana dalej Ustawą), na Agencji spoczywa obowiązek opiniowania projektów programów polityki zdrowotnej (do 31.12.2014 r. programów zdrowotnych). Programy te odnoszą się do interwencji z zakresu szeroko pojmowanego zdrowia publicznego, definiowanego jako troska o zachowanie i umacnianie zdrowia w wymiarze makrospołecznym, lokalnym i środowiskowym, polegająca na naukowym rozpoznawaniu stanu zdrowia i potrzeb zdrowotnych zbiorowości oraz inicjowaniu skoordynowanych wysiłków instytucji rządowych, samorządowych i pozarządowych w celu osiągnięcia pożądanego standardów zdrowia. Cel ten uzyskuje się przez kontrolę czynników ekologicznych, tworzenie sprzyjających zdrowiu warunków społeczno-ekonomicznych, zapobieganie zakaźnym i niezakaźnym chorobom społecznym, powszechny i równy dostęp do opieki medycznej, a przede wszystkim w wyniku kształtowania nawyków zdrowego stylu życia w ramach edukacji przedszkolnej i szkolnej oraz realizacji programów promocji zdrowia w środowisku ludzi dorosłych. Interwencje z zakresu zdrowia publicznego są zróżnicowane i obejmują działania odnoszące się zarówno do jednostki, jak i całego społeczeństwa.

2.1. Interwencje oraz działania z zakresu zdrowia publicznego

Ocena interwencji z zakresu zdrowia publicznego jest zazwyczaj wieloaspektowa. Interwencje z zakresu zdrowia publicznego są przeważnie opisywane jako zbiór komponentów, np. uwzględniający dietę, ćwiczenia, poradnictwo dotyczące redukcji masy ciała, organizowanie różnorodnych wydarzeń/spotkań o charakterze edukacyjnym dla społeczeństwa oraz zapewnienie dostępu do różnego rodzaju świadczeń jako składowe promowania zdrowia.

Złożoność wynika też ze zróżnicowanej charakterystyki populacji, punktów końcowych, które stanowią niezbędne elementy oceny, a także z faktu, że skuteczność bardzo często zależy od kontekstu prowadzonych działań. Mając powyższe na uwadze, żadna z konkretnych metod oceny nie może zostać uznana za uniwersalną.

Trzeba podkreślić, że proces oceny PPZ jest złożony i wykracza dalece poza metodologię HTA. Dzięki metodologii HTA można jedynie uzyskać odpowiedź na pytanie, czy wdrożenie danych interwencji może w skuteczny, bezpieczny i kosztowo-efektywny sposób przyczynić się do poprawy zdrowia społeczeństwa. Metodologia HTA staje się tym samym swoistą bazą, punktem wyjścia do wdrażania interwencji w życie. Podkreślić należy, że powodzenie realizacji danego PPZ, a tym samym osiągnięcie poprawy zdrowia lokalnej społeczności zależy od szeregu innych czynników i narzędzi, które niejednokrotnie są integralnym elementem PPZ. Mimo to celem niniejszego rozdziału jest jedynie przybliżenie metod HTA stosowanych do oceny interwencji z zakresu zdrowia publicznego. Ze względu na wspomnianą wyżej złożoność interwencji, tradycyjne kryteria tworzenia przeglądów systematycznych są zbyt ciasno zdefiniowane i tylko częściowo odpowiadają na pytania z zakresu zdrowia publicznego.

¹ zespół ekspercki przy Prezesie Agencji, który pełni funkcję opiniodawczo-doradczą, powoływany przez Ministra Zdrowia

3. Metodologia

3.1. Formulowanie pytania klinicznego

Poszukiwanie badań z zakresu zdrowia publicznego jest dużo bardziej złożone od pozyskiwania badań dot. technologii lekowych ze względu na dużo większe rozproszenie literatury.

W związku z szerokim zakresem i wieloaspektowością przeglądów z zakresu zdrowia publicznego istotne jest zaangażowanie społeczeństwa w tworzenie przeglądu systematycznego. W zdefiniowaniu pytania klinicznego pomocne jest przeprowadzenie konsultacji z odbiorcami interwencji (użytkownikami przeglądów) oraz innymi zaangażowanymi podmiotami; ustalenie kontekstu analizowanego aspektu zdrowia publicznego poprzez porozumienie się koordynatorów tworzących przegląd z odpowiednio dobraną grupą doradcą (np. eksperci w danym zakresie zdrowia publicznego, lekarze, stowarzyszenia pacjentów), których opinie mogą przyczynić się do ustalania precyzyjnego przedmiotu analizy (pytania klinicznego). Z dostępnych wyników badań i metaanaliz (1, 2) wynika, że zaangażowanie społeczeństwa w proces tworzenia przeglądu systematycznego z zakresu inicjatyw zdrowia publicznego może korzystnie wpływać na osiągnięte wyniki zdrowotne.

Ogólne rekomendacje dot. przeprowadzania przeglądów systematycznych (3):

1. Specjaliści w zakresie pozyskiwania informacji powinni stanowić integralną część zespołu przygotowującego raport HTA/przegląd systematyczny, od początku prac nad danym opracowaniem.
2. W procesie systematycznego wyszukiwania uwzględnia się wyszukiwanie w kierunku nieopublikowanej literatury celem identyfikacji zarówno nieopublikowanych badań jak i nieopublikowanych danych z opublikowanych badań.
3. Oprócz bazy MEDLINE należy przeszukiwać inne bazy danych takie jak EMBASE i The Cochrane Central Register of Controlled Trials (CENTRAL), celem identyfikacji wszystkich opublikowanych badań w danym zakresie.
4. Należy opracowywać indywidualne strategie wyszukiwania dla każdej bazy, stosując zarówno wyszukiwanie tekstowe oraz jeśli w danej bazie jest dostępne, także po głównych hasłach np. w MeSH².
5. Opracowywane strategie wyszukiwania powinny być poddawane ocenie przez drugą osobę (*peer reviewing*) celem zapewnienia im wysokiej jakości.

Proces wyszukiwania powinien być na bieżąco (w czasie rzeczywistym) dokumentowany i zapisywany w sposób przejrzysty i odtwarzalny.

3.2. Definiowanie problemu zdrowotnego oraz określenie wskaźników epidemiologicznych

Przystępując do wykonania raportu w sprawie zalecanych technologii medycznych, działań przeprowadzanych w ramach programu oraz warunków realizacji dla programów polityki zdrowotnej (dalej „Raport”), należy przedstawić problem zdrowotny, którego ma dotyczyć raport.

Opis problemu zdrowotnego, dokonany w oparciu o wiarygodne źródła informacji, powinien być zwięzły i zawierać podstawowe informacje istotne dla decyzji podejmowanych w dalszej części raportu.

Należy podać definicję jednostki chorobowej wraz z kodem ICD-10 bądź też DSM (4) (*Diagnostic and Statistical Manual of Mental Disorders*) oraz ogólną klasyfikację. Jeżeli przedmiotem raportu jest szczególna populacja docelowa (np. określone stadium zawiązania choroby), należy przedstawić zwięzły opis ogólnej jednostki chorobowej, po czym przejść do szczegółowej charakterystyki docelowego problemu zdrowotnego (np. określony stopień zaawansowania).

Działania w ramach danego problemu zdrowotnego powinny wpisywać się w priorytety zdrowotne określone przez Ministra Zdrowia zgodnie z obowiązującym rozporządzeniem (art. 31a, ust. 2 Ustawy) (5).

W opisie problemu należy uwzględnić etiologię i patogenezę, naturalny przebieg choroby, w tym objawy/zespoły objawów, czynniki ryzyka, rokowania, czynniki wpływające na przebieg choroby, a także obraz kliniczny choroby i jej rozpoznanie, leczenie oraz możliwe do wdrożenia działania profilaktyczne.

Opis problemu zdrowotnego powinien obejmować przyczyny i mechanizm rozwoju choroby oraz czynniki ryzyka jej wystąpienia.

Dodatkowo należy opisać kryteria rozpoznawania, w tym badania niezbędne do postawienia/potwierdzenia diagnozy, opierając się na wiarygodnych źródłach danych, np. rekomendacjach klinicznych. Jeśli w rozpoznawaniu stosowane są

² *Medical Subject Headings*; słownik synonimów używany do indeksowania artykułów dla PubMed

swoiste skale lub testy, należy je scharakteryzować z podaniem punktów odcięcia, informacji o ich walidacji oraz o istotności klinicznej wyników.

Należy opisać naturalny przebieg choroby/zaburzenia, główne objawy, czynniki rokownicze oraz wpływające na przebieg choroby/zaburzenia, skutki choroby/stanu zdrowia w zależności od rodzaju programu: prewencyjny – następstwa niepodjęcia działań prewencyjnych; przesiewowy – skutki choroby wykrywanej; leczniczy – skutki choroby leczzonej; poprawiający jakość życia – skutki niepodjęcia działań poprawiających jakość życia obywateli.

Opis zalecanego leczenia, a w szczególności działań profilaktycznych w ramach danego problemu zdrowotnego, powinien być oparty na aktualnych dowodach naukowych lub pozostawać w zgodzie z aktualnymi wytycznymi klinicznymi.

Ponadto, prawidłowo przedstawiony problem zdrowotny powinien zawierać rzetelnie przygotowany opis aktualnej sytuacji epidemiologicznej (zapadalność, chorobowość i umieralność ze szczególnym zwróceniem uwagi na dane dla polskiej populacji). Każdy problem zdrowotny będzie się charakteryzował inną specyfiką. Opis sytuacji epidemiologicznej powinien być oparty na Mapach potrzeb zdrowotnych (6) publikowanych przez Ministerstwo Zdrowia (tam gdzie jest to możliwe).

Należy także przedstawić w sposób ogólny problem zdrowotny z perspektywy zdrowia publicznego oraz wynikających z niego obciążeń społeczno-ekonomicznych.

3.3. Scoping

Określenie zakresu przeglądu oraz sformułowanie odpowiednich pytań klinicznych stanowi krytyczne stadium w procesie tworzenia przeglądu systematycznego (3). Opracowany schemat planowania służy pomocą zarówno w trakcie przygotowań, jak i realizacji przeglądu. Etap jego planowania poprzedza się analizą potrzeb zdrowotnych oraz sytuacji epidemiologicznej. Odpowiednio zaprojektowana i przeprowadzana procedura ustalania zakresu przeglądu (*scoping search*) stanowi pomocne narzędzie w:

- określeniu jego głównego obszaru,
- identyfikacji istniejących przeglądów systematycznych,
- badaniu zakresu literatury dla konkretnego zagadnienia, bez szczegółowego opisu wyników,
- ocenie, zidentyfikowaniu źródeł niezbędnych do wykonania przeglądu,
- identyfikacji obszarów, dla których brak jest danych w literaturze,
- badaniu zakresu literatury dla konkretnego zagadnienia, bez szczegółowego opisu wyników;
- identyfikacji odpowiednich parametrów przeglądu (zdefiniowaniu grupy docelowej, interwencji, interwencji alternatywnych, efektów i skuteczności);

Rezultaty wstępnego wyszukiwania (*preliminary search*) mogą być przydatne w opracowywaniu właściwej strategii wyszukiwania (generowanie listy słów kluczowych z odnalezionych istotnych publikacji, które mogą zostać użyte w budowaniu strategii wyszukiwania).

Proces *scopingu* stanowi specyficzny rodzaj przeglądu literatury oraz dowodów naukowych, oparty o 5 etapów: identyfikację pytania badawczego, identyfikację i selekcję badań, opis badań, zestawienie i prezentację wyników oraz ewentualne konsultacje. *Scoping review* różni się od przeglądu systematycznego. Do podstawowych cech należy: szeroki zakres pytania badawczego, kryteria włączenia/wyłączenia mogą być zdefiniowane *post hoc*, jakość dowodów nie jest podstawowym priorytetem, głównie jakościowa synteza wyników. Proces *scopingu* nie obejmuje oceny jakości odnalezionych badań, co ogranicza syntezę danych i interpretację wyników.

Podstawowym zadaniem jest tutaj sformułowanie pytania klinicznego lub pytań klinicznych. Zdefiniowanie pytania klinicznego stanowi punkt wyjścia w ustaleniu zakresu przeglądu, wpływa na kształt projektu strategii wyszukiwania oraz ostatecznie na wyniki przeprowadzanego przeglądu.

3.4. Definiowanie PICOS

Analiza problemu decyzyjnego pozwala docelowo poprawnie zbudować strategię wyszukiwania według schematu PICO (*population, intervention, comparison, outcome*):

- populacja, w której dana interwencja będzie stosowana (P);
- proponowana interwencja (I);
- proponowane komparatory (C);

- efekty zdrowotne, czyli punkty końcowe, względem których oceniana będzie efektywność kliniczna (O).

W formułowaniu pytania klinicznego możliwe jest zastosowanie schematu PICOS (populacja, w której dana interwencja ma być stosowana, interwencja, interwencja porównawcza, oceniane efekty interwencji, typ badań), mimo to mogą być potrzebne pewne modyfikacje np. waga kontekstu w jakim dana interwencja jest stosowana może skutkować koniecznością poszerzenia schematu o dodatkowe „C” (*context*) – PICOCS. Inną modyfikacją jest schemat PICOTS, poszerzający schemat podstawowy o zakres czasowy „T” (*timing*) oraz warunki/kontekst realizacji „S” (*setting*). Kryterium *setting* uwzględnia m. in.: aspekty organizacyjne, kluczowe czynniki towarzyszące mające wpływ na efekt końcowy (*conceptual factors*), działania konkurencyjne, wymagane zasoby, lokalizację geograficzną, warunki kliniczne oraz finansowe) (7).

- Populacja – w przeciwieństwie do badań klinicznych w badaniach z zakresu zdrowia publicznego populację docelową stanowią grupy, takie jak np. młodzież szkolna, osoby zamieszkujące szczególnie rejon, nie zaś poszczególne jednostki.
- Interwencja i komparatory – interwencje z zakresu zdrowia publicznego są często scharakteryzowane jako zbiór komponentów. Tego typu interwencje są określane mianem kompleksowych, jako że poszczególne składowe mogą działać niezależnie od siebie lub w powiązaniu i określenie „aktywnego składnika” może być dużo trudniejsze niż w innych obszarach badań. Jeśli pytanie badawcze dotyczy właśnie złożonych interwencji, ważne jest podjęcie decyzji czy w danym przypadku znajdzie zastosowanie rozdzielanie poszczególnych elementów. Czasem wykonanie takiego zabiegu może skutkować zgubieniem sensu interwencji.
- Zastosowanie teorii do opisu działania złożonej interwencji jest uważane za dobrą praktykę, gdyż może przyczynić się do wyjaśnienia skuteczności (lub też jej braku) danej interwencji.
- Punkty końcowe – interwencje z zakresu zdrowia publicznego często wywierają bardzo szeroki wpływ na osoby otrzymujące dane interwencje i jedne efekty mogą być uważane za bardziej istotne od innych. Pytanie o to, które z punktów końcowych powinny być priorytetowe może być przedyskutowane z grupą doradców np. ekspertów. Zbadanie dlaczego i w jaki sposób dana interwencja działa, będzie wymagało sformułowania pytań o proces dostarczania danej interwencji i sposób jej wdrożenia. Pomimo tego że interwencje z zakresu zdrowia publicznego wpływają generalnie na stan zdrowia ludności, wyniki wskazujące na poprawę mogą maskować różnice pomiędzy poszczególnymi grupami. Pomocą w rozwiązaniu tego problemu może być narzędzie PROGRESS (8) (narzędzie, którego zadaniem jest określenie, czy planowana interwencja uwzględnia nierówności wynikające z różnorodności, populacji np. pochodzenie etniczne, wykonywany zawód, płeć itp.)
- Kontekst
- Rodzaj badań

W kontekście ewaluacji interwencji zdrowia publicznego ważne jest uwzględnienie nie tylko informacji o skuteczności. Konieczne jest rozpatrzenie szerszej problematyki formułowania problemu klinicznego:

- skuteczności w określonych warunkach;
- jak dana interwencja jest realizowana;
- co usprawnia/wspomaga lub utrudnia jej wpływ;
- jaki proces wdrażania przeprowadzono;
- podejście do oceny interwencji złożonych;
- jakie czynniki mogły wpłynąć na jej skuteczność lub brak, zastosowanie wyników w lokalnych warunkach.

3.5. Selekcja informacji

Wielodyscyplinarny charakter zdrowia publicznego oznacza, że badania mogą pochodzić z wielu różnych obszarów i być odnalezione w wielu różnych elektronicznych bazach danych. Trudności dotyczące poszukiwania badań naukowych narastają w związku z istnieniem nieprecyzyjnej i ciągle zmieniającej się terminologii. Kryteria stosowane przy wyborze badań powinny przede wszystkim odzwierciedlać zadane pytanie/a kliniczne dla przeprowadzanego przeglądu, aniżeli z góry zdefiniowaną hierarchię tych badań. Decyzje dotyczące typów badań, które zostaną włączone, będą miały wpływ na kolejne etapy przeglądu – w szczególności na proces wyszukiwania, ocenę ryzyka błędu i analizy (szczególnie w przypadku metaanaliz). Z tego względu poszukiwanie publikacji z zakresu zdrowia publicznego może być bardzo złożonym zadaniem, wymagającym zastosowania innych metod, niż tylko przeszukiwanie elektronicznych baz danych w celu odnalezienia dowodów naukowych (9).

3.6. Podstawowe źródła danych

W ramach analizy klinicznej przeprowadza się systematyczne wyszukiwanie wszystkich danych naukowych, które dotyczą rozpatrywanego problemu.

W pierwszej kolejności wyszukuje się i przedstawia wyniki już istniejących, niezależnych raportów oceny technologii medycznych (raporty HTA), przeglądów systematycznych oraz metaanaliz dostępnych w:

- Bibliotece Cochrane (*Cochrane Library*);
- bazie MEDLINE;
- bazie EMBASE.

Kolejnym etapem analizy klinicznej powinno być przedstawienie wniosków płynących z odszukanych opracowań wtórnych oraz omówienie ograniczeń odnalezionych prac, w szczególności w kontekście celu i zakresu przeprowadzanej analizy klinicznej. Odnalezione opracowania można również wykorzystać jako źródło informacji o praktyce analitycznej w danym problemie decyzyjnym. Analizę kliniczną opartą tylko na wynikach odnalezionych przeglądów systematycznych i metaanaliz można przeprowadzić, jeżeli odnaleziony przegląd lub przeglądy są systematyczne, aktualne i odpowiadają na postawione pytanie badawcze, a metodyka spełnia wymogi jakościowe. W celu weryfikacji aktualności przeglądu przeprowadza się przeszukiwanie baz informacji medycznej ukierunkowane na identyfikację badań opublikowanych w okresie późniejszym niż objęty wyszukiwaniem w opublikowanym przeglądzie. Jakość odnalezionych przeglądów systematycznych ocenia się z wykorzystaniem stosownego narzędzia.

W przypadku nieodnalezienia opracowań wtórnych wykonuje się przegląd systematyczny badań pierwotnych, tj. odnalezienie wszystkich badań, które spełniają kryteria włączenia do analizy.

Podstawowymi bazami do poszukiwania badań pierwotnych są:

- Biblioteka Cochrane (*Cochrane Library*);
- MEDLINE;
- EMBASE.

W uzasadnionych przypadkach zalecane jest przeszukanie także innych baz informacji medycznej, zgodnie z wytycznymi EUnetHTA (10).

Należy zaznaczyć, że dla wielu interwencji w zakresie promocji zdrowia oraz zdrowia publicznego, kontrolowane próby kliniczne z randomizacją (RCT, *randomized controlled trial*) mogą być niedostępne (ze względu na trudności w ich przeprowadzeniu lub aspekty etyczne). Klasterowe badania randomizowane (*cluster randomized trials*) są coraz częściej wykorzystywane w zdrowiu publicznym, gdyż niektóre interwencje skierowane są do grup (inaczej klastrów, np. społeczności lokalne, pacjenci, uczniowie etc.), a nie pojedynczych osób. Badania te mogą stanowić wartościowe dowody, pod warunkiem poddania randomizacji wystarczającej liczby uczestników (ma to na celu zapewnienie równomiernego rozkładu między grupami potencjalnych czynników zakłócających). Koniecznym jest także poddanie konsultacjom z ekspertami klinicznymi technologii medycznych oraz działań z zakresu zdrowia publicznego, które mają być realizowane w ramach PPZ.

W uzasadnionych przypadkach sugeruje się uzupełnienie przeszukiwania baz informacji medycznych poprzez skorzystanie z innych źródeł, w tym (zgodnie z wytycznymi HTA Agencji):

- referencji bibliograficznych zawartych w publikacjach dotyczących oceny efektywności klinicznej oraz praktycznej;
- rejestrów badań klinicznych (co najmniej 2 rejestry; wymagane jest przeszukanie rejestru *clinicaltrials.gov* oraz *clinicaltrialsregister.eu*, dodatkowo *ICTRP Search Portal*) w celu odnalezienia badań zakończonych, ale nieopublikowanych;
- niesystematyczne wyszukiwanie danych publikowanych w czasopismach specjalistycznych zajmujących się ocenianą technologią i nieindeksowanych w wykorzystywanych bazach informacji medycznej;
- kontaktowanie się z autorami badań klinicznych, m.in. w celu uzyskania i włączenia do analizy szczegółowych danych niepublikowanych;
- wykorzystanie wyszukiwarek internetowych;
- konsultacje z podmiotami odpowiedzialnymi, w szczególności w zakresie informacji o zdarzeniach/działaniach niepożądanych (na podstawie okresowych raportów o bezpieczeństwie; *periodic safety update report*; PSUR);

- wykorzystanie danych z dokumentacji rejestracyjnej leku dostępnej na stronach internetowych agencji rejestracyjnych, tj. Urzędu Rejestracji Produktów Leczniczych, Wyrobów Medycznych i Produktów Biobójczych (URPL), Europejskiej Agencji Leków (EMA), Agencji Żywności i Leków (FDA), w tym analiza komunikatów dotyczących bezpieczeństwa stosowania danej technologii.

Dane odnoszące się do skuteczności klinicznej i efektywności kosztowej powinny pochodzić z wiarygodnych i rzetelnych źródeł danych. W każdym przypadku ocena efektywności powinna opierać się na danych naukowych cechujących się najwyższym poziomem wiarygodności (10).

3.7. Dodatkowe, elektroniczne bazy danych

W tabeli poniżej (Tabela 1) przedstawiono bazy danych, które mogą być przydatne w ocenie interwencji z zakresu zdrowia publicznego, z podziałem na wybrane dziedziny (11). Tabela ta nie wyczerpuje zbioru baz danych. Należy pamiętać, że wybór odpowiednich baz danych powinien być zawsze dostosowany do pytania klinicznego.

Tabela 1. Przykładowe bazy danych do wykorzystania w ocenie interwencji z zakresu zdrowia publicznego/tworzeniu PPZ

Kategoria	Bazy danych
Bazy przeglądów systematycznych i metaanaliz dot. zdrowia publicznego	<ul style="list-style-type: none"> <i>Health Evidence</i> (https://www.healthevidence.org/search.aspx)
Bazy danych związane z psychologią	<ul style="list-style-type: none"> PsycARTICLES (via EBSCOhost), PsycInfo (via EBSCOhost), HaPI (<i>Health and Psychosocial Instruments</i>)
Biomedyczne bazy danych	<ul style="list-style-type: none"> CINAHL (<i>Cumulative Index to Nursing and Allied Health literature</i>) – pielęgniarstwo i zagadnienia pokrewne LILACS (bireme.br), <i>Web of Science</i>, Medline, EMBASE, Cochrane Library, <i>Combined Health Information Database (CHID)</i>, SCOPUS, Trip DataBase, Research Gate (do zastosowania również w innych dziedzinach jak np. psychologia, socjologia etc.) OpenTrials FDA (dot. badań klinicznych)
Socjologiczne bazy danych	<ul style="list-style-type: none"> Sociofile, <i>Sociological Abstracts</i>, <i>Social services Abstracts</i>, <i>Social Science Citation Index</i>, <i>Social Policy and Practice</i>, <i>Social Care Online</i>
Bazy danych związane z edukacją	<ul style="list-style-type: none"> ERIC (<i>Educational Resources Information Center</i>), C2-SPECTR (<i>Campbell Collaboration Social, Psychological, Educational and Criminological Trial Register</i>), REEL (<i>Research Evidence in Education Library, EPPI-Centre</i>),

	<ul style="list-style-type: none"> • RDRB (<i>Research and Development Resource Base</i>)
Bazy danych dotyczące aktywności fizycznej	<ul style="list-style-type: none"> • SPORTDiscus, • WHO <i>European Database on Nutrition, Obesity and Physical Activity</i> (NOPA)
Bazy danych dotyczące rehabilitacji	<ul style="list-style-type: none"> • <i>Physiotherapy Evidence Database</i> (PEDro), • RehabData, • <i>Rehabilitation Reference Center</i>
Bazy danych dotyczące zdrowia publicznego	<ul style="list-style-type: none"> • TRoPHI (<i>Trials Register of Promoting Health Interventions</i>), • DoPHER (<i>Database of Promoting Health Effectiveness Reviews via EPPI-CENTER</i>), • Public Health electronic Library (NICE), • DARE (<i>Database of Abstracts of Reviews of Effectiveness</i>), • CDC <i>Community Health Improvement Navigator – Database of Interventions</i>. • <i>Global Health</i> via EBSCOhost
Inne	<ul style="list-style-type: none"> • POPLINE (zdrowie populacji, planowanie rodziny), • <i>Enviroline Database</i>, • AMED (<i>Alied and Complementary Medicine Databases</i>; pielęgniarstwo, nauki pokrewne do zdrowia), • MANTIS (<i>Manual Alternative and Natural Therapy Index System</i>; medycyna alternatywna), • OTseeker (<i>Occupational Therapy Systematic Evaluation of Evidences</i>), • TOXNET (toksykologia i zdrowie środowiskowe), • BNI (<i>British Nursing Index</i>; pielęgniarstwo i położnictwo)
Jakościowe	<ul style="list-style-type: none"> • <i>Database of Interviews on Patient Experience</i> (DIPEX), • UK <i>Data Service</i> (baza danych ekonomicznych i społecznych)
Dotyczące systemu ochrony zdrowia	<ul style="list-style-type: none"> • Health Systems Evidence (HSE)
Przeglądy systematyczne	<ul style="list-style-type: none"> • Medline, • EMBASE, • <i>Cochrane Database of Systematic Reviews</i>, • <i>The Campbell Collaboration Library of Systematic Reviews</i> • PROSPERO (międzynarodowy rejestr przeglądów)

Źródło: opracowanie własne

Poszerzone przeszukiwanie baz danych stanowi tylko częściowe rozwiązanie identyfikowania odpowiednich badań. Badania mogą pojawiać się w książkach, jako rozdziały, jako dokumenty robocze, w postaci raportów i tradycyjne wyszukiwanie w bazach danych może być niewystarczające do odnalezienia powyższych dokumentów.

3.8. Typy badań

Zdrowie publiczne to szeroko zdefiniowane działania, które można oceniać przy użyciu różnorodnych metod i badań. Nie istnieje uniwersalna metodologia, która pozwoliłaby na rzetelną ocenę wszystkich działań zdrowia publicznego zgodnych z zasadami EBM. Odpowiedni wybór typu badań włączanych do przeglądu systematycznego ma wpływ na ocenę kompletności przeglądu, jakości wyszukiwania oraz syntezę wyników. Badanie skuteczności interwencji w zakresie zdrowia publicznego nigdy nie odbywa się w ściśle kontrolowanych warunkach oraz wymaga uwzględnienia wpływu większej ilości

czynników. Kryteria wyboru typu badań powinny odzwierciedlać odpowiedź na postawione w przeglądzie pytanie badawcze (9).

Podstawą pomiaru skuteczności interwencji z zakresu zdrowia publicznego jest ocena efektywności (*effectiveness*) – stopnia, w jakim osiągnięto zakładane zmiany w danej populacji. Pojęcie to odnosi się do osiągania efektów zdrowotnych w warunkach rzeczywistych/naturalnych. Inną metodą oceny interwencji prozdrowotnej jest ocena skuteczności klinicznej (*efficacy*), a więc stopnia, w jakim dana procedura czy sposób leczenia przynosi korzyści w idealnych warunkach eksperymentalnych. Stosuje się metody jakościowe i ilościowe pomiaru skuteczności rozumianej jako *efficacy* i *effectiveness*. Kryteria stosowane do wyboru odpowiedniej metodologii badań, powinny być kompatybilne z danym problemem badawczym, stanowić odpowiedź na zadane w przeglądzie pytanie badawcze, a nie być konsekwencją z góry określonej hierarchii doniesień naukowych. Niemniej pewien zakres dowodów naukowych ze względu na ich metodologię i jakość, są preferowane i zwiększają wiarygodność interwencji. Decyzja o włączeniu do kolejnej fazy przeglądu danego badania, powinna być poprzedzona oceną jakości i przydatności dowodu.

Poniżej przedstawiono ogólną hierarchię badań wykorzystywanych w zdrowiu publicznym (9):

- Systematyczny przegląd dowodów naukowych (*systematic review*).
- Kontrolowana próba kliniczna z randomizacją (*randomised controlled trials, RCT*).
 - Dowody naukowe w przypadku których istnieje możliwość losowego przydzielania uczestników do interwencji oraz porównania wyników z grupą, która nie otrzymywała interwencji, stanowią użyteczne źródło dowodów skuteczności w zdrowiu publicznym. Jednakże dla wielu działań promocji zdrowia i zdrowia publicznego RCT mogą nie być dostępne, ze względu na problemy, takie jak studium wykonalności³ i etyki.
- Klasterowa kontrolowana próba kliniczna z randomizacją (*cluster-RCT*, klasterowe badanie RCT).
 - Klasterowe badania RCT są coraz częściej wykorzystywane w dziedzinie zdrowia publicznego. Czasami trudno przeprowadzić badania RCT ze względu na relatywnie bliski kontakt potencjalnych badanych między sobą i możliwość wymiany informacji lub istotne różnice w interwencji. W takich sytuacjach można randomizować nie poszczególnych pacjentów, ale całe istniejące już grupy: praktyki lekarskie, oddziały, populacje, klasy (w przypadku dzieci szkolnych) – jest to *cluster randomization*. Losowy przydział do grupy badanej jest tutaj zachowany, ale na nieco „wyższym” poziomie niż tylko pojedynczy pacjent. Jest to przydatna metoda do wykorzystania w dziedzinie zdrowia publicznego. Przykładem takiego badania jest ocena skuteczności, dostępności i jakości usług medycznych oraz programów profilaktycznych realizowanych na różnych populacjach/obszarach. RCT z *cluster* mogą wносить cenne dowody dla dziedziny zdrowia publicznego, o ile w próbie została uwzględniona wystarczająca liczba losowo wybranych jednostek (np. klastrów szkół, społeczności), w celu zapewnienia równomiernego rozkładu potencjalnych czynników pomiędzy grupami.
- Badania bez randomizacji (*non-randomized controlled trials, Non-RCT*), które mogą stanowić przydatne źródło informacji nt. skuteczności badanej interwencji. Są to badania obserwacyjne z grupą kontrolną oraz badanie opisowe (11).

W celu przedstawienia badań można zestawić w tabelach wszystkie dane odnoszące się do określonego problemu klinicznego. Wykaz ten powinien zawierać liczbę i typ włączonych badań oraz charakterystykę każdego z nich, uwzględniającą: okres obserwacji, liczbę ośrodków wykonujących badanie, listę sponsorów, liczebność badanej próby, charakterystykę pacjentów, szczegóły interwencji i uzyskane wyniki oraz inne informacje, które mają znaczenie dla oceny wiarygodności zewnętrznej.

Dla każdego badania włączonego do analizy powinno się przedstawić zwięzłą charakterystykę zamieszczoną w tabeli. Zestawienie powinno być wykonane w oparciu o wyżej opisaną hierarchię dowodów naukowych w zdrowiu publicznym oraz nomenklaturę zamieszczoną w Tabeli 2 wraz z podaniem typu włączonego badania.

Tabela 2. Klasyfikacja doniesień naukowych

Typ badania	Rodzaj badania	Opis podtypu
	IA	Metaanaliza na podstawie wyników przeglądu systematycznego RCT

³ oznacza ocenę i analizę potencjału projektu badania

Typ badania	Rodzaj badania	Opis podtypu
Przeгляд systematyczny RCT	IB	Systematyczny przegląd RCT bez metaanalizy
Badanie eksperymentalne	IIA	Poprawnie zaprojektowana kontrolowana próba kliniczna z randomizacją (RCT), w tym pragmatyczna próba kliniczna z randomizacją (<i>pragmatic randomized controlled trial</i> , pRCT)
	IIB	Poprawnie zaprojektowana kontrolowana próba kliniczna z pseudorandomizacją
	IIC	Poprawnie zaprojektowana kontrolowana próba kliniczna bez randomizacji (<i>clinical controlled trial</i> , CCT)
	IID	Badanie jednoramienne
Badanie obserwacyjne z grupą kontrolną	IIIA	Przeгляд systematyczny badań obserwacyjnych
	IIIB	Poprawnie zaprojektowane prospektywne badanie kohortowe z równoczesną grupą kontrolną
	IIIC	Poprawnie zaprojektowane prospektywne badanie kohortowe z historyczną grupą kontrolną
	IIID	Poprawnie zaprojektowane retrospektywne badanie kohortowe z równoczesną grupą kontrolną
	IIIE	Poprawnie zaprojektowane badanie kliniczno-kontrolne (retrospektywne)
Badanie opisowe	IVA	Seria przypadków – badanie pretest/posttest
	IVB	Seria przypadków – badanie posttest
	IVC	Inne badanie grupy pacjentów
	IVD	Opis przypadku
Opinia ekspertów	V	Opinia ekspertów w oparciu o doświadczenie kliniczne oraz raporty paneli ekspertów

Źródło: Wytyczne przeprowadzania Oceny Technologii Medycznych (HTA). Załącznik do Zarządzenia nr 40/2016 Prezesa Agencji Oceny Technologii Medycznych i Taryfikacji z dnia 13 września 2016 r. w sprawie wytycznych oceny świadczeń opieki zdrowotnej

Wartość materiałów dowodowych na każdym poziomie hierarchii zależy przede wszystkim od jakości metodycznej włączonych badań oraz spełnienia warunków oceny technologii medycznych (10).

Plan postępowania w procesie ekstrahowania danych z wyselekcjonowanych badań winien określać:

- rodzaje informacji wypisywanych z publikacji,
- liczbę osób dokonujących ekstrakcji i ich inicjały,
- formularz ekstrakcji danych.

Dodatkowo pod uwagę bierze się szereg innych elementów, które nie są uwzględniane w trakcie ekstrahowania danych z badań klinicznych: integralność interwencji, heterogeniczność, zagadnienia etyczne, związane w równością i nierównością, ciągłość, kontekst oraz stosowalność wyników (wyjaśnione w rozdz. 3.13 publikacji).

3.9. Strategia wyszukiwania

Tworzenie strategii wyszukiwania zostało szczegółowo opisane w Aneks 1 niniejszego opracowania. Niemniej jednak, dodatkowo w celu zredukowania błędu związanego z charakterystyką populacji (pochodzenie etniczne, warunki klimatyczne, uwarunkowania genetyczne, lokalna sytuacja epidemiologiczna) oraz zwiększenia specyficzności przeglądu systematycznego rekomendowane jest wykonywanie jak najbardziej czułego wyszukiwania danych, które jednocześnie będzie miało na celu zredukowanie błędu związanego z uwarunkowaniami geograficznymi (13).

Strategia wyszukiwania powinna zostać opracowana odpowiednio do zdefiniowanego problemu decyzyjnego oraz być zgodna z zaleceniami *Cochrane Handbook* oraz *Centre for Reviews and Dissemination* (CRD) wskazanymi przez EUnetHTA (3), dot. prawidłowego przeprowadzenia przeglądu systematycznego. Zaleca się stosowanie strategii wyszukiwania o możliwie najwyższej czułości. Jedynie w przypadku dużej liczby trafień można zwiększać swoistość przeszukiwania kosztem jego czułości. Jeżeli w różnych bazach zastosowano strategie różniące się istotnie czułością, wymagane jest uzasadnienie takiego postępowania. Kryteria wyszukiwania powinny uwzględniać elementy schematu PICOS. W strategii wyszukiwania nie zaleca się stosowania słów kluczowych odnoszących się do punktów końcowych.

Efektem końcowym wyszukiwania powinno być zgromadzenie wszystkich dostępnych badań i danych niezbędnych do wykonania rzetelnej oceny skuteczności klinicznej i praktycznej oraz profilu bezpieczeństwa przedmiotowej technologii medycznej lub działania w zakresie zdrowia publicznego.

Proces wyszukiwania danych i informacji szczegółowo opisuje się, tak aby można było ocenić jego poprawność oraz umożliwić jego powtórzenie w przypadku weryfikacji analizy (3). Przedstawienie wyników wyszukiwania powinno zawierać następujące informacje:

- słowa kluczowe i deskryptory zastosowane w czasie wyszukiwania;
- użyte operatory logiki Boole'a;
- użyte filtry;
- zastosowane ograniczenia (limity) np. język;
- przedział czasowy objęty wyszukiwaniem/datę przeprowadzenia ostatniego wyszukiwania;
- liczbę odnalezionych rekordów oddzielnie dla poszczególnych zapytań (kwerend) użytych w strategii wyszukiwania.

Proces weryfikacji, mający na celu ocenę, czy odnalezione doniesienia naukowe spełniają kryteria włączenia do analizy, przeprowadza się etapowo. Pierwszy etap obejmuje selekcję na podstawie tytułów i streszczeń, a dalsze – na podstawie pełnych tekstów publikacji. Selekcja badań powinna być dokonywana na podstawie przyjętych przed rozpoczęciem wyszukiwania kryteriów włączenia i wyłączenia, zgodnych ze zdefiniowanym schematem PICOS (protokół przeglądu systematycznego).

W sytuacji gdy populacja docelowa zdefiniowana na etapie analizy problemu decyzyjnego nie odpowiada próbie ocenianej w odnalezionym materiale dowodowym, dopuszcza się przeprowadzenie analizy klinicznej w populacji zbliżonej do docelowej. W takiej sytuacji można przedyskutować potencjalny wpływ różnic pomiędzy populacjami na wyniki uzyskane w analizie klinicznej.

Selekcja badań pierwotnych powinna dotyczyć publikacji w językach: angielskim, polskim, a w uzasadnionych przypadkach – także w innych.

Na wszystkich etapach selekcja badań w ramach przeglądu systematycznego powinna być wykonywana przez co najmniej dwóch analityków, którzy pracują niezależnie. W tym celu podaje się stopień zgodności wewnętrznej (14) między analitykami dokonującymi selekcji na etapie analizy pełnych tekstów publikacji (np. w postaci współczynnika *Kappa*

Cohena). Preferowaną metodą rozstrzygnięcia niezgodności jest uzyskiwanie konsensusu. W odpowiednich miejscach opracowania podaje się inicjały analityków wykonujących poszczególne zadania.

Analiza powinna przejrzysto informować o liczbie dostępnych doniesień naukowych na poszczególnych etapach wyszukiwania i selekcji badań. Proces prowadzący do ostatecznej selekcji należy przedstawić w postaci diagramu zgodnego z zaleceniami PRISMA (3). Powinno się także przedstawić przyczyny wykluczenia badań na poszczególnych etapach selekcji oraz szczegółowe przyczyny wykluczenia – w przypadku publikacji ocenianych na podstawie pełnego tekstu.

Wszystkie skale i kwestionariusze powinny być przedstawione w załącznikach do przeglądu systematycznego.

3.10. Ocena jakości informacji

Ocena jakości badań w zakresie zdrowia publicznego (w szczególności badań pierwotnych) może być utrudniona, ze względu na duże zróżnicowanie rodzaju tych badań (15). Przed wyżej wspomnianą oceną rozważa się kryteria, które powinny być uwzględnione w ocenie jakości już na etapie planowania przeglądu. Kryteria te będą zależne od rodzaju badań uwzględnionych w przeglądzie.

3.11. Ocena jakości informacji badań pierwotnych

W przypadku badań z randomizacją można przeprowadzić ocenę ich wiarygodności zgodnie z procedurą oceny ryzyka błędu systematycznego opisaną w *Cochrane Handbook* (16). Jednym z najpopularniejszych narzędzi oceny jakości badań klinicznych jest skala Jadad (17) (celem ustalenia minimalnego poziomu badań włączonych do przeglądu systematycznego/metaanalizy; Aneks 2). Badanie, zgodnie z daną skalą, może otrzymać wartości od 0 (niska jakość) do 5 punktów (najwyższa jakość). Jednak *Cochrane Handbook for Systematic Reviews of Interventions* nie zaleca jej samodzielnego stosowania ze względu na pominięcie w niej tzw. *allocation concealment* (ukrycia kodu alokacji) i zwrócenia uwagi jedynie na raportowanie, a nie faktyczne przeprowadzenie badania klinicznego (18). Dlatego też zalecane jest stosowanie ww. skali wraz z narzędziem *Cochrane* do oceny ryzyka błędu systematycznego (16) (odpowiedź negatywna na sformułowane pytania odnośnie poszczególnych kryteriów oznacza wysokie ryzyko błędu; Aneks 3).

Badania prospektywne z grupą kontrolną i randomizacją można oceniać za pomocą narzędzia *Cochrane Collaboration* dla badań z randomizacją, w innych przypadkach (badania bez randomizacji lub badania retrospektywne) za pomocą kwestionariusza *Newcastle-Ottawa Scale* (NOS) (19), natomiast badania jednoramienne w skali NICE (20).

3.12. Ocena jakości informacji badań wtórnych

Przeglądy systematyczne z zakresu zdrowia publicznego mogą opierać się zarówno o wyniki pochodzące z badań pierwotnych, jak i wyniki pochodzące z innych przeglądów systematycznych z interesującego nas obszaru tematycznego.

O ile włączanie do przeglądu systematycznego badań pierwotnych nie budzi żadnych wątpliwości czy kontrowersji, o tyle tworzenie przeglądów systematycznych przeglądów systematycznych literatury jest sprawą relatywnie nową. Przegląd systematyczny przeglądów systematycznych może być bardzo przydatny w przypadku, gdy pytanie kliniczne jest bardzo szerokie, a ponadto istnieją przeglądy systematyczne dotyczące tego zagadnienia. W teorii przeglądy systematyczne włączone do przeglądu systematycznego powinny pokryć większość dostępnych badań pierwotnych. Biorąc pod uwagę często występujące ograniczenia czasowe – takie postępowanie uznać należy za prawidłowe.

W kierunku oceny metodologicznej jakości przeglądów systematycznych dot. problematyki zdrowia stosuje się narzędzia: *Assessment of Multiple Systematic Reviews* (AMSTAR) (21) czy *Health EvidenceTM – Quality Assessment Tool – Review Articles* (22) (Aneks 8). Należy mieć na uwadze, że zgodnie z aktualnym stanowiskiem *National Collaborating Centre for Methods and Tools* (23) warto rozważyć zastosowanie aktualizacji ww. skali AMSTAR, mianowicie AMSTAR-2 (24), która uwzględnia też ocenę metodologicznej jakości przeglądów systematycznych badań obserwacyjnych (a nie tylko randomizowanych jak jej oryginalna wersja).

Narzędzie AMSTAR (Aneks 4), jako odnoszące się w szczególności do interwencji z zakresu zdrowia publicznego (25), jest kwestionariuszem składającym się z 11 pozycji, oceniającym obecność następujących elementów: wcześniejsze zaprojektowanie, selekcja duplikatów i ekstrakcja danych, kompleksowe przeszukiwanie literatury, odniesienie się do statusu publikacji jako jednego z kryteriów włączenia, przedstawienie listy włączonych/wykluczonych publikacji, przedstawienie charakterystyki włączonych badań, udokumentowanie oceny jakościowej badań włączonych do przeglądu systematycznego, odpowiednie sformułowanie wniosków, odpowiednie zastosowanie metod służących syntezy wyników, ocena prawdopodobieństwa błędów metodologicznych włączonych badań, udokumentowanie konfliktu interesów (26, 27). Narzędzie to powinno być aktualizowane w miarę postępu wiedzy nt. czynników wpływających na jakość przeglądów systematycznych.

Wady i ograniczenia dotyczące projektowania lub przeprowadzenia przeglądu systematycznego mogą skutkować potencjalnym błędem w otrzymanych wynikach. Błąd ten może narastać we wszystkich stadiach tworzenia przeglądu systematycznego. Osoba korzystająca z przeglądu systematycznego powinna być świadoma tych błędów w trakcie

interpretowania wyników i wniosków. Mimo istnienia kilku narzędzi służących do oceny jakości metodologicznej przeglądów systematycznych, żadne z nich nie odnosi się do oceny ryzyka błędu systematycznego przeglądu systematycznego. W tym właśnie celu zostało stworzone narzędzie ROBIS (28, 29) (Aneks 5).

Należy mieć na uwadze, że ocena jakości danych pozwala określić ich wiarygodność wewnętrzną (*internal validity*) oraz zewnętrzną (*external validity*). Ocena jakości informacji dostarczanych przez badania włączone do analizy wymaga rozważenia szeregu czynników:

- metodyki przeprowadzania poszczególnych badań,
- ryzyka błędu systematycznego (*bias*),
- spójności pomiędzy wynikami poszczególnych badań włączonych do przeglądu,
- stopnia, w jakim wyniki stwierdzone w badaniach naukowych można przenieść (uogólnić) na populację, której ma dotyczyć analiza.

Wśród narzędzi najczęściej stosowanych w klasyfikacji rekomendacji tworzonych *de novo* (jakość danych i siła zaleceń) wskazuje się skalę *Grading of Recommendations Assessment, Development and Evaluation* (GRADE (30); Aneks 6). Polega ona na zdefiniowaniu pytań, wybraniu punktów końcowych i przypisaniu im wagi, ocenie jakości danych i zinterpretowaniu tych danych, głównie w kontekście oszacowanych efektów względnych i bezwzględnych.

Do oceny wytycznych, a w szczególności kontroli ich jakości powszechnie używanym narzędziem jest instrument *Appraisal of Guidelines for Research and Evaluation* (AGREE, 2017; Aneks 7). Narzędzie to składa się z 23 pozycji zgrupowanych w sześciu domenach, tj. zakres i cel, udział osób zainteresowanych, rzetelność opracowywania wytycznych, przejrzystość i sposób przedstawienia, użyteczność oraz niezależność redakcyjna. Każda pozycja jest oceniana według 7-stopniowej skali Likerta, gdzie 1 oznacza najmniejszą jakość, a 7 dotyczy najwyższej jakości zaleceń. Wyniki dla każdej z domen są sumą punktów, jakie uzyskały poszczególne ich pozycje. Są również niezależne i nie mogą być sumowane w jeden wynik całościowy.

3.13. Dodatkowe elementy brane pod uwagę przy tworzeniu przeglądu systematycznego interwencji z zakresu zdrowia publicznego

Zagadnienia etyczne i nierówności

Przeglądy systematyczne mogą określać skuteczność podejmowanych działań z zakresu zdrowia publicznego. Jednak działania te powinny uwzględniać również względy etyczne i społeczne. Standardowe postępowanie oceniające skuteczność interwencji odnosi się do ogólnej charakterystyki populacji docelowej, nie biorąc pod uwagę możliwych nierówności w zdrowiu wynikających z czynników innych niż zdrowotne. Takie podejście może zniekształcać wyniki analizowanych punktów końcowych w poszczególnych populacjach. Jest to ważne, ponieważ określenie wpływu interwencji na grupy znajdujące się w niekorzystnej sytuacji, może zwrócić uwagę na konieczność przyjęcia strategii mających na celu zmniejszenie nierówności zdrowotnych. Informacje uzyskane w ramach przeglądu systematycznego mogą dostarczyć informacji na temat oddziaływania skutków interwencji na nierówności zdrowotne. Uzyskane informacje można następnie wykorzystać do rozwiązania problemu nierówności. Niekorzystna sytuacja prowadząca do nierówności może być rozpatrywana w aspekcie miejsca zamieszkania (*place of residence*), pochodzenia etnicznego (*race, ethnicity*), wykonywanego zawodu (*occupation*), płci (*gender*), wyznania (*religion*), poziomu wykształcenia (*education*), statusu socjoekonomicznego (*socio-economic position*) i kapitału społecznego (*social capital*) – znane jako akronim PROGRESS (7, 32). Autorzy przeglądu powinni dokładnie rozważyć, które z tych czynników są istotne w uwzględnianej przez nich populacji. Wymaga to zwrócenia uwagi nie tylko na korzyści lub szkody wynikające z badanej interwencji, lecz także na ich dystrybucję tj. kto czerpie największe korzyści, kto jest poszkodowany, kto jest wykluczony z możliwości korzystania z interwencji?

W przeglądzie dot. efektywności badanej interwencji w odniesieniu do nierówności zdrowotnych wymagane jest uwzględnienie trzech elementów:

- pomiaru stanu zdrowia (lub zmiany stanu zdrowia),
- pomiaru statusu socjoekonomicznego,
- metody statystycznej podsumowującej wielkość różnic zdrowotnych między osobami w różnych grupach.

Przeprowadzanie analiz dotyczących nierówności jest skomplikowane nie tylko ze względu na ograniczone informacje dot. różnic między grupami, ale także przez ograniczony udział grup osób mogących odnieść mniejsze korzyści (grup defaworyzowanych) w badaniach. Pomimo tych barier, przeglądy systematyczne mogą odgrywać istotną rolę w podnoszeniu świadomości w zakresie nierówności zdrowotnych. *Cochrane Health Equity Field and Campbell Equity Methods Group* (33) zidentyfikowała przeglądy istotne w aspekcie równości, mogące stanowić dodatkowe wskazówki dla autorów przyszłych

opracowań. Zidentyfikowanie i pozyskanie badań dot. nierówności, może wymagać od autorów przeglądu szerokiego wyszukiwania i kontaktowania się z autorami badań celem uzyskania dalszych informacji dot. statusu socjoekonomicznego (opublikowane badania często nie przedstawiają informacji o rozkładzie statusu socjoekonomicznego wśród uczestników danego badania).

W tym przypadku, aby określić, czy przegląd systematyczny uwzględni niekorzystne czynniki mogące wpływać na zdrowie warto zastosować narzędzie PROGRESS.

PROGRESS jest narzędziem, którego zadaniem jest określenie czy planowana interwencja uwzględni nierówności wynikające z różnorodności populacji. PROGRESS ma na celu przedstawienie wielowymiarowości rozkładu zdrowia w populacji. Odnalezienie czynników wpływających na zdrowie daje możliwość skierowania środków zwalczających nierówności w zdrowiu w sposób precyzyjny i wymierny. Narzędzie to w przeglądach systematycznych może być używane jako ramy określania niekorzyści/niedogodności, które wpływają na zdrowie. W dalszej perspektywie PROGRESS może być pomocny w zwalczaniu nierówności zdrowotnych.

Narzędzie to odnosi się do ośmiu głównych zagadnień:

- miejsca występowania,
- rasy/pochodzenia kulturowego, etnicznego/używanego języka,
- wykonywanego zawodu,
- płci,
- wyznania,
- wykształcenia,
- statusu socjoekonomicznego.

PROGRESS-PLUS jest rozwinięciem podstawowej skali PROGRESS m.in. o ogólną pozycję socjoekonomiczną uwzględniającą wykształcenie, zatrudnienie i miejsce zamieszkania, wiek, niepełnosprawność, orientację seksualną, inne ważne czynniki (np. nastoletni rodzice, osoby nie uczęszczające do szkół, byli więźniowie, ofiary przemocy).

PROGRESS (32, 34) składa się z 7 bloków pytań, z czego poszczególne odnoszą się do:

1. Rozwinięcia/stworzenia logicznego modelu postępowania.
2. Zdefiniowania jakie występują ujemne strony /braki/niekorzyści oraz do kogo skierowana jest interwencja.
3. Włączenia odpowiednich badań.
4. Zidentyfikowania punktów końcowych.
5. Ewaluacji procesu oraz wskazaniu spójności.
6. Przeanalizowania i prezentacji danych.

Kontekst

Typ implementowanej interwencji, a co za tym idzie jej następny sukces lub niepowodzenie, w dużym stopniu zależą od społecznego, ekonomicznego i politycznego kontekstu, w jakim przedmiotowe interwencje są realizowane. Przykład interwencji, której sukces zależy od kontekstu realizacji zamieszczono w Aneks 9.

Przy prowadzeniu przeglądu systematycznego należy zachować ostrożność przy przenoszeniu wniosków z jednego kontekstu na inny (w jednych warunkach dany program przynosi korzyści dla odbiorców, w innych można nie zaobserwować pozytywnego efektu). Standardowo, punkty końcowe w badaniach przypisywane są interwencji. Jednak punkty te odnotowane w badaniach, mogą w rzeczywistości wynikać z wcześniej istniejącego kontekstu (określonych czynników), w którym interwencja była prowadzona. Dlatego kontekst powinien być rozważany i mierzony jako modyfikator efektu w badaniach. Szersze spektrum kontekstu mogą obejmować aspekty systemu, w którym dana interwencja ma być wdrożona. Kontekst ma istotne znaczenie przy implementowaniu danej interwencji/rozwiązania np. z innego kraju, w którym ta interwencja miała zastosowanie.

Należy zwrócić uwagę, na to, że od klasterowych badań z randomizacją można oczekiwać (teoretycznie) równomiernego rozłożenia, wyrównania lub zredukowania istotnych aspektów danego kontekstu, pod warunkiem wystarczającej wielkości próby.

Należy również zaprezentować informacje związane z kontekstem jeśli są dostępne:

- aspekty dotyczące organizatora i personelu: np. liczebność, doświadczenie, morale, wiedza pracowników, historia organizacji dotycząca innowacji, rozmiar organizacji, status organizacji programu;
- aspekty związane z systemem, np. płaca i wynagrodzenie, system wynagradzania, stopień wyspecjalizowania w dostarczaniu świadczeń;
- charakterystyka populacji docelowej (czynniki socjoekonomiczne, kulturowe, związane z miejscem zamieszkania i poziomem alfabetyzacji).

Informacje dot. kontekstu interwencji są częściej przedmiotem analiz jakościowych aniżeli ilościowych.

Ciągłość, trwałość

Ciągłość (*sustainability*) odnosi się do kontynuowania interwencji lub programu i/lub jego wyników. Zagadnienie związane z ciągłością powinno być ważną częścią rozważań w ramach przeglądu systematycznego. Użytkownicy przeglądu systematycznego są często zainteresowani wiedzą dotyczącą tego, kiedy korzyści płynące z interwencji będą się utrzymywały nawet po zakończeniu interwencji.

Interwencje z zakresu zdrowia publicznego i promocji zdrowia są zazwyczaj złożone, a zmiany w zachowaniach zdrowotnych często powolne i trudne. Z tego względu wyniki dotyczące wpływu interwencji na zachorowalność czy śmiertelność mogą być dostępne dopiero po bardzo długim czasie od zastosowania interwencji.

W przypadku rozważania zagadnień związanych z ciągłością należy:

- Zastanowić się w jakich przypadkach ciągłość wyników jest odpowiednia do zastosowanej interwencji. Jeśli rozważanie ciągłości jest w danym przypadku zasadne, należy zastanowić się, które punkty końcowe powinno się uwzględnić, w jakim okresie czasu i jaki jest wzorzec dla wyników interwencji w dłuższej perspektywie czasu.
- Informacje powinny być rozważane zarówno w odniesieniu do kontekstu jak i charakterystyki projektu, co może wyjaśnić zakres ciągłości. W przypadku, gdy w badaniu nie odniesiono się do ciągłości, można zbadać potencjał punktów końcowych/wyników do utrzymywania się w czasie. Do oceny ciągłości mogą być przydatne cztery następujące ramy:
 1. Czynniki opisane przez Bossert (35):
 - ekonomiczne i polityczne zmienne w zakresie implementacji oraz ewaluacji interwencji;
 - siła instytucji wdrażającej daną interwencję;
 - pełna integracja interwencji w istniejącym programie/świadczeniu itd.;
 - uwzględnienie przez program elementu szkoleniowego;
 - zaangażowanie społeczeństwa/uczestnictwo w programie.
 2. Ramy opracowane przez Swerissen i Crisp (36) dotyczą prawdopodobieństwa stabilności interwencji/programu i efektów na różnych poziomach organizacji społecznej. Ramy te określają związek pomiędzy poziomem interwencji, strategiami i prawdopodobną trwałością interwencji/programu i efektów.
 3. Ramy opracowane przez Shediac-Rizkallah i Bone (37) prezentują kryteria użyteczne dla konceptualizacji ciągłości. W tym kontekście kluczowe aspekty ciągłości programu są zdefiniowane jako 1) utrzymanie korzyści zdrowotnych płynących z programu; 2) instytucjonalizacja programu wewnątrz organizacji; oraz 3) tworzenie potencjału w społeczności odbiorców. Kluczowe czynniki wpływające na ciągłość definiowane są jako 1) czynniki w szerszym środowisku; 2) czynniki w otoczeniu warunków organizacyjnych; oraz 3) projektowanie i czynniki związane z wdrażaniem.
 4. *The Centre for Health Promotion, University of Toronto* (38) również stworzyło dokument podkreślający 4 integralne elementy ciągłości.
- W ocenie jakości badań włączonych do przeglądu systematycznego powinno się rozważyć kilka kwestii odnoszących się do ciągłości:
 1. Szeroka konceptualizacja potencjalnego wpływu na ciągłość odnosząca się do standardowych ram;
 2. Zakres w jakim dane dotyczące aktywności programu, wyników programu oraz czynników, które mogą wpływać na ciągłość są prospektywnie zbierane poza czasem trwania interwencji;
 3. Zakres w jakim program jest modyfikowany po pewnym czasie oraz zakres w jakim program może być uważany za nowy lub inny w wyniku modyfikacji;

4. Zakres w jakim jakość i ilość aktywności podejmowanych w ramach programu jest kontynuowana w trakcie procesu instytucjonalizacji;
5. Potencjalny wpływ naturalnego progresu choroby i starzenia się kohort na potencjalne utrzymanie korzyści zdrowotnych.

Stosowalność

Zakres, w którym oczekuje się, że interwencja z zakresu zdrowia publicznego będzie działać, może być mniej przewidywalny niż w przypadku interwencji klinicznych. Z tego względu niezwykle ważne jest zbadanie szczegółów dotyczących procesu wdrażania interwencji, jak i kontekstu przed ekstrapolowaniem wyników z poszczególnych badań i przeprowadzeniem dalszej syntezy wyników. Pierwszą trudnością jest to, że duża liczba interwencji z zakresu zdrowia publicznego nie uwzględnia w badaniach randomizacji. Przez to, że w badaniach są często słabiej zdefiniowane kryteria kwalifikacji, warunki realizacji, interwencje, utrudnia to ustalenie stosowalności (*applicability and transferability*). Drugą kwestią jest, że interwencje w zdrowiu publicznym i promocji zdrowia zazwyczaj składają się z wielu komponentów. Dlatego utrudnia to określenie, który dokładnie komponent interwencji charakteryzował się znaczącym i zauważalnym efektem, oraz ocenę synergii między komponentami. Trzeci aspekt dotyczy tego, że w interwencjach realizowanych w określonych wspólnotach/społecznościach, wdrażanie i utrzymanie się realizacji tej interwencji, może być znacznie trudniejsze do osiągnięcia i zmierzenia. Utrudnia to także interpretację i zastosowanie otrzymanych wyników. W interwencjach z zakresu zdrowia publicznego, zasadnicze cechy socjologiczne i kulturowe danych społeczności są złożone i trudne do zmierzenia. Dlatego też trudno jest określić u kogo i w jakim stopniu zastosowano daną interwencję, przez co trudniej jest ustalić stosowalność. *Cochrane Handbook* zwraca jednak uwagę, że ta heterogeniczność może zwiększać zastosowanie, ponieważ wcześniej zdefiniowane populacje, warunki realizacji oraz interwencje mogą być w dużym stopniu zróżnicowane, zwiększając tym samym prawdopodobieństwo tego, że będzie można to wykorzystać w szerszym zakresie. Autorzy wykonywanego przeglądu mogą podsumować różne aspekty odnoszące się do potencjalnych odbiorców interwencji, pochodzące z odnalezionych dowodów i odnieść je do warunków, w których ma być wprowadzana badana interwencja.

Przeprowadzenie podsumowania wyników różnych badań przeprowadzonych w różnych warunkach i w różnych populacjach jest samo w sobie testem stosowalności wyników. Jeśli wyniki z różnych badań, przeprowadzanych w różnych okolicznościach są do siebie zbliżone wtedy prawdopodobieństwo, że wyniki można przetransferować (*transferability*) wzrasta. Jeśli jednak wyniki różnią się w znaczący sposób, to informacja może być przydatna do zrozumienia w jakich okolicznościach wyniki z badania znajdą zastosowanie. Według Wang 2005 (39) stosowalność (*applicability*) definiowana jest jako stopień, w jakim interwencja mogłaby być wdrożona w innych warunkach, natomiast możliwość przenoszenia (*transferability*) jako stopień, w jakim mierzona skuteczność wdrożonej już interwencji, może zostać osiągnięta w innych warunkach. Różnice i kontrast między tymi dwoma pojęciami zostały przedstawione w Aneks 10. Wdrożenie jakiegokolwiek interwencji z zakresu zdrowia publicznego stawia szereg wymagań w zakresie kontekstu np. sprzyjająca sytuacja polityczna i/lub zgodność z normą społeczną. Dlatego aby móc dokonać oceny stosowalności, można opracować listę możliwych jej atrybutów, które mogłyby być opracowane wyłącznie na podstawie rzetelnej wiedzy na temat danej interwencji w zakresie zdrowia publicznego. Atrybuty mogą być następnie oceniane według ich obecności w danych warunkach lokalnych przez osoby o dobrej znajomości tych warunków. Istnieje wiele zewnętrznych i wewnętrznych czynników kontekstowych, które mogą mieć wpływ na skuteczność interwencji. Wspomniane czynniki obejmują: sytuację epidemiologiczną (występowanie danej choroby i jej rozszczylenie w różnych obszarach geograficznych i/lub grupach ludności), zdolność do wdrożenia interwencji (w tym dostępność zasobów, umiejętności danej grupy osób, czynniki organizacyjne oraz otoczenie społeczne i polityczne) oraz charakterystyka populacji docelowej (w tym uwarunkowania kulturowe).

CRD oraz *Cochrane Handbook* wskazuje na narzędzie pozwalające na ocenę stosowalności wyników z badań naukowych w ramach PPZ, jak i ich transferowania opracowane przez Wang 2005 (39). Wskazywane są co najmniej dwa podejścia w tym zakresie. Pierwsze obejmuje metodę delficką – zaangażowanie ekspertów (eksperci odpowiedni dla danego problemu zdrowotnego, epidemiolodzy, socjologowie, ekonomiści zdrowia, specjaliści w zakresie zdrowia publicznego i przedstawiciele populacji docelowej) w proces wdrażania interwencji od projektu do oceny wyników. Następnie inni eksperci, którzy znają kontekst lokalny mogą określić czy i w jakim stopniu interwencję można wdrożyć i dostosować do warunków lokalnych, i czy otrzymane w badaniach wyniki skuteczności interwencji, mogą zostać przeniesione na warunki lokalne. Drugie podejście opiera się na literaturze. Osoby wykonujące przegląd systematyczny podejmują decyzje w oparciu o dostępne dowody naukowe, a także swoje doświadczenie i wiedzę w zakresie związku między wdrażaną interwencją a sytuacją lokalną. Następnie osoby wykonujące przegląd literatury mogą opracować listę parametrów w zakresie stosowalności, opierając się na dostępnych informacjach pochodzących z literatury.

Cochrane Handbook wskazuje na poniższe pytania jako pomocne dla autorów przeglądów przy rozważaniu kwestii dot. stosowalności interwencji w zakresie zdrowia publicznego i promocji zdrowia.

1. Stosowalność (*applicability*):
 - Czy środowisko polityczne w danej społeczności lokalnej przychyli się do wdrożenia interwencji?
 - Czy istnieje jakakolwiek bariera polityczna utrudniająca wdrożenie interwencji?

- Czy populacja ogólna oraz populacja docelowa akceptuje interwencję? Czy jakkolwiek aspekt interwencji jest wbrew lokalnym normom społecznym? Czy jest etycznie akceptowalna?
- Czy wszystkie elementy interwencji mogą być dostosowane do lokalnych uwarunkowań kulturowych?
- Czy niezbędne zasoby do wdrożenia interwencji są dostępne w warunkach lokalnych?
- Czy populacja docelowa ma wystarczający poziom wykształcenia do zrozumienia interwencji?
- Która organizacja będzie odpowiedzialna za prowadzenie interwencji w na danym obszarze?
- Czy istnieje jakkolwiek bariera we wdrożeniu interwencji ze względu na organizację?
- Czy podmiot odpowiedzialny za realizację interwencji w warunkach lokalnych posiada umiejętność jej realizacji? Jeśli nie, to czy będą dostępne szkolenia?

2. Możliwość przetransferowania (*transferability*):

- Jaka jest podstawowa częstość występowania problemu zdrowotnego na danym obszarze? Jaka jest różnica w częstości występowania pomiędzy badaniem a lokalną sytuacją?
- Czy charakterystyka populacji docelowej jest porównywalna między badaniem a warunkami lokalnymi? W odniesieniu do poszczególnych aspektów, które będą przedmiotem interwencji, czy możliwe jest, że cechy populacji docelowej, takie jak pochodzenie etniczne, status społeczno-ekonomiczny, poziom wykształcenia itp., będą miały wpływ na skuteczność interwencji?
- Czy możliwość wdrożenia interwencji jest porównywalna z warunkami badania w takich kwestiach jak uwarunkowania polityczne, akceptacja społeczna, zasoby, struktura organizacyjna i kompetencje lokalnych podmiotów odpowiedzialnych za realizację?

3.14. Przedstawienie badań włączonych do przeglądu systematycznego i ekstrakcja danych

Interwencje w zakresie zdrowia publicznego są bardzo często złożonymi działaniami. Skierowane są do zróżnicowanej populacji osób badanych oraz charakteryzują się niejednorodnymi punktami końcowymi czy też zróżnicowanymi zagadnieniami metodologicznymi. Skuteczność danej interwencji może być modyfikowana przede wszystkim poprzez kontekst. W wyniku wyżej wspomnianej złożoności należy spodziewać się znacznej heterogeniczności w badaniach i tym samym rozważyć wyżej wymienioną kwestię podczas syntezy wyników (40).

Synteza wyników badań ma za zadanie uzyskanie możliwie skoncentrowanej informacji oraz określenie poziomu związanej z nią niepewności. Obejmuje ona systematyczny przegląd literatury (wraz z metaanalizą lub bez), jak również jakościowe podsumowanie.

Cochrane Handbook (41) zaznacza, że istnieją dwa podejścia celem zintegrowania wyników jakościowych i ilościowych, tj.: synteza wielopoziomowa oraz synteza równoległa. W wielopoziomowej analizie dowodów jakościowych (synteza 1) i dowodów ilościowych (synteza 2) może być prowadzona osobnym strumieniem lub rozdzielnie, ale ostatecznie łącząc się w syntezę 3 (tj. przeglądy i inne badania każdej z syntez są łączone). Synteza równoległa zakłada dodatkowo, że syntezę jakościową (synteza 1) można wykorzystać równolegle i zestawzić z wynikami syntezy ilościowej (synteza 2) celem ułatwienia interpretacji wyników włączonych badań.

Jeśli odnaleziono wystarczająco wiarygodne badania, w których zastosowano porównywalne miary wyników w zbliżonych populacjach (zredukowana heterogeniczność badań), zalecane jest opracowanie wyników za pomocą metaanalizy (synteza ilościowa). Jeśli badania te są niedostępne/ nie zostały odnalezione, analiza może być ograniczona do przeglądu jakościowego, w którym krytycznie ocenia się poszczególne badania i przedstawia ich wyniki w tabelach. Na podstawie syntezy wyników wyciągane są stosowne wnioski.

3.15. Synteza jakościowa

Kluczową kwestią w przeglądzie systematycznym przeglądów/badań naukowych z uwzględnieniem interwencji z zakresu zdrowia publicznego jest identyfikacja publikacji o odpowiedniej jakości. Wykonuje się adekwatne zestawienie danych dotyczących skuteczności oraz bezpieczeństwa rozpatrywanej technologii medycznej oraz działań z zakresu zdrowia publicznego (w jednolitym formacie tabelarycznym, umożliwiającym porównanie wyników badań dla określonych punktów końcowych). Zestawienie to powinno uwzględniać dokonaną uprzednio ocenę wiarygodności źródeł i jakości danych. Należy zestawzić wyniki wszystkich doniesień naukowych spełniających kryteria włączenia do przeglądu systematycznego. W przypadku stwierdzenia wysokiej heterogeniczności uzyskanych wyników konieczne jest prześledzenie i opisanie różnic.

Celem wykonania prawidłowego wyszukiwania przeglądów, które miałyby zostać uwzględnione w planowanym przeglądzie systematycznym, zaleca się stosowanie odpowiednich filtrów wyszukiwania. Przykładowe filtry zostały opisane w Aneks 1 i dostępne są też m.in. na stronie internetowej grupy *InterTASC Information Specialists' Sub-Group* (ISSG), pn. *ISSG Search Filters Resource* (42).

Należy mieć na uwadze, że sam fakt nazwania w tytule publikacji przeglądem systematycznym, nie oznacza wcale, że nim jest. Dostępne w bazach naukowych przeglądy (tj. *reviews*) powinny być traktowane jako przeglądy systematyczne wyłącznie, gdy podczas ich opracowywania zastosowano minimalny zestaw kryteriów kwalifikowalności włączanych publikacji. Kryteria te mogą się różnić w zależności od źródła ich pochodzenia, niemniej jednak prawidłowo opisany przegląd zawiera pełną i odpowiednio do zagadnienia sprecyzowaną strategię wyszukiwania publikacji, predefiniowane kryteria włączenia badań, ocenę ryzyka błędu możliwego do wystąpienia w uwzględnionych badaniach, a także prawidłową syntezę wyników włączonych do opracowania.

Prawidłowe postępowanie w tworzeniu przeglądu systematycznego z przeglądów systematycznych to uwzględnienie wyłącznie tych przeglądów, które spełniają powyższe kryteria jakości oraz odnoszą się bezpośrednio do opracowywanego zagadnienia, a nie uwzględnienie wszystkich opublikowanych dotychczas przeglądów na dany temat. Ocena adekwatności badań (już pod kątem samego zagadnienia, a nie jakości badań) powinna zostać oceniona na podstawie schematu PICO(T), tj. określenia populacji, interwencji, komparatorów, efektu końcowego, ale również czasu (*Time*, T).

W celu prawidłowej oceny odnalezionej w przeglądach metodologii zalecane jest zastosowanie odpowiednich narzędzi. Należy mieć na uwadze, że sama ocena jakości odnalezionych przeglądów systematycznych nie mówi nic o ryzyku błędów bądź sile dowodów uwzględnionych w danym przeglądzie (SEO). Dlatego też, w miarę możliwości, zaleca się ocenę jakości badań pierwotnych włączanych do odnalezionych przeglądów systematycznych podczas opracowywania nowego przeglądu systematycznego (43). Wystąpienie niewielkich różnic w zakresie analizowanego ryzyka nie powinno martwić, jeśli ogólna ocena dowodów naukowych została określana jako o niskim ryzyku wystąpienia błędu.

Kolejną kwestią jest szczegółowe przedstawianie wyników przeglądu systematycznego z przeglądów. Zaleca się, aby przedstawić nie tylko wnioski wtórne, ale i pochodzące bezpośrednio z badań pierwotnych. Takie postępowanie znajduje uzasadnienie w momencie, gdy autorzy przeglądu z przeglądów będą wyodrębniać inne dane, niż te wyodrębnione podczas tworzenia indywidualnych przeglądów systematycznych. Zestawione wyniki powinny bezwzględnie odnosić się do siły odnalezionych dowodów naukowych. W zestawieniach powinny znaleźć się również informacje dot. liczby badań włączonych do przeglądu systematycznego, liczby uczestników, osiągane efekty wraz z przedziałem ufności. Jednakże, gdy wnioski przeglądu z przeglądów dotyczą tego samego aspektu, co w odnalezionych przeglądach systematycznych, nie jest konieczne przywoływanie danych z badań pierwotnych.

Zgodnie z konsensem eksperckim grupy *Evidence-based Practice Centers* (EPC) (44), decyzja o wykonaniu syntezy ilościowej/jakościowej powinna zależeć nie od samej jakości odnalezionych przeglądów systematycznych, ale od jakości uwzględnionych w nich badań pierwotnych. Dlatego też dominującym postępowaniem w tej kwestii jest podejście ujmujące siłę dostępnych dowodów naukowych, tj. ocena: ograniczeń, spójności, dokładności oraz raportowania błędów. Celem tworzenia przeglądu z przeglądów jest aktualizacja wyników kolejnych przeglądów systematycznych poprzez uwzględnianie w nich coraz to nowszych wyników badań pierwotnych. Jeśli w ostatnim czasie nie pojawiły się nowe badania pierwotne w wybranym zakresie, wówczas wykonywanie nowej analizy jakościowej/ilościowej nie ma większego sensu. Zasadność przeprowadzenia nowej analizy zaistnieje wówczas, gdy pojawienie się nowych badań pierwotnych może zmienić dotychczasowe wnioski wynikające z odnalezionych przeglądów systematycznych. Kwestia ta powinna być jednak jednoznacznie udokumentowana i opisana w tworzonym przeglądzie przeglądów.

3.16. Synteza ilościowa

Ilościowa synteza wyników pochodzących z niezależnych badań z grupą kontrolną/przeglądów systematycznych, tj. metaanaliza, stawiana jest najwyżej w hierarchii wiarygodności dowodów naukowych EBM. Pozwala ona na precyzyjną ocenę wpływu działania analizowanej interwencji oraz rozszerzyć wnioski z pojedynczych badań/przeglądów na szerszą populację. Przed wdrożeniem wspomnianej syntezy ilościowej istotne jest, aby mieć pewność, że wyniki poszczególnych badań zostały zebrane poprawnie (np. kierunek standaryzowanej średniej różnic może być przypadkowo błędnie określony).

Przed zastosowaniem statystycznych metod syntezy oszacowuje się poziom i źródła (przyczyny) heterogeniczności statystycznej wyników badań ujętych w metaanalizie (45) (Aneks 11). Do najczęściej pojawiających się źródeł heterogeniczności badań w metaanalizie zalicza się m.in. odmienną metodologię badań, zróżnicowane próby badawcze, różnorodne okresy trwania badania, zróżnicowane wskaźniki i sposoby estymacji wyników. Celem usunięcia wyżej wspomnianych czynników stosuje się analizę wrażliwości (odnalezienie badań odmiennych), metaregresję (odnalezienie czynnika powodującego heterogeniczność), bądź analizuje się wyniki w różnorodnych populacjach badań (analiza w podgrupach). Statystyczna różnorodność wyników badań może być zbadana za pomocą modelu efektów losowych (*random effects model*). W przypadku jednorodnych badań wykorzystuje się model stały (*fixed effects model*).

Do oprogramowania stosowanego w danych analizach należy m.in. *RevMan*, programy statystyczne jak *pakiet R*, *Statsdirect* czy np. *Excel*.

Wśród głównych etapów tworzenia metaanalizy należy wyróżnić:

1. formułowanie problemu i typu badania,
2. sformułowanie kryteriów włączania analizy,
3. identyfikację odpowiednich badań,
4. wyłączenie badań źle przeprowadzonych i obciążonych błędami metodologicznymi,
5. ocenę, agregację i interpretację wyników (według tej samej skali).

Graficzną formą przedstawienia wyników metaanalizy jest wykres typu *forest plot*.

3.17. Efektywność kosztowa interwencji z zakresu zdrowia publicznego

Efektywność kosztowa interwencji z zakresu zdrowia publicznego może być oceniana za pomocą przeglądu systematycznego pierwotnych analiz ekonomicznych. W ocenie efektywności kosztowej interwencji z zakresu zdrowia publicznego znajdują zastosowanie wszystkie rodzaje analiz ekonomicznych. Mimo to metody oceny ekonomicznej jakie są stosowane w HTA w klasycznym ujęciu nie pokrywają wszystkich kosztów i korzyści odpowiednich do oceny interwencji z zakresu zdrowia publicznego (46).

Rozszerzenie zakresu zastosowania analiz ekonomicznych (z analiz dot. leków) wiąże się z wieloma problemami metodologicznymi. Trudności związane z ekonomiczną oceną programów z zakresu zdrowia publicznego wynikają z ich specyficznego charakteru (np. kompleksowości, zróżnicowanej populacji docelowej w obrębie danego programu), który w wielu sytuacjach uniemożliwia zastosowanie metod wykorzystywanych przy ocenie efektywności kosztowej leków lub procedur klinicznych (46, 47).

Poniżej zostały przedstawione zalecenia odnoszące się do poszczególnych etapów i/lub elementów procesu ekonomicznej ewaluacji programów zdrowia publicznego. Są to po kolei: wybór perspektywy analizy, wybór komparatora, horyzont czasowy, rodzaj badania, problematyka szacowania kosztów i efektów oraz uwzględnienie kontekstu programu.

1. Perspektywa analizy określa, które koszty i konsekwencje muszą zostać uwzględnione. Ocena ekonomiczna interwencji z zakresu zdrowia publicznego może odbywać się z perspektywy uczestnika, dostawcy usług, płatnika lub z perspektywy całego społeczeństwa (46). W odniesieniu do analiz ekonomicznych programów zdrowia publicznego zasadnym jest przyjęcie perspektywy społecznej (a nie płatnika jak ma miejsce najczęściej w przypadku decyzji refundacyjnych dot. leków), która zapewnia uwzględnienie wszystkich możliwych kosztów i rezultatów, bez względu na to, kto generuje koszty, a kto jest odbiorcą rezultatów/efektów. Przy perspektywie społecznej możliwe jest uwzględnienie wielowymiarowych celów programu – nie tylko odnoszących się do samego zdrowia (np. zwiększenie integracji w danej społeczności, poprawa warunków socjalno-bytowych). Należy jednak zaznaczyć, że perspektywa społeczna generuje wiele trudności metodologicznych, które są związane z szacowaniem kosztów i efektów.
2. Komparator – zakłada interwencje alternatywne, do których oceniany program będzie porównywany. Należy zaznaczyć, że jest to bardzo ważny aspekt analizy, ponieważ określa wartość dodaną/korzyść, jaką niesie za sobą wdrożenie ocenianego programu (tj. nadwyżka kosztów oraz efektów ocenianego programu nad kosztami i efektami alternatywy). W przeprowadzanej analizie ekonomicznej, daną ocenianą interwencję porównuje się przynajmniej z dotychczasową praktyką (standardem postępowania w danym zakresie). Porównanie do wszystkich istniejących alternatyw, mimo, że byłoby idealne, generowałoby duże zasoby, koszty oraz czas wykonania takiej analizy. Jak opisano w rozdziale dot. opcjonalnych technologii medycznych, w przypadku programów z zakresu zdrowia publicznego znalezienie właściwego komparatora może stanowić trudne zadanie. Jeśli cele programu wykraczają poza aspekty czysto zdrowotne (np. poprawę zdrowia w danym zakresie w konkretnej populacji docelowej) i obejmują wpływ na warunki środowiskowe i/lub wzmacnianie społeczności lokalnej – odpowiedni komparator może nie występować. W takim przypadku zaleca się porównanie programu (ocenianych interwencji bądź działań) z tak zwaną *do nothing option*, a więc brakiem jakiegokolwiek interwencji. Dlatego istotnym jest aby wybór komparatora za każdym razem szczegółowo uzasadnić.
3. Horyzont czasowy analizy określa, w jakim czasie (jak długo) koszty i efekty ocenianej interwencji/działania w ramach programu będą poddawane obserwacji w czasie (jak długo trwa tzw. okres *follow-up*). Należy zaznaczyć odmienną przyjmowania horyzontu czasowego w analizie ocenianych programów ze zdrowia publicznego w porównaniu do interwencji klinicznych np. w badaniu klinicznym horyzont czasowy będzie obejmował czas trwania interwencji i jej efektów (np. roczny *follow-up* w badaniu klinicznym). Natomiast efekty programu z zakresu zdrowia publicznego (np. programu edukacji zdrowotnej lub interwencji z zakresu redukcji określonych czynników ryzyka) mogą się pojawić nawet po kilkunastu latach. Dlatego też wieloletni horyzont czasowy generuje także

konieczność dyskontowania przy ocenie kosztów i efektów. Dodatkowo należy zwrócić uwagę na to, że aktualna (teraźniejsza) wartość danego efektu, który pojawi się za kilkanaście lat, zawsze będzie niższa od wartości teraźniejszej takiego samego efektu, który pojawi się w krótszej perspektywie czasu. W praktyce, gdy porównywana jest efektywność kosztowa programów z zakresu zdrowia publicznego z interwencjami klinicznymi, istnieje ryzyko, że programy zdrowia publicznego mogą być określane jako mniej opłacalne. Jednym z możliwych rozwiązań problemu dyskontowania, jest wykonanie analizy wrażliwości, która będzie uwzględniać różne wartości stopy dyskontowej (w tym także wartości zerowej).

4. Rodzaj badania odnosi się do sposobu dostarczenia interwencji z zakresu zdrowia publicznego, wyboru grupy docelowej do programu oraz metody zbierania danych. Zgodnie z hierarchią dowodów naukowych najwyżej wśród badań pierwotnych znajduje się poprawnie zaprojektowana kontrolowana próba kliniczna z randomizacją (RCT). Ten rodzaj badania jest jednak bardzo trudny do przeprowadzenia w aspekcie interwencji ze zdrowia publicznego. Zasadniczą trudność polega na zaobserwowaniu wielowymiarowych efektów programu z zakresu zdrowia publicznego w badaniu pierwotnie zaprojektowanym do oceny wąsko sprecyzowanych interwencji klinicznych (np. leków), przeznaczonych dla pojedynczych pacjentów – badanie takie przebiega także w konkretnych kontrolowanych warunkach klinicznych, z relatywnie krótkim okresem obserwacji (*follow-up*). Jak wspomniano wcześniej, często stosowanym rozwiązaniem w zdrowiu publicznym jest zastosowanie klasterowej kontrolowanej próby klinicznej z randomizacją, w której randomizacji podlegają nie pojedyncze osoby, lecz całe grupy np. klasy dzieci w szkole, konkretne społeczności. Jednak przeprowadzenie RCT klasterowego nie rozwiązuje problemu kosztochłonności wykonania takiego badania ani nie niweluje trudności dot. zastosowania wieloletniego horyzontu czasowego. Dlatego też przy ocenianiu interwencji/ działań z zakresu zdrowia publicznego zasadne i dopuszczalne może być wykorzystanie wyników badań nieeksperymentalnych (np. długofalowego badania obserwacyjnego), które pomimo zajmowania niższego miejsca w hierarchii dowodów naukowych, może lepiej odzwierciedlać specyfikę programów z zakresu zdrowia publicznego. Innym rozwiązaniem problemu dostępności danych wykorzystywanych w wieloletnim horyzoncie czasowym, może być wykorzystanie technik modelowania. Umożliwiają one rozszerzenie wyników, które pochodzą z pojedynczego badania (np. ekstrapolację poza horyzont czasowy) i/lub kojarzenie dowodów pochodzących z różnych źródeł. Znane metody modelowania obejmują np. drzewo decyzyjne, model Markova, symulację zdarzeń dyskretnych. Modelowanie takie może stanowić bardzo pomocne narzędzie umożliwiające uchwycenie wielowymiarowych efektów programów zdrowia publicznego.
5. Szacowanie kosztów w interwencjach z zakresu zdrowia publicznego jest ściśle związane z aspektami analizy opisanymi powyżej w pkt. 1-4. Poprzez przyjęcie perspektywy społecznej konieczna jest identyfikacja, pomiar i wycena wszystkich kosztów związanych z ocenianym programem, bez względu na to, kto je ponosi. Koszty te sklasyfikowane są jako koszty bezpośrednie (medyczne i niemedyczne) oraz pośrednie (utracona produkcja z powodu chorobowości i umieralności). Oprócz uniwersalnych trudności obserwowanych w analizach ekonomicznych, dotyczących wyceny kosztów opieki nieformalnej, utraconej produktywności, uzyskanych lat życia, w aspekcie programów z zakresu zdrowia publicznego, dodatkową problematyczną kwestią jest szacowanie kosztów międzysektorowych. Dla przykładu można podać program dot. zapobiegania uzależnieniom od środków odurzających – może on zapobiegać wystąpieniu przyszłych kosztów nie tylko w sektorze ochrony zdrowia (koszty leczenia), lecz także w sektorze opieki społecznej czy sektorze więziennictwa. W takiej sytuacji powinno uwzględnić się wszystkie rodzaje kosztów. Rodzaje kosztów, gdzie nie jest możliwy ilościowy i/lub monetarny opis kosztów, można przedstawić w sposób opisowy. W takiej sytuacji wskazane jest zastosowanie analizy kosztów i konsekwencji (*costs-consequence analysis*, CCA), gdzie koszty i efekty mogą być szacowane osobno dla różnych sektorów i/lub podmiotów zaangażowanych w program. W analizie CCA nie wylicza się finalnej miary efektywności kosztowej programu (np. współczynnika kosztów i efektów), dlatego też nie ma możliwości sporządzenia zestawienia porównywanych programów. Należy zaznaczyć, że mimo wszystko analiza CCA daje możliwość przedstawienia pełnego obrazu kosztów, a jej wyniki mogą posłużyć do przeprowadzenia analizy kosztów i efektów – oddzielnie dla każdego z sektorów, na który oddziałuje oceniany program.

Alternatywną formą prezentacji kosztów międzysektorowych może być wyliczenie korzyści netto z danej interwencji (*net-benefit approach*) dla każdego z sektorów zaangażowanych w program. Polega to na tym, że gdy w wyniku programu w jednym sektorze otrzymuje się dodatnie korzyści netto, a w innym sektorze – ujemne efekty (straty), można zastosować tzw. test kompensacyjny. Przy jego wykorzystaniu dany program może być uznany za efektywny kosztowo, a co za tym idzie zalecany do realizacji, pod warunkiem gdy sektor ochrony zdrowia, który otrzymuje dodatnie korzyści netto, może skompensować straty innego sektora. Należy jednak zaznaczyć, że w praktyce taka rekompensata może nie być możliwa do uzyskania w krótkim czasie, co będzie wymagało prowadzenia bilansu relokacji środków między sektorami w określonym przedziale czasu.

6. Szacowanie efektów uzależnione jest przyjętą perspektywą oceny oraz rodzajem analizy. W przypadku programów z zakresu zdrowia publicznego najistotniejsze znaczenie ma wielowymiarowy charakter efektów. Zgodnie z klasyfikacją Nutbeam 1998 (48) efekty programu promocji zdrowia można podzielić na trzy następujące po sobie w czasie poziomy. Pierwszy poziom dotyczy bezpośrednich efektów realizacji programu tj. wpływu na

indywidualne, społeczne i środowiskowe czynniki kształtujące zdrowie (np. na poziomie jednostkowym – podniesienie poziomu wiedzy i świadomości w zakresie zachowań prozdrowotnych; na poziomie społeczności – zwiększenie kontroli nad lokalnymi uwarunkowaniami zdrowia, kształtowanie, promowanie lokalnych polityk promujących zdrowie). Efekty bezpośrednie programu kształtują oraz wpływają na wystąpienie pośrednich efektów zdrowotnych tj. poziomu drugiego, co za tym idzie – zmian w modyfikowalnych determinantach zdrowia (np. zmiany w stylu życia czy poprawa usług zdrowotnych). Dopiero końcowe efekty programu (poziomu trzeciego) odnoszą się bezpośrednio do zdrowia (poziom chorobowości, umieralności, niepełnosprawności) i efektów społecznych (jakość życia, samodzielność, redukcja nierówności w zdrowiu).

Dla decydenta istotna jest informacja o koszcie uzyskania jednostki efektu końcowego (np. koszt uniknięcia jednego przypadku zgonu; koszt uzyskania jednego roku życia korygowanego jakością – QALY). Należy jednak zaznaczyć, że oszacowanie efektów końcowych wymaga analizy uwzględniającej długofalowy horyzont czasowy lub wykorzystanie technik modelowania. Jednak zawsze należy uwzględnić zakłócający wpływ czynników zewnętrznych. Najbardziej rozpowszechnione miary efektów końcowych np. QALY mogą nie być odpowiednio czułe na wychwycenie specyficznego, najczęściej wieloaspektowego, charakteru programu z zakresu zdrowia publicznego (46, 47). Podobnie jak przy kosztach, wspomniane oddziaływanie programu z zakresu zdrowia publicznego, na wiele sektorów, wskazuje na konieczność uwzględnienia wielu odbiorców i różnych miar efektów badanych interwencji. Kolejnym utrudnieniem w analizach ekonomicznych programów z zakresu zdrowia publicznego jest to, że specyfika pomiaru efektów polega także na tym, że bezpośrednie oraz pośrednie efekty zdrowotne (programu) mogą być równie ważne co efekty końcowe. Niektóre efekty mogą być trudne do zmierzenia np. zmiany behawioralne, które w konsekwencji wpływają na poprawę zdrowia. Jednak celem zapewnienia kompleksowości analizy uwzględnienie takich efektów ma istotne znaczenie. Dlatego analogicznie jak w przypadku kosztów, przedstawia się wszystkie miary efektów i/lub zastosować analizę CCA. W sytuacji, gdy nie jest możliwe przedstawienie miary efektów w sposób ilościowy, dopuszcza się opisową formę ich przedstawienia.

Dodatkowym aspektem dotyczącym szacowania efektów programów z zakresu zdrowia publicznego, jest uwzględnienie celu równości w zdrowiu. Ze względu na to, że zmniejszenie nierówności w zdrowiu jest jednym z istotniejszych celów polityki w zdrowiu publicznym, niezmiernie ważnym jest uwzględnienie tego kryterium w analizach ekonomicznych. Dlatego w procesie ewaluacji ekonomicznej programów z zakresu zdrowia publicznego należy podjąć próbę oceny wpływu ocenianego programu na poziom nierówności w zdrowiu, przy zastosowaniu metod opisowych lub jeśli to możliwe także analitycznych.

7. Kontekst w jakim program ma być zaimplementowany może wpływać na jego efektywność kosztową. W aspekcie kontekstu istotne są dodatkowe uwarunkowania socjoekonomiczne, polityczne, środowiskowe populacji docelowej programu, które mogą oddziaływać na efektywność programu. Kontekst programu z zakresu zdrowia publicznego obejmuje także obecność innych komplementarnych programów prozdrowotnych lub efektów programów, które były wdrażane w przeszłości. Dlatego też możliwości przeniesienia wyników oceny efektywności kosztowej danego programu może być ograniczona i/lub wymagać dodatkowych prac uwzględniających specyfikę konkretnej populacji docelowej. Dlatego w procesie ewaluacji ekonomicznej programów z zakresu zdrowia publicznego należy szczegółowo opisać kontekst ocenianej interwencji z łącznym przedstawieniem wyników oceny ekonomicznej.

Edwards 2013 (47) w przeglądzie systematycznym dot. zaleceń z zakresu analiz ekonomicznych w zdrowiu publicznym stworzył checklistę, która może być pomocna podmiotom wykonującym analizy (decydom, ekonomistom, innym zainteresowanym), w wyborze i rozważaniu właściwej metodologii oceny ekonomicznej w odniesieniu do specyfiki interwencji z zakresu zdrowia publicznego. 12 następujących pytań obejmuje:

1. Jakie są ramy teoretyczne danej analizy? Czy bliższe są np. analizie kosztów-korzyści tj. perspektywa badawcza *welfarist* czy analizie efektywności kosztów tj. perspektywa badawcza *extra-welfarist* lub podejściu zorientowanym na możliwości (*capability approach*).
2. Jakie są warunki realizacji interwencji poddawanej ocenie? Np. osiągnięcie danej zmiany środowiskowej, kontrola chorób zakaźnych, badania przesiewowe, zmiana zachowań wpływających na poprawę sytuacji zdrowotnej, wspieranie zmian legislacyjnych/polityki zdrowotnej.
3. Czy interwencja poddawana ocenie stanowi profilaktykę pierwotną, wtórną, trzeciorzędową?
4. Jaki podmiot jest odpowiedzialny za wdrożenie ocenianej interwencji (np. rząd, system opieki zdrowotnej, samorząd lokalny, wolontariat) i kim są główni interesariusze?
5. Jeśli oceniana interwencja ma na celu zmianę danych zachowań, to jakie są kluczowe narzędzia wprowadzające te zmiany? Np. zmiany w ustawodawstwie, zmiany w przyjętych normach społecznych.
6. Jaki jest odpowiedni horyzont czasowy analizy i najbardziej odpowiednia stopa dyskontowa dla kosztów i efektów?
7. Jeśli oceniana interwencja ma na celu „przesunięcie krzywej”, to czy jesteśmy najbardziej zainteresowani wynikami znajdującymi się w jej centrum czy w „ogonie” otrzymanego rozkładu?

8. W jaki sposób oceniana interwencja może oddziaływać na nierówności zdrowia np. dostęp do świadczeń w obszarze wiejskim i miejskim?
9. Czy analiza podgrup pomoże w identyfikacji zakresu szacunków efektywności kosztowej w obrębie różnych ustawień analizy, przyjętych metod i grup populacji?
10. Jakie są główne miary wyników np. QALY, DALY lub szeroka gama mierników związanych ze zdrowiem i aspektami społecznymi?
11. Jak ważna jest wycena kosztów, korzyści i zysków w kategoriach pieniężnych? Czy od ocenianej interwencji można rozsądnie oczekiwać, że wygeneruje oszczędności w krótkiej, średniej i długiej perspektywie czasu?
12. Jak istotne będzie zestawienie ICER z progiem opłacalności przyjętym w danym kraju?

Przeprowadzając ewaluację ekonomiczną programów lub działań z zakresu zdrowia publicznego, należy zawsze uwzględnić indywidualną specyfikę danego programu/ działania, a w konsekwencji przygotować zindywidualizowane podejście metodologiczne (46, 47).

4. Organizacja oraz monitorowanie i ewaluacja programów polityki zdrowotnej

4.1. Opcjonalne technologie medyczne

W przypadku tworzonego Raportu weryfikuje się czy oceniana technologia/ działanie ma opcjonalną (-e) technologię (-e) medyczną (-e). W zakresie definiowania technologii opcjonalnych, na podstawie odnalezionych danych naukowych porównuje się skuteczność, bezpieczeństwo a także efektywność kosztową ocenianej interwencji z interwencją opcjonalnie stosowaną w tej samej populacji docelowej (jeśli to możliwe). Na podstawie odnalezionych danych wykazuje się, czy oceniana technologia medyczna/działanie przeprowadzane w ramach PPZ, niesie za sobą wartość dodaną w stosunku do interwencji opcjonalnych. Należy jednak podkreślić, że bardzo często wykorzystywana jest „*do-nothing option*”, a więc brak komparatora.

Źródłem informacji na temat interwencji opcjonalnych mogą być:

- konsultacje z ekspertami klinicznymi w szczególności odpowiednimi dla danej choroby lub danego problemu zdrowotnego,
- informacje pochodzące z odnalezionych rekomendacji i/lub badań,
- wykazy świadczeń gwarantowanych.

4.2. Stan finansowania interwencji ze środków publicznych w Polsce

Przy tworzeniu Raportu przedstawia się aktualny stan finansowania technologii/działań w ramach PPZ ze środków publicznych w Polsce, z uwzględnieniem informacji czy np. stanowią świadczenia gwarantowane, interwencje w ramach Programu Szczepień Ochronnych, PPZ realizowane przez Ministra Zdrowia lub programy zdrowotne.

Dodatkowo opisuje się aktualny stan finansowania ze środków publicznych w Polsce opcjonalnych technologii medycznych (jeśli zostały zdefiniowane).

Przy określaniu aktualnego stanu finansowania należy uwzględnić czy dane interwencje realizowane są np. w ramach podstawowej, ambulatoryjnej opieki zdrowotnej, opieki paliatywnej, lecznictwa szpitalnego, uzdrowiskowego.

4.3. Warunki realizacji interwencji w ramach programów

Należy opisać warunki realizacji ocenianych interwencji w ramach PPZ dotyczących danego problemu zdrowotnego.

Źródłem informacji na temat warunków realizacji ocenianych technologii/działań oraz interwencji opcjonalnych mogą być:

- wytyczne praktyki klinicznej,
- konsultacje z ekspertami klinicznymi (o specjalności adekwatnej do danego problemu zdrowotnego),
- warunki realizacji danych interwencji opisane w wykazach świadczeń gwarantowanych,
- informacje z odnalezionych badań (z uwzględnieniem aspektu stosowalności).

Przy określaniu warunków realizacji należy uwzględnić warunki polskie lub posłużyć się podsumowaniem wyników odnalezionych badań (jeżeli były dostępne) przeprowadzonych w różnych warunkach i w różnych populacjach. Ważne jest zbadanie szczegółów dotyczących procesu wdrażania interwencji jak i kontekstu przed ekstrapolowaniem wyników z poszczególnych badań i przeprowadzeniem dalszej syntezy wyników. Aspekty związane ze stosowalnością, czyli stopniem w jakim dana interwencja mogłaby być wdrożona w innych warunkach, oraz możliwością przenoszenia czyli stopnia, w jakim mierzona skuteczność wdrożonej już interwencji, może zostać osiągnięta w innych warunkach, zostały szczegółowo opisane w rozdziale 3.13.

4.4. Monitorowanie i ewaluacja

Programy z zakresu zdrowia publicznego mają na celu zapobieganie lub kontrolę danej choroby (wystąpienie choroby), urazu, niepełnosprawności lub zgonu. Coraz częściej programy odnoszą się w sposób kompleksowy do wybranych problemów zdrowotnych. Dlatego też ważne jest angażowanie danej społeczności i organizacji oraz tworzenie szerokich koalicji/ porozumień. Najczęściej problemy zdrowia publicznego wiążą się ze znaczącymi i trudnymi zmianami postaw i zachowań ryzykownych, ale i zachowań prozdrowotnych odbiorców określonych interwencji i/lub ich realizatorów. Ponadto kontekst, w którym realizowane są programy z zakresu zdrowia publicznego, stał się bardziej złożony. Programy, które z powodzeniem realizowane są w pewnych warunkach, zawodzą w innych np. z powodów finansowych, socjoekonomicznych, demograficznych czy organizacyjnych. Także ze względu na dużo większą kompleksowość programów, wzrasta odpowiedzialność decydentów i innych zainteresowanych stron. Wyżej wspomniane zmiany warunków,

w których realizowane są programy, wskazują na to, że ewaluacja programów jest niezbędna. Należy zaznaczyć, że nie ma jednego uniwersalnego sposobu ewaluacji. Istotnym jest dokumentowanie i pomiar realizacji danego programu – ma to znaczenie w powodzeniu osiągnięcia zamierzonych celów, a także przy rozliczaniu się z kluczowymi interesariuszami (49).

Proces ewaluacji danego programu zachęca realizatorów i podmioty odpowiedzialne do weryfikacji funkcjonowania programu. Szczególnie w obrębie prowadzonych działań, podmiotów realizujących i tego, kto w rezultacie osiąga korzyści. Dzięki ewaluacji programu można oszacować czy interwencje/działania są realizowane z wcześniej założonym planem, a także zidentyfikować mocne i słabe strony programu oraz obszary wymagające poprawy. Większość programów z zakresu zdrowia publicznego ma na celu zmianę zachowań w danej grupie docelowej i stworzenie środowiska, które wzmacnia trwałe wdrożenie tych zmian. Oczekuje się, że zmiany w danym środowisku i zachowaniach będą w stanie zapobiegać i kontrolować określone problemy zdrowotne, choroby i urazy. Dzięki ewaluacji można śledzić oczekiwane zmiany i oceniać skuteczność oraz wpływ konkretnego programu, interwencji lub strategii w ich tworzeniu. Powody, dla których powinno się prowadzić ewaluację programów z zakresu zdrowia publicznego są następujące:

- monitorowanie postępu w zakresie celów programu,
- określenie czy dane komponenty programu generują pożądany postęp w zakresie oczekiwanych wyników,
- umożliwienie porównania działania danej interwencji w obrębie danych grup docelowych (przede wszystkim wśród populacji o nieproporcjonalnie wysokich czynnikach ryzyka i niekorzystnych efektach zdrowotnych),
- uzasadnienie potrzeby dalszego finansowania i wsparcia danego programu,
- określenie możliwości dla ciągłej poprawy jakości danego programu,
- zagwarantowania, że skuteczne i efektywne programy będą utrzymywane, a zasoby nie będą marnowane na nieskuteczne działania (49).

Źródło wskaźników służących do monitorowania i ewaluacji zalecanych technologii medycznych, działań przeprowadzanych w ramach programów polityki zdrowotnej mogą stanowić:

- rekomendacje/wytyczne kliniczne,
- punkty końcowe uwzględnione w odnalezionych opracowaniach wtórnych (przeglądach literatury) lub badaniach pierwotnych (np. RCT),
- konsultacje z ekspertami klinicznymi w szczególności odpowiednimi dla danego problemu zdrowotnego,
- konsultacje z podmiotami wdrażającymi i realizującymi programy polityki zdrowotnej (np. poprzez przygotowaną ankietę na danej platformie internetowej),
- projekty programów polityki zdrowotnej (zakończone raportem końcowym).

Istotnym elementem jest to, aby określone wskaźniki monitorowania i ewaluacji były możliwe do zaimplementowania przez podmioty, które będą wdrażały i realizowały programy polityki zdrowotnej.

W zakresie monitorowania powinny zostać określone wskaźniki, które będą odnosić się do:

- zgłaszalności (odsetek lub liczba: uczestników programu polityki zdrowotnej; osób, którzy nie zostali objęci działaniami programu z powodów zdrowotnych lub innych);
- jakości świadczeń (np. ankieta satysfakcji).

W związku z tym, że często granica między monitorowaniem i ewaluacją jest płynna (49), informacje pochodzące z monitorowania mogą zostać wykorzystane do ewaluacji programu.

W zakresie ewaluacji programu powinno odnieść się do efektów zdrowotnych uzyskanych i utrzymujących się po zakończeniu programu (np. odsetek nowo zdiagnozowanych chorych, odsetek osób u których obniżono masę ciała itp.) (50). Efekty zdrowotne są to zmiany, które z założenia będą wynikać z wdrażanego programu z zakresu zdrowia publicznego. W kompleksowych programach ukierunkowanych na rozbudowane problemy zdrowia publicznego, końcowy efekt jest często ambitny i długotrwały np. całkowite wyeliminowanie danego problemu lub stanu, poprawa jakości życia osób dotkniętych danym problemem zdrowotnym. W Aneks 12 przedstawiono potencjalną hierarchię efektów danego programu. Zaczynając od podstawy hierarchii:

1. działania programu mają na celu uzyskanie uczestnictwa w obrębie populacji docelowej;
2. reakcje/oddziaływania uczestników (populacji docelowej) na działania programu wpływają na ich naukę (edukację) – wiedzę, opinie, umiejętności i ambicje;

- poprzez proces nauki, osoby oraz organizacje podejmują działania, które skutkują zmianą warunków społecznych, behawioralnych i/lub środowiskowych, które wpływają na uzyskanie długoterminowych efektów zdrowotnych w danej społeczności.

Należy mieć na uwadze, że efekty znajdujące się wyżej w hierarchii stanowią zazwyczaj główny powód, dla którego program został utworzony. Mimo tego że koszty i trudności związane ze zbieraniem danych wzrastają wraz z kolejnym poziomem w hierarchii efektów. Informacje pochodzące z niższych poziomów pomagają w zrozumieniu wyników otrzymanych na wyższych poziomach, możliwych do zaobserwowania w dłuższej perspektywie.

Schemat poprawnie skonstruowanej ewaluacji programów z zakresu zdrowia publicznego, zgodnie z zaleceniami *Centers for Disease Control and Prevention* (49), powinien zawierać sześć jasno określonych etapów działań (Rysunek 1).



Rysunek 1. Struktura ewaluacji programu w zakresie zdrowia publicznego (sześć etapów) wg CDC

Źródło: opracowanie własne na podstawie CDC 2011. Introduction to program evaluation for public health programs: A self-study guide

Pierwszy etap zalecanego schematu ewaluacji stanowi zaangażowanie potencjalnych interesariuszy (tj. osób/ organizacji, które inwestują swoje zasoby w realizację danego programu, podmiotów zainteresowanych wynikami ewaluacji). Ich perspektywa może istotnie wpływać na każdy etap struktury ewaluacji programu. Kolejnym etapem jest kompleksowy opis programu (projekt programu), który ma na celu wyjaśnienie wszystkich jego komponentów i oczekiwanych efektów. Dzięki rzetelnemu przygotowaniu opisu, proces ewaluacji będzie skoncentrowany na najważniejszych aspektach programu. Wśród jego komponentów wymienia się: potrzeby (czyli jak duże znaczenie ma wybrany program z perspektywy zdrowia publicznego), cele, oczekiwane efekty, działania/interwencje, produkty, zasoby/wkład wejściowy, związek zastosowanych działań z uzyskaniem oczekiwanych efektów programu, etap rozwoju (czy realizacja programu dopiero się rozpocznie, czy już jest wdrażany, czy też trwa już od dłuższego czasu) oraz kontekst (jakie czynniki/ogólne trendy mogą wpływać na sukces bądź fiasko programu). Należy podkreślić, że na danym etapie procesu ewaluacji można zastosować narzędzie jakim jest macierz logiczna (*logic models*), jak również uwzględnić się tzw. kryteria ewaluacji, opisane poniżej.

Wytyczne CDC (49) podkreślają znaczenie czterech kryteriów ewaluacji (Aneks 13), tj. użyteczności (*utility*), wykonalności/opłacalności (*feasibility*), stosowności (*propriety*) oraz trafności (*accuracy*). Kryteria te mają ścisły związek z planowanymi składowymi programu (ww. opisem), jak i jego macierzą logiczną. Należy podkreślić, że modele logiczne stanowią graficzny opis zależności między działaniami w programie, a jego oczekiwanymi efektami. Mogą one przedstawiać wszystkie lub tylko niektóre z elementów opisu programu z zakresu zdrowia publicznego, w zależności od przeznaczenia:

- wkład (dane wejściowe; *inputs*) – zasoby wchodzące w skład programu i uzależnione od jego działań,
- interwencje/działania (*activities*) – konkretne działania/wydarzenia wykonywane w trakcie programu,
- wydajność/produkty (*outputs*) – bezpośrednie efekty działań/interwencji w programie, mają charakter statyczny i opisują stan po programie, często mierzone w kategoriach policzalnych, jak np. liczba sesji terapeutycznych,

- punkty/rezultaty końcowe (*outcomes*) – zmiany wynikające z działań/interwencji w programie oraz ich bezpośrednich efektów, mają charakter dynamiczny i obrazują zmianę między okresem przed programem i po jego realizacji, często wyrażane jako rezultaty/ oddziaływania krótkoterminowe, pośrednie lub długoterminowe.

Przykład wykorzystania wyżej opisanego modelu oraz jego graficzny schemat przedstawiono w Aneks 14.

Należy podkreślić, że w zakresie monitorowania programu z zakresu zdrowia publicznego nie uwzględnia się elementu modelu logicznego, jakim jest oddziaływanie długoterminowe.

Trzecim etapem w opisywanej strukturze ewaluacji jest skoncentrowanie się już na samym procesie oceny, gdzie zarówno potencjalni interesariusze, jak i osoby odpowiedzialne za realizację programu rozumieją jego założenia i wspólnie konstruują stosowne pytania ewaluacyjne. Etap ten skupia się na wyborze właściwego typu ewaluacji: dotyczący implementacji programu oraz dotyczący jego skuteczności i punktów końcowych/ efektów zdrowotnych.

Ewaluacja dotycząca implementacji programu dokumentuje czy program został wdrożony zgodnie z jego przeznaczeniem. W procesie ewaluacji można sprawdzić, czy zaplanowane działania są realizowane, kto je prowadzi, kto czerpie z nich korzyści oraz czy zmobilizowano i zapewniono odpowiednie zasoby/ wkład. We wszystkich przypadkach proces ewaluacji sprawdza, czy rzeczywista wydajność programu była zgodna z zamierzeniem. Takie pomiary mogą odnosić się do rzeczywistej i zaplanowanej pierwotnie wydajności i realizacji, oraz mogą obejmować:

- obszar, gdzie realizowane są działania programu (wiejski, miejski),
- liczbę osób objętych działaniami programu,
- status ekonomiczny osób korzystających z programu,
- jakość działań programu,
- rzeczywiste wydatki na realizację programu,
- koszty bezpośrednie i wydatki majątkowe na realizację działań programu,
- personel realizujący działania programu,
- liczbę przeprowadzonych interwencji/działania w ramach programu.

W sytuacji, gdy źródła/ zasoby na ewaluację są ograniczone, mogą zostać uwzględnione jedynie najbardziej istotne elementy w zakresie ewaluacji dot. implementacji. W tym przypadku, w procesie ewaluacji można skoncentrować się na:

- przeniesieniu odpowiedzialności: w sytuacji gdy działania programu nie przynoszą zamierzonych rezultatów i inny podmiot (osoba lub organizacja) podejmie odpowiednie działania - następuje przeniesienie odpowiedzialności na nowy podmiot;
- intensywności działań: poprzez założone efekty płynące z realizacji interwencji programu można założyć progowy poziom uczestnictwa w programie lub ekspozycji na wdrażane interwencje;
- dostępie do programu: gdy założone rezultaty programu zakładają nie tylko wzrost zapotrzebowania (popyt) na dane interwencje, ale także zwiększenie podaży tych usług - wtedy proces ewaluacji może obejmować pomiar dostępności do świadczeń programu;
- kompetencjach personelu: gdy założone rezultaty programu zakładają dobrze zaprojektowane działania programu realizowane przez personel, który posiada odpowiednie kompetencje oraz został odpowiednio dobrany do odbiorców – wtedy w procesie ewaluacji można uwzględnić stopień „dopasowania” personelu do populacji docelowej odbiorców.

Ewaluacja dot. punktów końcowych/efektów zdrowotnych ocenia, w jakim stopniu założone efekty programu zostały osiągnięte w trakcie jego trwania. W programach często stosuje się następujące określenia dla sekwencji punktów końcowych/efektów tj. efekty krótkoterminowe, pośrednie, długoterminowe lub bliższe i dalsze. W zależności od etapu programu i celu ewaluacji, ewaluacja efektów zdrowotnych może obejmować dowolne lub wszystkie wyniki w obrębie wyżej opisanych sekwencji, w tym:

- zmiany w nastawieniu osób uczestniczących w programie oraz ich przekonaniach,
- zmiany w zachowaniach ryzykownych i ochronnych,
- zmiany w środowisku, w tym polityki, formalnych i nieformalnych sposobach egzekwowania określonych regulacji oraz wpływ norm społecznych i innych uwarunkowań społecznych,
- zmiany we wskaźnikach zachorowalności i umieralności.

W zakresie ewaluacji procesu implementacji i w odniesieniu do efektów zdrowotnych istnieje kilka innych typów pytań ewaluacyjnych, które obejmują:

- efektywność: czy realizacja programu jest zapewniona przy minimalnym wykorzystaniu zasobów zarówno finansowanych, jak i czasu pracy?; jaka jest wielkość „produktów” (np. wykonanych interwencji) programu generowanych przez zasoby przeznaczone na program?
- efektywność kosztów: czy wartość lub korzyść wynikająca z efektów/ rezultatów programu przewyższa koszt ich wytworzenia?
- przypisanie (autorstwa): czy uzyskane wyniki programu są związane z programem, czy może to być efekt innych wydarzeń/ działań (nie mających związku z programem) prowadzonych w tym samym czasie?

Powyższe pytania odnoszą się jednak do części, a nie całego modelu logicznego stosowanego w ewaluacji.

Ustalenie właściwego kierunku ewaluacji jest decyzją indywidualną w przypadku każdego programu. Kilka zaleceń inspirowanych głównymi kryteriami ewaluacji tj. użyteczności i wykonalności/opłacalności, może pomóc w określeniu najlepszego kierunku ewaluacji (Aneks 15).

Kolejnym etapem w ramach struktury ewaluacji - w wyniku nakreślenia konkretnego typu/obszaru ewaluacji, wyboru pytań ewaluacyjnych i opracowania matrycy logicznej wraz z rezultatami/punktami końcowymi – jest zgromadzenie rzetelnych dowodów wpływających na ostateczne wnioski płynące z ewaluacji. Zebranie dowodów powinno uwzględniać zarówno wskaźniki, źródła dowodów i ich metodologię, opis ich jakości oraz ilości. Kluczową kwestią jest to, czy istnieją już odpowiednie źródła danych (wtórne gromadzenie danych), czy do zmierzenia interesujących nas wskaźników niezbędne będzie zebranie nowych informacji (pierwotne gromadzenie danych). W zależności od założonych pytań ewaluacyjnych i określonych wskaźników, niektóre istniejące już źródła danych (dane wtórne) mogą być odpowiednie i przydatne. Wśród istniejących już źródeł danych przydatnych w ocenie punktów końcowych w zakresie programów ze zdrowia publicznego CDC 2011 wymienia:

- aktualne dane demograficzne pochodzące np. ze spisu powszechnego,
- *Behavioral Risk Factor Surveillance System* (BRFSS),
- *Youth Risk Behavior Survey* (YRBS),
- *Pregnancy Risk Assessment Monitoring System* (PRAMS),
- rejestry nowotworów,
- bazy danych gromadzące informacje w zakresie obserwacji pacjentów,
- *National Health Interview Survey* (NHIS),
- inne istotne źródła statystyczne.

W odniesieniu do programów planowanych do realizacji na terenie Polski, należy dostosować i przeszukać źródła odpowiednie dla populacji polskiej i dla danego zakresu tematycznego programu. Przed wykorzystaniem wtórnych źródeł danych, należy upewnić się, że spełniają założone wcześniej oczekiwania. CDC 2011 zwraca uwagę na to, że chociaż duże systemy danych mają zalety polegające na rutynowym gromadzeniu danych, niektóre z nich charakteryzują się niewielką elastycznością, co może sprawiać trudności w wykorzystaniu ich w ewaluacji zaplanowanego programu.

Metody pierwotnego zbierania danych można również podzielić na kilka szerokich kategorii. Do najbardziej powszechnych należą:

- ankiety, obejmujące wywiad osobisty, wywiad przez telefon, informacje uzupełnione przez respondenta otrzymane pocztą lub drogą elektroniczną,
- grupy dyskusyjne, grupy tematyczne,
- obserwacja,
- źródła dokumentacji medycznej.

Przedostatnim elementem struktury ewaluacji wg CDC 2011 jest przedstawienie i uzasadnienie wniosków płynących ze wcześniejszych etapów procesu. Niezależnie od głównego założenia ewaluacji programu, np. wykazania skuteczności programu, poprawy programu czy też pociągnięcia do odpowiedzialności stosownych jednostek, analizuje się i interpretuje wszystkie dowody zebrane w etapie czwartym ewaluacji. W związku z tym, przeprowadza się dokładną analizę i syntezę punktów końcowych, porównać je z wcześniejszymi zidentyfikowanymi/określonymi standardami (jak np. potrzeby uczestników, regulacje prawne, efektywność podobnych programów), przedstawić wnioski płynące z ww. analiz i ich uzasadnienie.

Rzetelnie przygotowane wnioskowanie na temat ww. punktów końcowych i standardów stanowi podstawę dla ostatniego etapu ewaluacji, jakim jest wykorzystanie wyników przez interesariuszy i realizatorów programu. Rezultaty te mogą zostać wykorzystane do przedstawienia skuteczności programu, określenia sposobów jego udoskonalenia/poprawy, modyfikacji w zakresie planowania kolejnych programów czy wykazania odpowiedzialności i uzasadnienia potrzeby finansowania.

5. Bibliografia

1. Oliver, K., Rees, R., Brady, L. M., Kavanagh, J., Oliver, S., & Thomas, J. (2015). Broadening public participation in systematic reviews: a case example involving young people in two configurative reviews. *Research synthesis methods*, 6(2), 206-217
2. O'Mara-Eves, A., Brunton, G., Oliver, S., Kavanagh, J., Jamal, F., & Thomas, J. (2015). The effectiveness of community engagement in public health interventions for disadvantaged groups: a meta-analysis. *BMC public health*, 15(1), 129
3. EUnetHTA (2017). Guidelines: Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness. Version 1.2, December 2017 https://www.eunetha.eu/wp-content/uploads/2018/01/Guideline_Information_Retrieval_V1-2_2017.pdf (dostęp w dniu 18.12.18)
4. American Psychiatric Association. (2017). The Diagnostic and Statistical Manual of Mental Disorders (DSM-5). <http://www.dsm5.org/psychiatrists/practice/dsm> (dostęp w dniu 18.12.18)
5. Rozporządzenie Ministra Zdrowia z dnia 27 lutego 2018 r. w sprawie priorytetów zdrowotnych, Dz.U. 2018 poz. 469
6. Mapy Potrzeb Zdrowotnych, Ministerstwo Zdrowia <http://www.mpz.mz.gov.pl/> (dostęp w dniu 18.12.18)
7. CfRa, D. (2009). CRD's guidance for undertaking reviews in health care. York Publishing Services Ltd, 32
8. O'Neill, J., Tabish, H., Welch, V., Petticrew, M., Pottie, K., Clarke, M., ... & Tugwell, P. (2014). Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of clinical epidemiology*, 67(1), 56-64
9. Cochrane Collaboration. (2011). Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Chapter 6. <http://handbook-5-1.cochrane.org/> (dostęp w dniu 18.12.18)
10. EUnetHTA Joint Action 2, (2016). Work Package 8. HTA Core Model ® version 3.0 for the full assessment of Diagnostic Technologies, Medical and Surgical Interventions, Pharmaceuticals and Screening Technologies. s. 146-148. <http://eunetha.eu/sites/5026.fedimbo.belgium.be/files/HTACoreModel3.0.pdf> (dostęp w dniu 18.12.18)
11. Cochrane Collaboration. (2011). Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Chapter 13. <http://handbook-5-1.cochrane.org/> (dostęp w dniu 18.12.18)
12. Wytyczne przeprowadzania Oceny Technologii Medycznych (HTA). Załącznik do Zarządzenia nr 40/2016 Prezesa Agencji Oceny Technologii Medycznych i Taryfikacji z dnia 13 września 2016 r. w sprawie wytycznych oceny świadczeń opieki zdrowotnej http://www.aotm.gov.pl/www/wp-content/uploads/wytyczne_hta/2016/20160913_Wytyczne_AOTMiT.pdf (dostęp w dniu 18.12.18)
13. Stansfield, C., Kavanagh, J., Rees, R., Gomersall, A., & Thomas, J. (2012). The selection of search sources influences the findings of a systematic review of people's views: a case study in public health. *BMC medical research methodology*, 12(1), 55.
14. McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276-282
15. Cochrane Collaboration. (2011). Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Chapter 21.4 Assessment of study quality and risk of bias. <http://handbook-5-1.cochrane.org/>, (dostęp w dniu 18.12.18)
16. Cochrane Collaboration. (2017). Cochrane handbook for systematic reviews of interventions. Version 5.2.0. Chapter 8.5. <https://training.cochrane.org/handbook/pdf-versions> (dostęp w dniu 18.12.18)
17. Halpern, S. H., & Douglas, M. (2005). Jadad scale for reporting randomized controlled trials. *Evidence-based obstetric anesthesia*, 237-238
18. Cochrane Collaboration. (2017). Cochrane handbook for systematic reviews of interventions. Version 5.2.0. Chapter 8. Assessing risk of bias in included studies. <https://training.cochrane.org/handbook/pdf-versions> (dostęp w dniu 18.12.18)
19. EUnetHTA (2017). Guidelines: Internal validity of non-randomised studies (NRS) on interventions. https://www.eunetha.eu/wp-content/uploads/2018/01/Internal-validity-of-non-randomised-studies-NRS-on-interventions_Guideline_Final-Jul-2015.pdf (dostęp w dniu 18.12.18)
20. NICE. (2003). Quality assessment for case series. NICE guidelines (CG3); June 2003. <https://www.nice.org.uk/guidance/ng45/evidence/appendix-o-cg3-full-guideline-pdf-87258149466> (dostęp w dniu 18.12.18)

21. EUnetHTA. (2017). Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness. s. 32. http://www.eunetha.eu/sites/default/files/Guideline_Information_Retrieval_V1-2_2017.pdf (dostęp w dniu 18.12.18)
22. Quality Assessment Tool – Review Articles <https://www.healthevidence.org/documents/our-appraisal-tools/quality-assessment-tool-dictionary-en.pdf> (dostęp w dniu 18.12.18)
23. National Collaborating Centre for Methods and Tools (2017). AMSTAR2: Appraisal tool for systematic reviews of randomized and observational studies (non-randomized studies). Hamilton, ON: McMaster University. (Updated 22 February, 2018). <http://www.nccmt.ca/knowledge-repositories/search/307> (dostęp w dniu 18.12.18)
24. Shea, B., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., et al. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358, j4008.
25. National Collaborating Centre for Methods and Tools. (2011). AMSTAR: assessing methodological quality of systematic reviews. On: McMaster University. <http://www.nccmt.ca/registry/view/eng/97.html> (dostęp w dniu 18.12.18)
26. Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., ... & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, 7(1), 10
27. Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., ... & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of clinical epidemiology*, 62(10), 1013-1020
28. Whiting, P., Savović, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., ... & Churchill, R. (2016). ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *Journal of clinical epidemiology*, 69, 225-234.
29. Robinson, K. A., Whitlock, E. P., Oneil, M. E., Anderson, J. K., Hartling, L., Dryden, D. M., ... & Lin, J. S. (2014). Integration of existing systematic reviews into new reviews: identification of guidance needs. *Systematic reviews*, 3(1), 60
30. Atkins, D., Best, D., Briss, P. A., Eccles, M., Falck-Ytter, Y., Flottorp, S., ... & Hill, S. (2004). Grading quality of evidence and strength of recommendations. *BMJ (Clinical research ed.)*, 328(7454), 1490-1490
31. AGREE Next Steps Consortium. (2017). Appraisal of guidelines for research & evaluation II. AGREE II Instrument. The Agree Research Trust. AGREE II Update September 2017. <https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf> (dostęp w dniu 18.12.18)
32. O'Neill, J., Tabish, H., Welch, V., Petticrew, M., Pottie, K., Clarke, M., ... & Tugwell, P. (2014). Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of clinical epidemiology*, 67(1), 56-64
33. Cochrane Health Equity Field and Campbell Equity Methods Group. <http://methods.cochrane.org/equity/our-publications> (dostęp w dniu 18.12.18)
34. Equity Checklist for Systematic Review Authors – version 2012-10-04. Equity Checklist for Systematic Review Authors. <http://methods.cochrane.org/sites/methods.cochrane.org/equity/files/public/uploads/EquityChecklist2012.pdf> (dostęp w dniu 18.12.18)
35. Bossert, T. J. (1990). Can they get along without us? Sustainability of donor-supported health projects in Central America and Africa. *Social science & medicine*, 30(9), 1015-1023
36. Swerissen, H., & Crisp, B. R. (2004). The sustainability of health promotion interventions for different levels of social organization. *Health promotion international*, 19(1), 123-130
37. Shediach-Rizkallah, M. C., & Bone, L. R. (1998). Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. *Health education research*, 13(1), 87-108
38. Health Communication Unit, Centre for Health Promotion, University of Toronto (2001). Overview of Sustainability Version 8.2 April 31, 2001. <http://www.usask.ca/healthsci/che/chep/summer2007/day4/Overview%20of%20Sustainability.pdf> (dostęp w dniu 18.12.18)

39. Wang, S., Moss, J. R., & Hiller, J. E. (2005). Applicability and transferability of interventions in evidence-based public health. *Health promotion international*, 21(1), 76-83
40. Armstrong, R., Waters, E., & Jackson, N. (2007). *Systematic reviews of health promotion and public health interventions*. Melbourne: University of Melbourne
41. Cochrane Collaboration. (2017). *Cochrane handbook for systematic reviews of interventions*. Version 5.2.0. Chapter 20.3.2.5 - Approaches to integrating qualitative and quantitative evidence syntheses. <https://training.cochrane.org/handbook/pdf-versions> (dostęp w dniu 18.12.18)
42. ISSG. (2018). The InterTASC Information Specialists' Sub-Group Search Filter Resource. <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home> (dostęp w dniu 18.12.18)
43. Agency for Healthcare Research and Quality (2014). *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Chapter 9. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions, s. 193-211. https://ahrq-ehc-application.s3.amazonaws.com/media/pdf/cer-methods-guide_overview.pdf (dostęp w dniu 18.12.18)
44. Agency for Healthcare Research and Quality (2014). *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. https://ahrq-ehc-application.s3.amazonaws.com/media/pdf/cer-methods-guide_overview.pdf (dostęp w dniu 18.12.18)
45. Cochrane Collaboration. (2017). *Cochrane handbook for systematic reviews of interventions*. Version 5.2.0. Chapter 9. Analysing data and undertaking meta-analyses. <http://handbook-5-1.cochrane.org/> (dostęp w dniu 18.12.18)
46. Dubas-Jakóbczyk, K., Kocot, E., Kissimova-Skarbek, K., Huter, K., & Rothgang, H. (2016). Ekonomiczna ewaluacja programów z zakresu zdrowia publicznego-przegląd wytycznych międzynarodowych/Economic evaluation of public health programs-review of international guidelines. *Zeszyty Naukowe Ochrony Zdrowia. Zdrowie Publiczne i Zarządzanie*, 14(4), 259
47. Edwards, R. T., Charles, J. M., & Lloyd-Williams, H. (2013). Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues. *BMC public health*, 13(1), 1001
48. Nutbeam, D. (1998). Evaluating health promotion—progress, problems and solutions. *Health promotion international*, 13(1), 27-44
49. U.S. Department of Health and Human Services Centers for Disease Control and Prevention. Office of the Director, Office of Strategy and Innovation (2011). *Introduction to program evaluation for public health programs: A self-study guide*. Atlanta, GA: Centers for Disease Control and Prevention.. <https://www.cdc.gov/eval/guide/cdcevalmanual.pdf> (dostęp w dniu 19.12.18)
50. Agencja Oceny Technologii Medycznych i Taryfikacji (2016). *Praktyczne wskazówki dotyczące planowania, wdrażania oraz realizacji programów polityki zdrowotnej*. Wersja 1.1. http://www.aotm.gov.pl/www/wp-content/uploads/2016/08/Instrukcja_PPZ.pdf (dostęp w dniu 19.12.18)
51. Checklist PRESS. EUnetHTA (2017). *Guidelines: Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness*. Version 1.2, December 2017. Annexe 4. Example: Ultrasound screening for abdominal aortic aneurysms, s. 59-60.
52. McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of clinical epidemiology*, 75, 40-46. Table 9: PRESS 2015 evidence-based checklist s. 39-40
53. Halpern, S. H., & Douglas, M. (2005). Jadad scale for reporting randomized controlled trials. *Evidence-based obstetric anesthesia*, 237-238
54. Cochrane Collaboration. (2017). *Cochrane handbook for systematic reviews of interventions*. Version 5.2.0. Chapter 10: Addressing reporting biases. <http://handbook-5-1.cochrane.org/> (dostęp w dniu 19.12.18)
55. AGREE Next Steps Consortium. (2017). *Appraisal of guidelines for research & evaluation II. AGREE II Instrument*. The Agree Research Trust. AGREE II Update September 2017. <https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf> (dostęp w dniu 19.12.18)

Aneks 1. Tworzenie strategii wyszukiwania

Formułowanie pytania klinicznego

Implementacja wyszukiwania w poszczególnych bazach informacji naukowych

Przeprowadzenie wstępnego wyszukiwania (*preliminary searches*):

1. Przed opracowaniem ostatecznej strategii wyszukiwania wykonuje się wstępne wyszukiwanie celem odnalezienia wysokiej jakości przeglądów systematycznych, które dokładnie pokrywają się z zadaniem pytaniem klinicznym. W tym celu przeszukuje się *Cochrane Library*. Wstępne wyszukiwanie powinno być uproszczone.
2. Inne źródła, które wykorzystuje się we wstępnym wyszukiwaniu to: CRD *database*, strony agencji HTA (np. NICE, AHRQ), wyszukiwarka PubMed.
3. Odnalezione przeglądy systematyczne, a także istotne pozycje referencyjne, mogą być wykorzystane przy budowie właściwej strategii wyszukiwania.

Określanie struktury strategii wyszukiwania:

1. Przed opracowaniem ostatecznej strategii wyszukiwania należy zdefiniować strukturę wyszukiwania. Wymaga to jasnego sformułowania pytania klinicznego opartego o PICOS. Pytanie kliniczne jest powszechnie dzielone na poszczególne kwerendy (które mogą obejmować wskazanie, populację, interwencje, typ badań) i tylko najbardziej istotne są wykorzystywane przy budowie strategii wyszukiwania. Istotnym jest aby nie wprowadzać zbyt wielu kwerend, które mogłyby nie mieć adekwatnego odzwierciedlenia w tytułach, abstraktach i *subject headings* odnalezionych publikacji.
2. Strategia wyszukiwania może uwzględniać populację, interwencję (-e) i typ badań. Najczęściej punkty końcowe nie są uwzględniane w wyszukiwaniu systematycznym. Może być konieczne zastosowanie kilku kombinacji kwerend, które objęłyby temat przeglądu. W zakresie jednej kwerendy stosuje się operator logiczny Boolean OR łączący zarówno słowo kluczowe wpisywane tekstowo oraz z wykorzystaniem *subject headings*.
 - Przykład: badania przesiewowe USG w kierunku tętniaka aorty brzusznej (*ultrasound screening for abdominal aortic aneurysms*); budowanie struktury wyszukiwania z wykorzystaniem konceptów:
 - Koncept 1 (stanowi wskazanie): tętniak aort brzusznej (*abdominal aortic aneurysm*)
 - Koncept 2 (stanowi interwencje): badanie USG (*ultrasound screening*)
 - Koncept 3 (stanowi typ badań): randomizowane próby kliniczne (RCTs)
3. Stosowanie limitów np. język, data publikacji powinno być wykorzystywane z ostrożnością, ponieważ może to wprowadzić błąd systematyczny. Wykorzystanie limitów powinno być rozważane indywidualnie dla każdej przeszukiwanej bazy.

Wybór źródeł informacji:

1. Powinno się przeprowadzić wyszukiwanie w przynajmniej 3 bazach: MEDLINE (via Ovid), Embase (via Ovid), CENTRAL (Wiley).
2. Dodatkowo można przeszukać PubMed w celu wyszukania nieindeksowanych informacji. W zależności od tematu wykonywanego przeglądu systematycznego, dodatkowe bazy danych odpowiednie dla tego tematu mogą okazać się przydatne (dodatkowe bazy wymienione w *Chapter 6 Cochrane Handbook*).

Opracowanie strategii wyszukiwania: identyfikacja słów kluczowych (*search terms*):

1. W budowaniu strategii wyszukiwania wymagane jest korzystanie z *subject headings* oraz ze słów wpisywanych ręcznie. W literaturze opisywane są różne podejścia do identyfikacji słów kluczowych.
2. Pierwsze podejście – tradycyjne lub koncepcyjne. Źródła stosowane w tym podejściu obejmują słownik MeSH, słowniki medyczne, przeglądanie odpowiednich dla danego tematu publikacji lub konsultowanie się z ekspertami. Ma to na celu identyfikację szerokiego zakresu słów kluczowych (z wykorzystaniem *subject headings* oraz słów wpisywanych ręcznie). Dodatkowo, jedna lub kilka kluczowych publikacji jest często wybierana jako punkt wyjścia w identyfikacji dalszych istotnych słów (zastosowanie metody „*growing pearl*”). Zazwyczaj proces ten powtarza się do momentu kiedy nie znajduje się więcej istotnych publikacji/słów kluczowych.
3. Drugie podejście – przedmiotowe. W celu opracowania strategii wyszukiwania stosuje się procedury tekstowo-analityczne, za pomocą których w analizie frekwencji identyfikuje się słowa kluczowe wpisywane ręcznie oraz *subject headings*. W tym kontekście, istotne znane wcześniej publikacje lub nowo zidentyfikowane artykuły, są systematycznie analizowane. Do tego celu opracowanych zostało wiele programów różniących się między sobą

funkcjonalnością oraz kosztami np. PubReMiner, Wordstat, TerMine, EndNote, Text Mining Package of R, AntConc, Leximancer.

- Wybrane słowa kluczowe są następnie przypisywane do pojedynczych kwerend w strategii wyszukiwania, niezależnie od tego, które podejście identyfikacji słów było wybrane. Aby uniknąć powtórzeń, słowa wpisywane ręcznie, powinny być skrócone do rdzenia danego słowa (np. aort* zamiast aorta/aortic) oraz w przypadku *subject headings* powinno zastosować się funkcję „explode”, jeśli będzie to istotne.

Tabela 3. Zastosowanie przykładowych fraz i ich następstwa w stosowaniu w budowaniu strategii wyszukiwania w bazie MEDLINE via OVID

Wyodrębnione najczęściej występujące słowa kluczowe	Następstwa	Składnia słów do zastosowania w strategii w Ovid
Abdominal aortic aneurysm (s)	Słowa zazwyczaj występują we wskazanej grupie słów; dlatego też te 3 słowa połączone są operatorem „przylegania” <i>adj⁴</i>	Abdominal adj1 aortic adj1 aneurysm (wstępna koncepcja)
Abdominal aortic/ aorta aneurysm (s)	<ul style="list-style-type: none"> „aneurysm” używany jest zarówno w liczbie pojedynczej jak i mnogiej; dlatego też słowo to należy skrócić. „aorta” jest także używana w formie „aortic”; dlatego należy wyodrębnić rdzeń tego słowa. 	Abdominal adj1 aort* adj1 aneurysm* (wstępna koncepcja)
Aneurysm od the abdominal aorta	Określenie może być stosowane także w formie innej sekwencji lub z większym dystansem między słowami; dystans do słowa „aneurysm” uległ zwiększeniu.	Abdominal adj1 aort* adj3 aneurysm* (ostateczna koncepcja)

- Określenia dotyczące typu badań nie muszą być za każdym razem na nowo określane, jeśli dobrze sprawdzone i wysokiej jakości filtry są dostępne. Cochrane Handbook oraz *Health Information Research Unit of McMaster University* zapewniają wysokiej jakości filtry dla badań RCT.

Opracowanie strategii wyszukiwania: dostosowanie składni słów

- Po opracowaniu struktury strategii wyszukiwania, zdefiniowaniu słów kluczowych dla poszczególnych baz danych, można opracować ostateczną strategię wyszukiwania. Zamiast stosować wyszukiwanie krzyżowe (*cross-database search*) w kilku bazach jednocześnie, należy przeszukać każdą bazę indywidualnie. W tym celu, wcześniej zidentyfikowane słowa kluczowe wpisywane ręcznie, mogą być zastosowane w każdej bazie. *Subject headings* muszą być określone indywidualnie dla każdej bazy z osobna. Zasadnym jest aby pierwszą (najczęściej w bazie MEDLINE) opracowaną strategię wyszukiwania adaptować do pozostałych baz. Należy również zauważyć, że pewne opcje/możliwości są odmiennie w ramach różnych interfejsów poszczególnych baz danych np. inne operatory logiczne, inne funkcje skracania słów, funkcja „explode”.

Tabela 4. Tagi specyficzne dla baz danych oraz ich interfejsów dla wyszukiwania ręcznego tekstowego

Baza danych (interfejs)	Wyszukiwanie ręczne tekstowe
MEDLINE i Embase (Ovid)	(abdominal adj1 aort* adj3 aneurysm*).ti,ab.
Cochrane (Wiley)	(abdominal NEAR/1 aort* NEAR/3 aneurysm*).ti,ab
PubMed (NLM)	abdominal*[tiab] AND aort*[tiab] AND aneurysm* [tiab]

⁴ Operator ADJ „wyrzuca” wyniki z szukanymi słowami położonymi obok siebie w określonym porządku. Operator ADJ może być stosowany w połączeniu z liczbami od 1 do 99: ADJ1 „wyrzuci” wyniki z szukanymi słowami znajdującymi się obok siebie, w dowolnej kolejności; ADJ2 „wyrzuci” wyniki z szukanymi słowami znajdującymi się obok siebie z maksymalnie jednym słowem między nimi, w dowolnej kolejności; ADJ3 „wyrzuci” wyniki z szukanymi słowami znajdującymi się obok siebie z maksymalnie dwoma słowami między nimi, w dowolnej kolejności itd. Przykład: *physician adj5 relationship* wyszuka wyniki zawierająca słowa *physician* oraz *relationship* z maksymalnie czterema słowami między nimi, otrzymane frazy mogą wyglądać następująco: *physician patient relationship*, *patient physician relationship*, lub *relationship between cancer patient and physician*.

Tabela 5. Tagi specyficzne dla baz danych oraz ich interfejsów dla subject headings

Baza danych (interfejs)	Subject headings
MEDLINE (Ovid)	Aortic Aneurysm, Abdominal/
Embase (Ovid)	Abdominal Aorta Aneurysm/
Cochrane (Wiley)	MeSH descriptor: [Aortic Aneurysm, Abdominal] this term only
PubMed (NLM)	Nie stosuje się <i>subject headings</i> dla nowych, niezaindeksowanych pozycji.

- Przed „uruchomieniem” ostatecznej strategii wyszukiwania należy taką strategię poddać sprawdzeniu (*peer reviewing*). Do tego celu służy narzędzie PRESS (*Peer Review of Electronic Search Strategies*). Narzędzie to identyfikuje błędy i może zwiększyć liczbę otrzymanych istotnych publikacji. Sprawdzenie draftu strategii wyszukiwania składa się z dwóch części:
 - weryfikacja za pomocą checklisty PRESS (51, 52) przez drugą osobę,
 - sprawdzenie kompletności strategii.

Opracowanie strategii wyszukiwania: wyszukiwanie, pozyskiwanie otrzymanych wyników, zarządzanie referencjami

- Strategia wyszukiwania powinna być zapisana indywidualnie dla każdej bazy.
- Wszystkie otrzymane wyniki są zapisywane i przygotowywane do procesu ich przeglądania. Do tego celu stosuje się następujące programy np. *EndNote, RefWorks, Mendeley, RefMan*.
- Należy usunąć duplikaty powtarzające się w przeszukanych bazach danych.
- Proces selekcji publikacji optymalnie powinien być przeprowadzany przez dwie niezależne osoby (selekcja badań do przeglądu systematycznego także powinna być wykonywana przynajmniej przez dwie osoby). W pierwszej kolejności przegląda się publikacje po tytułach i abstraktach w celu oceny, czy spełniają bądź nie kryteria kwalifikacji do włączenia.

Opracowanie strategii wyszukiwania: dokumentowanie i raportowanie procesu wyszukiwania

- Proces wyszukiwania powinien być dokumentowany w czasie rzeczywistym, a nie retrospektywnie – dotyczy to zarówno opracowywania strategii jak i samego procesu przeprowadzania wyszukiwania.
- Jasne i przejrzyste raportowanie wszystkich aspektów wyszukiwania umożliwia ocenę jego jakości i kompletności oraz powtórzenie wyszukiwania w przypadku aktualizacji danego tematu. W tym zakresie wyszczególniono osiem najpowszechniejszych elementów, które obejmują: bazy danych, nazwę interfejsów, okres wyszukiwania, datę ostatniego wyszukiwania, pełną strategię wyszukiwania, ograniczenia językowe, wyszukiwania uzupełniające, kwalifikacje osób wykonujących wyszukiwanie (52).
- Wyniki wyszukiwania, usuwanie duplikatów, selekcja odnalezionych publikacji powinna być przedstawiona w postaci diagramu (*flowchart*) PRISMA.
- Wszystkie włączone oraz wykluczone publikacji powinny być przedstawione w tabeli wraz z powodem wykluczenia np. P – populacja, I – interwencja, C – komparator, O – punkty końcowe, S – typ badania, metodologia, brak pełnego tekstu.

Opracowanie strategii wyszukiwania: aktualizacja wyszukiwania

- Ze względu na fakt, że strategia wyszukiwania opracowywana jest na początku prac nad danym przeglądem systematycznym, może się okazać, że wyniki wyszukiwania ulegną dezaktualizacji przed opublikowaniem przeglądu. Z tego względu ostatnie wyszukiwanie powinno być przeprowadzone najpóźniej 12 miesięcy przed publikacją. Dlatego też, często przeprowadza się aktualizację wyszukiwania przed planowaną datą publikacji. Ustawienie auto alertów w poszczególnych bazach pomaga zidentyfikować istotne publikacje od razu po ich publikacji. Metoda ta nie zastępuje aktualizacji wyszukiwania, jednak może stanowić istotny sygnał dla przeprowadzenia ewentualnej aktualizacji.
- Aktualizacja wyszukiwania powinna identyfikować jedynie publikacje, które zostały dodane do danej bazy po dacie ostatniego wyszukiwania. W celu ograniczenia czasu wyszukiwania, powinno się brać pod uwagę datę zamieszczenia danej publikacji w bazie, aniżeli datę samej publikacji danego artykułu. Drugą metodą jest

zastosowanie operatora NOT, za którym wpisywana jest pierwotna strategia. Trzecia technika polega na pozyskaniu wszystkich wyników z aktualizacji i usuwaniu duplikatów powtarzających się w wyszukiwaniu pierwotnym.

Implementacja wyszukiwania w rejestrach badań

Ogólne zasady:

1. W celu odnalezienia badań zakończonych, ale nieopublikowanych zalecane jest przeszukanie rejestrów badań klinicznych (co najmniej 2 rejestry; wymagane jest przeszukanie rejestru *clinicaltrials.gov* oraz *ICTRP Search Portal*; dodatkowo *clinicaltrialsregister.eu*).
2. Wyszukiwanie badań w rejestrach powinno być proste, czułe i nastawione na wyszukiwanie w kierunku jednej kwerendy (interwencja lub wskazanie).
3. Weryfikacja (*peer reviewing*) opracowanej strategii wyszukiwania powinna przebiegać podobnie jak w przypadku baz informacji naukowej. Do tego celu można wykorzystać narzędzie PRESS.
4. Otrzymane wyniki z poszczególnych rejestrów zapisuje się, aby następnie przeprowadzić ich selekcję (przez dwie osoby).
5. Jeśli na początku wyszukiwania odnaleziono badania będące w trakcie (*ongoing studies*), to należy sprawdzić ich aktualny status podczas aktualizacji wyszukiwania.

Zgodnie z wytycznymi HTA Agencji, można także rozważyć potrzebę uzupełniającej identyfikacji danych naukowych następującymi sposobami (wybór źródeł informacji do identyfikacji niepublikowanych badań powinien opierać się na kompletności i wiarygodności danych):

1. konsultacje z ekspertami klinicznymi;
2. poszukiwanie nieopublikowanych dokumentów np. raportów badań;
3. wykorzystanie danych z dokumentacji rejestracyjnej leku dostępnej na stronach internetowych agencji rejestracyjnych, tj. URPL, EMA, FDA (w tym analiza komunikatów dotyczących bezpieczeństwa stosowania danej technologii);
4. kontaktowanie się z autorami badań klinicznych, m.in. w celu uzyskania i włączenia do analizy szczegółowych danych niepublikowanych;
5. niesystematyczne wyszukiwanie danych publikowanych w czasopiśmie specjalistycznych zajmujących się ocenianą technologią i nieindeksowanych w wykorzystywanych bazach informacji medycznej;
6. wykorzystanie wyszukiwarek internetowych;
7. wyszukiwanie abstraktów konferencyjnych (nie zaleca się jednak rutynowego poszukiwania abstraktów konferencyjnych; na pierwszym miejscu należy dążyć do odnalezienia pełnego raportu danego badania).

Przykładowe filtry dot. typów publikacji (oprócz filtrów automatycznych wbudowanych w wyszukiwarce)

1. Baza Medline via PubMed

Systematic Review

Lp.	Słowo kluczowe	Wynik
#5	(((review*[Title/Abstract]) OR "Review" [Publication Type])) AND systematic*[Title/Abstract]	166589
#4	(review*[Title/Abstract]) OR "Review" [Publication Type]	3160788
#3	"Review" [Publication Type] Sort by: PublicationDate	2345592
#2	review*[Title/Abstract]	1754816
#1	systematic*[Title/Abstract]	357451

Meta Analysis

Lp.	Słowo kluczowe	Wynik
#5	((("Meta-Analysis" [Publication Type] OR "Meta-Analysis as Topic"[Mesh])) OR Meta-Analysis[Title/Abstract]) OR MetaAnalysis[Title/Abstract]) OR "Meta Analysis"[Title/Abstract]	137535
#4	"Meta Analysis"[Title/Abstract]	105229
#3	MetaAnalysis[Title/Abstract]	1467
#2	Meta-Analysis[Title/Abstract]	105229
#1	"Meta-Analysis" [Publication Type] OR "Meta-Analysis as Topic"[Mesh] Sort by: PublicationDate	99511

RCT

Lp.	Słowo kluczowe	Wynik
#9	(((((("Randomized Controlled Trials as Topic"[Mesh] OR "Randomized Controlled Trial" [Publication Type])) OR "Random Allocation"[Mesh]) OR "Double-Blind Method"[Mesh])) AND (((study[Title/Abstract] OR studies[Title/Abstract]))) OR trial*[Title/Abstract]) OR experiment*[Title/Abstract])	534109
#8	(((((study[Title/Abstract] OR studies[Title/Abstract]))) OR trial*[Title/Abstract]) OR experiment*[Title/Abstract])	9646579
#7	((("Randomized Controlled Trials as Topic"[Mesh] OR "Randomized Controlled Trial" [Publication Type])) OR "Random Allocation"[Mesh]) OR "Double-Blind Method"[Mesh]	657531
#6	experiment*[Title/Abstract]	1901738
#5	trial*[Title/Abstract]	855835
#4	(study[Title/Abstract] OR studies[Title/Abstract])	8289142
#3	"Double-Blind Method"[Mesh] Sort by: PublicationDate	144019
#2	"Random Allocation"[Mesh] Sort by: PublicationDate	93233
#1	"Randomized Controlled Trials as Topic"[Mesh] OR "Randomized Controlled Trial" [Publication Type] Sort by: PublicationDate	562410

Guidelines, Recommendations

Lp.	Słowo kluczowe	Wynik
#5	((("Practice Guideline" [Publication Type]) OR "Consensus Development Conference" [Publication Type]) OR "Health Planning Guidelines"[Mesh]) OR ((guideline*[Title/Abstract] OR guidance*[Title/Abstract] OR recommendation*[Title/Abstract] OR standard*[Title/Abstract] OR consensus*[Title/Abstract]))	1625513
#4	(guideline*[Title/Abstract] OR guidance*[Title/Abstract] OR recommendation*[Title/Abstract] OR standard*[Title/Abstract] OR consensus*[Title/Abstract])	1615868

Lp.	Słowo kluczowe	Wynik
#3	"Health Planning Guidelines"[Mesh] Sort by: PublicationDate	3973
#2	"Consensus Development Conference" [Publication Type] Sort by: PublicationDate	10841
#1	"Practice Guideline" [Publication Type] Sort by: PublicationDate	23624

2. Baza Embase via Ovid

Systematic review

Lp.	Słowo kluczowe	Wynik
1	"systematic review"/	155497
2	"systematic*".ab,kw,ti.	409058
3	"review*".ab,kw,ti.	2064531
4	2 and 3	186681
5	1 or 4	240290

Meta Analysis

Lp.	Słowo kluczowe	Wynik
1	(metaanalysis or meta-analysis or meta analysis).ab,kw,ti.	140444
2	meta analysis/	137576
3	1 or 2	185169

RCT

Lp.	Słowo kluczowe	Wynik
1	randomized controlled trial/	469053
2	(random* or mask* or blind* or control*).ab,kw,ti.	4653325
3	(study or studies or trial* or experiment*).ab,kw,ti.	10756954
4	2 and 3	3150297
5	1 or 4	3245547

Guidelines, Recommendations

Lp.	Słowo kluczowe	Wynik
1	practice guideline/	325732
2	consensus development/	22764
3	good clinical practice/	8418

4	(guideline* or guidance* or recommendation* or standard* or consensus*).ab,kw,ti.	2138287
5	1 or 2 or 3 or 4	2286505

Aneks 2. Skala Jadad

Skala Jadad (53) to oksfordzki system oceny jakości badania klinicznego powstały celem ustalenia minimalnego poziomu badań włączonych do przeglądu systematycznego/ metaanalizy. Badanie może otrzymać wartości od 0 (niska jakość) do 5 punktów (najwyższa jakość).

W wyniku analizy danego badania klinicznego należy odpowiedzieć na pytania uwzględnione poniżej:

1. Czy badanie opisywane jest jako randomizowane?

Jeśli tylko wspomniano o randomizacji w badaniu otrzymuje ono 1 pkt, jeśli wskazano również metodę randomizacji i była ona właściwa przyznaje się dodatkowy punkt.

2. Czy badanie opisano jako podwójnie zaślepione?

Jeśli tylko wspomniano o badaniu z podwójnie ślełą próbą należy dodać 1 pkt, jeśli metoda zaślepienia była właściwa przyznaje się dodatkowy punkt.

3. Czy w badaniu znajduje się opis pacjentów, którzy wycofali się lub zostali wycofani z badania (opis wykluczenia)?

Jeśli opisano, należy przyznać 1 pkt.

Nie ma ocen pośrednich. Za każdy brak w zakresie ww. informacji, należy odjąć 1 pkt.

Aneks 3. Narzędzie Cochrane do oceny ryzyka *bias* oraz rodzaje *bias*

Tabela poniżej przedstawia narzędzie Cochrane (16) do oceny ryzyka błędu systematycznego, gdzie odpowiedź negatywna na sformułowane pytania odnośnie poszczególnych kryteriów sugeruje wysokie ryzyko błędu.

Tabela 6. Kryteria *Cochrane risk of bias*

Kryterium	Opis	Ocena
Określenie kolejności przydzielania chorych do grup – kodu randomizacyjnego	Opisać możliwie najdokładniej metodę wykorzystaną do opracowania sekwencji przydziału chorych do grup tak, aby umożliwić ocenę jej skuteczności, to jest wygenerowania wyjściowo porównywalnych grup.	Czy prawidłowo określono sposób przydzielania chorych do grup?
Ukrycie kodu alokacji	Opisać możliwie najdokładniej metodę wykorzystaną do utajnienia informacji o przydziale chorych do grup tak, aby umożliwić ocenę możliwości przewidzenia przydziału chorych do grup przed lub w trakcie alokacji.	Czy kolejność przydzielania do grup była prawidłowo utajniona?
Metoda zaślepienia w zakresie każdego ważnego efektu terapeutycznego	Opisać wszystkie metody wykorzystane do zaślepienia uczestników badania i badaczy, co do badanego i stosowanego u chorych leczenia.	Czy metoda zaślepienia była efektywna? Czy informacja o przydziale do grup określonych różnym stosowanym leczeniem była tajna w trakcie badania?
Analiza danych niepełnych w zakresie każdego ważnego efektu terapeutycznego	Opisać w jakim stopniu dane o efektach terapeutycznych były kompletne oraz jak przedstawiały się wyłączenia chorych z badania z poszczególnych przyczyn.	Czy publikacje z badania nie są oparte na selektywnym doborze chorych i tym samym danych o efektach terapeutycznych?
Selektywne raportowanie wyników o skuteczności klinicznej	Opisać możliwość selektywnego raportowania wyników.	Czy publikacje z badania nie są oparte na selektywnym raportowaniu wyników o efektach zdrowotnych?
Inne źródła błędów	Opisać inne aspekty badania mogące być źródłem błędów.	Czy badanie miało inne uchybienia metodologiczne mogące zwiększać ryzyko błędu?

Źródło: opracowanie własne na podstawie Cochrane Handbook

Wśród przykładowych rodzajów błędu systematycznego (54) wymienia się:

- błąd systematyczny związany z wybiórczym publikowaniem (*publication bias*) - wynika z większej szansy opublikowania wyników badania, w którym uzyskano istotny efekt, niż badań z wynikami nierozstrzygniętymi;
- błąd systematyczny związany z wielokrotnym publikowaniem wyników tego samego badania (*duplicate/multiple publication bias*) - wynika z większej szansy na wielokrotną publikację wyników badań, w których wykazano efekt istotny statystycznie, niż badań z wynikami nierozstrzygniętymi (może być powodem włączenia tych samych wyników więcej niż raz do metaanalizy);
- błąd systematyczny opóźnienia publikacji wyników badań (*time-lag bias*) - możliwość błędnej interpretacji wyników w wyniku szybszej lub opóźnionej publikacji badania;
- błąd systematyczny związany z wybiórczym cytowaniem badań (*citation bias*) - wynika z większej szansy na cytowanie niektórych badań, np. opublikowanych przez znanego autora lub w ogólnodostępnym czasopiśmie (łatwiej dostępnym);
- błąd systematyczny związany z dobieraniem badań opublikowanych tylko w niektórych językach (*language bias*) - wynika z włączenia do przeglądu systematycznego badań opublikowanych najczęściej po angielsku (może spowodować pominięcie badań, w których nie wykazano efektu lub wykazany efekt był mały);

- błąd systematyczny związany z dobieraniem publikacji wyników badań w czasopismach o różnej dostępności (np. uwzględnianie głównie łatwo dostępnych, tzw. szara literatura) lub różnych poziomach indeksowania w bazach danych (*location bias*);
- błąd systematyczny związany z wybiórczym raportowaniem wyników wpływający na ich ostateczną interpretację i kierunek (*outcome reporting bias*) – wynika z większej szansy pełnego przedstawienia wyników statystycznie istotnych niż nie spełniających istotności statystycznej.

Aneks 4. Skala AMSTAR

Skala AMSTAR jest istotnym i łatwym do zastosowania narzędziem służącym do oceny przeglądów systematycznych. Narzędzie to zostało przedstawione poniżej i zgodnie z zaleceniami autorów powinno być aktualizowane w miarę postępu wiedzy nt. czynników wpływających na jakość przeglądów systematycznych.

Lp.	Pytanie wraz z uzasadnieniem	Odpowiedź
1.	<p><u>Czy badanie zaprojektowano <i>a priori</i>?</u></p> <p>Badane zagadnienie oraz kryteria włączenia powinny zostać ustalone przed przystąpieniem do wykonywania przeglądu.</p> <p><i>Odpowiedź twierdzącą należy zaznaczyć, jeżeli w ocenianym przeglądzie znajdują się odniesienia do protokołu, akceptacji etycznej badania lub informacje nt. wstępnie ustalonych lub wcześniej opublikowanych celów badawczych.</i></p>	
2.	<p><u>Czy selekcja badań oraz ekstrakcja danych zostały powielone?</u></p> <p>Powinno być co najmniej dwóch niezależnych badaczy, a w przypadku niezgodności pomiędzy badaczami powinien być stosowany proces dochodzenia do konsensusu.</p> <p><i>Dwie osoby wybierają badania, dwie wykonują ekstrakcję danych lub jedna osoba kontroluje pracę drugiej osoby. Warto zwrócić uwagę na liczbę autorów przeglądu, być może kilka osób uczestniczyło w procesie selekcji badań.</i></p>	
3.	<p><u>Czy wykonano kompleksowy przegląd dostępnej literatury?</u></p> <p>Przeszukane powinny być co najmniej dwa źródła elektroniczne. Autorzy powinni podać przedział czasu objęty wyszukiwaniem oraz podać dokładnie, jakie bazy zostały przeszukane.</p> <p>Powinni podać też słowa kluczowe/terminy MeSH a tam gdzie to możliwe powinna być przedstawiona strategia wyszukiwania dowodów naukowych. Wyszukiwanie powinno zostać uzupełnione o przegląd referencji odnalezionych doniesień, informacje przedstawione w opublikowanych przeglądach, podręcznikach, wyspecjalizowanych rejestrach, bądź opinie ekspertów z danej dziedziny naukowej.</p> <p><i>Jeśli zostały przeszukane co najmniej dwa źródła danych oraz zastosowano jedną strategię uzupełniającą należy zaznaczyć odpowiedź twierdzącą (Cochrane register/Central liczą się jako dwa źródła, natomiast grey literature zaliczana jest do uzupełniającej strategii).</i></p>	
4.	<p><u>Czy status publikacji (tzw. <i>grey literature</i>, <i>GL</i>) był stosowny jako kryterium włączenia?</u></p> <p>Autorzy powinni podkreślić, że wyszukiwali badań niezależnie od rodzaju ich publikacji. Autorzy powinni również podać, czy wykluczali jakiegokolwiek doniesienia (z przeglądu systematycznego) w oparciu o status publikacji, język itp.</p> <p><i>Jeśli w otmn przeglądzie jest zaznaczone, że poszukiwano także GL lub literatury nieopublikowanej należy zaznaczyć odpowiedź twierdzącą. Baza danych Europejskiego Systemu Informacji o Szarej Literaturze (SIGLE), dysertacje, materiały konferencyjne czy rejestry badań klinicznych uznawane są za GL. W przypadku wyszukiwania prowadzonego w źródle zawierającym zarówno GL, jak i wiarygodną literaturę, autorzy powinni sprecyzować, że poszukiwali GL lub nieopublikowanej literatury.</i></p>	
5.	<p><u>Czy przedstawiono wykaz badań włączonych oraz wykluczonych?</u></p> <p><i>W przypadku badań wykluczonych z przeglądu, mogą one zostać przedstawione w postaci referencji. W przypadku automatycznych odnośników, jeśli link się nie otwiera, należy zaznaczyć odpowiedź „nie”.</i></p>	
6.	<p><u>Czy przedstawiono charakterystykę badań włączonych do przeglądu?</u></p> <p>Dane pochodzące z oryginalnych badań dot. uczestników, interwencji oraz wyników zdrowotnych powinny zostać przedstawione w formie zagregowanej (np. w tabeli – przykładowe dane: wiek, płeć, stan i czas trwania choroby, informacje odnośnie innych schorzeń).</p> <p><i>Charakterystyka badań nie musi być koniecznie przedstawiana w tabeli, o ile została opisana jak wyżej.</i></p>	

7.	<p><u>Czy dokonano oceny jakości badań włączonych do przeglądu?</u></p> <p><i>Można uwzględnić narzędzia oceny jakości lub listę kontrolną (tj. checklist), np. skala oceny Jadad, ocena ryzyka błędu, analiza wrażliwości itp. lub opis elementów jakościowych, z pewnego rodzaju wynikiem dla każdego badania.</i></p>	
8.	<p><u>Czy jakość włączonych badań do przeglądu została odpowiednio wykorzystana przy formułowaniu wniosków?</u></p> <p><i>W przypadku, gdy zaznaczono odpowiedź negatywną dla pytania nr 7, a autor przeglądu wskazuje, że „wyniki należy interpretować z ostrożnością ze względu na słabą jakość badań” nie można na postawione pytanie odpowiedzieć twierdząco.</i></p>	
9.	<p><u>Czy użyto właściwych metod w celu wykonania metaanalizy wyników?</u></p> <p>Celem sprawdzenia homogeniczności analizowanych wyników autorzy przeglądu powinni wykonać odpowiednie testy (np. jednorodności Chi²). W przypadku heterogeniczności, powinni zastosować model losowy (np. ANOVA).</p> <p><i>Odpowiedź twierdzącą należy zaznaczyć, jeśli autorzy wspomnieli o heterogeniczności lub opisali heterogeniczność badań włączonych do przeglądu, tj. jeżeli wyjaśnili, że nie mogą wykonać metaanalizy ze względu na heterogeniczność badań.</i></p>	
10.	<p><u>Czy oceniano prawdopodobieństwo tendencyjności publikacyjnej (ang. <i>publication bias</i>)?</u></p> <p>Ocena błędu publikacji powinna zawierać połączenie pomocy graficznych (np. wykres lejkowy, inne dostępne testy) lub testów statystycznych (np. test regresji Eggera, test Hedgesa-Olkena).</p> <p><i>W przypadku braku wartości testowych lub wykresu lejkowego należy zaznaczyć odpowiedź negatywną. Wynik twierdzący można zaznaczyć, jeżeli w przeglądzie wspomniano, że nie można było ocenić tendencyjności publikacyjnej, ponieważ do analizy włączono mniej niż 10 badań.</i></p>	
11.	<p><u>Czy uwzględniono konflikt interesów?</u></p> <p>Potencjalne źródła wsparcia powinny zostać wyraźnie wskazane zarówno w przeglądzie systematycznym, jak i włączonych badaniach.</p> <p><i>Należy wykazać źródła finansowania lub wsparcia dla przeglądu systematycznego oraz dla każdego z badań włączonych do przeglądu, aby udzielić odpowiedzi twierdzącej na dane pytanie.</i></p>	

Źródło: opracowanie własne na podstawie <https://amstar.ca/>

Interpretacja narzędzia AMSTAR:

W każdym z pytań możliwe odpowiedzi to: TAK, NIE, NIE MOŻNA ODPOWIEDZIEĆ (jednak tą odpowiedź należy wybrać, jeśli oceniana pozycja jest istotna, ale nie została opisana przez autorów), NIE DOTYCZY (należy wybrać, gdy oceniana pozycja nie jest istotna, np. metaanaliza nie była możliwa do wykonania, lub autorzy nie podjęli próby wykonania metaanalizy).

Wynik poniżej 5 odpowiedzi „TAK” sugeruje, że jest to przegląd o niskiej jakości. Wynik ≥ 5 sugeruje, że jest o przegląd o umiarkowanej jakości. Wynik ≥ 9 to przegląd o wysokiej jakości.

Aneks 5. Narzędzie ROBIS

Tabela 7. Narzędzie ROBIS

Etap 1 – Ocena związku z tematem⁵ (opcjonalnie)		
Czy pytania uwzględnione w przeglądach są dopasowane/powiązane z pytaniem docelowym podsumowania/analizy/wytycznych? TAK/ NIE/ NIEJASNE		
Kategoria	Pytanie docelowe podsumowania/przeglądu/wytycznych	Pytania w zakresie przeglądów będących przedmiotem oceny
Przeglądy dot. interwencji		
Populacja/-e (P)		
Interwencja/-e (I)		
Komparator/-y (C)		
Punkt/-y końcowy/-e (O)		
Przeglądy dot. etiologii		
Pacjenci/populacja		
Ekspozycja i komparatory		
Punkty końcowe		
Przeglądy dot. dokładności diagnostycznej testu (DTA)		
Pacjenci		
Badany wskaźnik/wskaźniki		
Norma referencyjna (wartości referencyjne)		
Stan docelowy		
Przeglądy dot. prognozy w zakresie analizowanego problemu zdrowotnego		
Pacjenci		
Przewidywane punkty końcowe		
Przeznaczenie modelu (zamierzone wykorzystanie modelu)		
Zaplanowany/przewidywany moment w czasie (okres czasu)		
Etap 2 - Identyfikacja problemów związanych z procesem weryfikacji przeglądów		

⁵ Ocena ryzyka błędu w przeglądach z pytaniami badawczymi dot. interwencji, etiologii, dokładności diagnostycznej testu, prognozowania w zakresie analizowanego problemu zdrowotnego. Należy określić pytanie docelowe podsumowania/przeglądu/wytycznych i pytania badawcze w przeglądach będących przedmiotem oceny.

Domena 1. Kryteria kwalifikacji badań					
<i>Proszę opisać kryteria kwalifikowalności badania, wszelkie ograniczenia w ww. zakresie oraz czy istnieją przesłanki, że cele i kryteria kwalifikowalności zostały z góry określone na podstawie pytań zawartych w tabeli. Należy zaznaczyć odpowiednią odpowiedź.</i>					
	TAK	RACZEJ TAK	BRAK INFORMACJI	RACZEJ NIE	NIE
Czy przegląd był zgodny z wcześniej zdefiniowanymi celami i kryteriami kwalifikowalności?					
Czy kryteria kwalifikowalności były adekwatne do postawionego pytania badawczego w przeglądzie?					
Czy kryteria kwalifikowalności zostały jednoznacznie określone?					
Czy wszystkie ograniczenia dot. kryteriów kwalifikowalności badania były właściwe (np. okres realizacji, data badania, rozmiar próby, jakość badania, pomiar wyników etc.)?					
Czy ww. ograniczenia w zakresie kryteriów kwalifikowalności były oparte o rzetelne źródła informacji (np. została wskazana data dostępu, stan publikacji lub jej format, język etc.)?					
Obawy dot. określenia kryteriów kwalifikowalności do badań	MAŁE / DUŻE / NIEJASNE				
Uzasadnienie ww. obaw					
Domena 2. Identyfikacja i selekcja badań					
<i>Należy opisać metodologię identyfikacji i selekcji badań (np. liczba zaangażowanych recenzentów itp.).</i>					
	TAK	RACZEJ TAK	BRAK INFORMACJI	RACZEJ NIE	NIE
Czy wyszukiwanie badań zostało przeprowadzone w odpowiednim/niezbędnym zakresie baz danych/ elektronicznych źródeł informacji dot. opublikowanych i nieopublikowanych raportów?					
Czy w metodologii uwzględniono dodatkowe wyszukiwanie w innych bazach danych celem wyszukania stosownych raportów/badań?					
Czy zastosowane terminy/słowa kluczowe oraz struktura strategii wyszukiwania pozwoliły na uzyskanie możliwie największej liczby badań na założony temat?					
Czy zawężenie/ograniczenie wyszukiwania było zależne od daty, rodzaju publikacji lub języka?					
Czy podjęto starania celem zminimalizowania ryzyka błędu selekcji badań?					
Obawy dot. metodologii w zakresie identyfikacji i selekcji badań	MAŁE / DUŻE / NIEJASNE				
Uzasadnienie ww. obaw					
Domena 3. Gromadzenie/ekstrakcja danych i ich ocena					
<i>Należy opisać metodę gromadzenia danych, jakie dane zostały wyciągnięte z badań lub zebrane za pomocą innych narzędzi/działań, w jaki sposób oszacowano ryzyko błędu (np. liczba zaangażowanych recenzentów) oraz narzędzia stosowane do oceny ryzyka błędu stronności</i>					

	TAK	RACZEJ TAK	BRAK INFORMACJI	RACZEJ NIE	NIE
Czy podjęto starania celem zminimalizowania błędów w zakresie gromadzenia danych?					
Czy udostępniono wystarczającą charakterystykę badań celem umożliwienia interpretacji wyników czytelnikom/autorom przeglądów?					
Czy zostały zebrane wszystkie potrzebne wyniki badań celem wykorzystania w syntezie?					
Czy ryzyko błędu systematycznego (lub jakości metodologicznej) zostało ocenione przy użyciu odpowiednich kryteriów?					
Czy podjęto wysiłki w celu zminimalizowania błędów w ocenie ryzyka wystąpienia błędu systematycznego?					
Obawy dot. gromadzenia/ekstrakcji danych i ich oceny	MAŁE / DUŻE / NIEJASNE				
Uzasadnienie ww. obaw					
Domena 4. Synteza badań i wyniki					
	TAK	RACZEJ TAK	BRAK INFORMACJI	RACZEJ NIE	NIE
Czy synteza dotyczy/zawiera wszystkie możliwe badania, które powinna uwzględnić?					
Czy przedstawiono metodę włączenia badań i powód ich ewentualnego wykluczenia?					
Czy określono heterogeniczność włączonych badań celem syntezy ilościowej?					
Czy obciążenie/błędy były minimalizowane					
Czy zmienność między badaniami (heterogeniczność) była minimalna lub zaadresowana w syntezie					
Czy obciążenie/błędy w badaniach pierwotnych były minimalizowane lub zostały wskazane w danej syntezie?					
Obawy dot. syntezy badań i wyników	MAŁE / DUŻE / NIEJASNE				
Uzasadnienie ww. obaw					
Etap 3 - Ocenianie ryzyka błędu systematycznego					
Należy podsumować obawy uwzględnione w etapie 2, a następnie ocenić ryzyko błędu systematycznego.					
Ryzyko błędu systematycznego w przeglądzie					
Opisz pod względem zgromadzonych dowodów.					
	TAK	RACZEJ TAK	BRAK INFORMACJI	RACZEJ NIE	NIE
Czy znaleziono wytłumaczenie dla problemów/obaw w zakresie Domen 1-4?					

Czy odpowiednio uwzględniono związek z tematem (trafność) włączonych badań do postawionego pytania klinicznego?					
Czy recenzenci/badacze unikali uwydatniania wyników opartych na ich istotności statystycznej?					
Ryzyko błędu systematycznego w danym przeglądzie	MAŁE / DUŻE / NIEJASNE				

Źródło: opracowanie własne na podstawie University of Bristol – ROBIS Tool (<https://www.bristol.ac.uk/population-health-sciences/projects/robis/>)

Aneks 6. Skala GRADE

Nie do wszystkich etapów procesu tworzenia wytycznych klinicznych odnosi się bezpośrednio system GRADE (30). Rola GRADE polega na zdefiniowaniu pytań, wybraniu punktów końcowych i przypisaniu im wagi, ocenie jakości danych i zinterpretowaniu tych danych, głównie w kontekście oszacowanych efektów względnych i bezwzględnych.

Ogólny stopień wiarygodności danych (siła zaleceń):

- A. dane pochodzące z licznych badań z randomizacją lub z metaanaliz
- B. dane pochodzące z jednego badania z randomizacją lub z dużych badań bez randomizacji
- C. uzgodniona opinia ekspertów lub dane pochodzące z małych badań, lub z badań retrospektywnych bądź rejestrów

W tabeli poniżej przedstawiono system klasyfikacji dowodów i zaleceń wg GRADE.

Tabela 8. System klasyfikacji dowodów i zaleceń wg GRADE

System klasyfikacji dowodów i zaleceń
<p>Stopień 1A - Silna rekomendacja (stosowana u większości pacjentów w większości przypadków bez zastrzeżeń), dowód o wysokiej jakości:</p> <ul style="list-style-type: none">• korzyści wyraźnie przewyższają ryzyko i obciążenia,• dowód o wysokiej jakości pochodzący z badania typu RCT bez istotnych ograniczeń lub silny dowód oparty na wynikach badań obserwacyjnych.
<p>Stopień 1B - Silna rekomendacja (może być zastosowana u większości pacjentów w większości przypadków bez zastrzeżeń), dowód o średniej jakości:</p> <ul style="list-style-type: none">• korzyści wyraźnie przewyższają ryzyko i obciążenia,• dowód o średniej jakości pochodzący z badania typu RCT z istotnymi ograniczeniami (niezgodność wyników, błędy metodologiczne, pośrednie lub niesprecyzowane), bądź wyjątkowo silny dowód oparty na wynikach badań obserwacyjnych.
<p>Stopień 1C - Silna rekomendacja (ale może ulec zmianie, gdy będzie dostępny dowód o wysokiej jakości), dowód o niskiej lub bardzo niskiej jakości:</p> <ul style="list-style-type: none">• korzyści wyraźnie przewyższają ryzyko i obciążenia,• dowód o niskiej lub bardzo niskiej jakości pochodzący z badań obserwacyjnych lub opisu serii przypadków.
<p>Stopień 2A - Słaba rekomendacja (najlepsza metoda działania może być różna w zależności od okoliczności lub wartości uznawanych przez pacjenta bądź społeczeństwo), dowód o wysokiej jakości:</p> <ul style="list-style-type: none">• korzyści równoważą ryzyko i obciążenia,• badania typu RCT bez poważnych ograniczeń lub silny dowód pochodzący z badań obserwacyjnych.
<p>Stopień 2B - Słaba rekomendacja (najlepsza metoda działania może być różna w zależności od okoliczności lub wartości uznawanych przez pacjenta lub społeczeństwo), dowód o średniej jakości:</p> <ul style="list-style-type: none">• korzyści równoważą ryzyko i obciążenia,• dowód pochodzący z badań typu RCT z poważnymi ograniczeniami (niezgodność wyników, błędy metodologiczne, pośrednie lub niesprecyzowane) lub bardzo silny dowód oparty na wynikach badań obserwacyjnych.
<p>Stopień 2C - Bardzo słaba rekomendacja (zastosowanie innych metod alternatywnych może być równie uzasadnione), dowód o niskiej lub bardzo niskiej jakości:</p> <ul style="list-style-type: none">• niepewność w oszacowaniu korzyści, ryzyka i obciążeń lub korzyści, ryzyko i obciążenia mogą się ściśle równoważyć• dowód o niskiej lub bardzo niskiej średniej jakości pochodzący z badań obserwacyjnych lub opisu serii przypadków.

Źródło: opracowanie własne na podstawie Atkins 2004. Grading quality of evidence and strength of recommendations

Aneks 7. Instrument AGREE

Narzędzie AGREE składa się z 23 pozycji zgrupowanych w sześciu domenach (55). Każda pozycja jest oceniana według 7-stopniowej skali Likerta, gdzie 1 oznacza najmniejszą jakość, a 7 dotyczy najwyższej jakości zaleceń. Wyniki dla każdej z domen są sumą punktów, jakie uzyskały poszczególne ich pozycje. Wyniki każdej z domen są niezależne i nie mogą być sumowane w jeden wynik całościowy.

W tabeli poniżej przedstawiono kryteria jakości wytycznych/rekomendacji wg AGREE.

Tabela 9. Kryteria jakości wytycznych/rekomendacji wg AGREE

Domena	Uzasadnienie
Zakres i cel	<ul style="list-style-type: none"> Ogólne cele, pytania kliniczne i populacja pacjentów, do której odnoszą się wytyczne, są szczegółowo opisane. <ol style="list-style-type: none"> Ogólny cel (lub cele) wytycznych jest (są) precyzyjnie opisany (opisane). Pytanie (pytania) kliniczne objęte wytycznymi jest (są) precyzyjnie opisane. Populacja (pacjenci, społeczeństwo itd.), do której wytyczne się odnoszą, jest precyzyjnie opisana.
Udział osób zainteresowanych	<ul style="list-style-type: none"> Zespół opracowujący wytyczne obejmuje przedstawicieli wszystkich grup zawodowych, określa jasno docelowych użytkowników, umożliwia wypróbowanie wytycznych przez przyszłych użytkowników i uwzględnia punkt widzenia i preferencje pacjentów. <ol style="list-style-type: none"> W skład zespołu opracowującego wytyczne wchodzi przedstawiciele wszystkich grup zawodowych, których one dotyczą. Uwzględniono punkt widzenia i preferencje populacji (pacjentów, społeczeństwa itd.). Użytkownicy wytycznych są jasno określani.
Rzetelność opracowywania wytycznych	<ul style="list-style-type: none"> Metody systematyczne stosujące jasne i jawne kryteria selekcji oraz sposób formułowania zaleceń zostały jasno opisane. Istnieje wyraźny związek między wykorzystanymi dowodami a wnioskami i zaleceniami. Przy opracowywaniu zaleceń wzięto pod uwagę korzyści zdrowotne, działania niepożądane i zagrożenia. Zalecenia zostały wyraźnie powiązane z dowodami, na których je oparto. Istnieje jasna metoda uaktualniania wytycznych. <ol style="list-style-type: none"> Wyszukiwanie danych przeprowadzono w sposób systematyczny. Kryteria doboru danych są jasno opisane. Zalety i ograniczenia wszystkich danych zostały przejrzysto opisane. Metody formułowania zaleceń są jasno opisane. Przy formułowaniu zaleceń uwzględniono korzyści zdrowotne, skutki uboczne i zagrożenia. Zalecenia zostały wyraźnie powiązane z danymi, na podstawie których je sformułowano. Przed publikacją wytyczne zostały zrecenzowane przez zewnętrznych ekspertów. Podano procedurę aktualizacji wytycznych.
Przejrzystość i sposób przedstawienia	<ul style="list-style-type: none"> Zalecenia są szczegółowe i jednoznaczne, przedstawiają też różne opcje leczenia. Udostępnione są narzędzia ułatwiające korzystanie z wytycznych. <ol style="list-style-type: none"> Zalecenia są konkretne i jednoznaczne. Różne możliwości postępowania w danej sytuacji są jasno przedstawione. Główne zalecenia są łatwe do zidentyfikowania.
Użyteczność	<ul style="list-style-type: none"> Omówiono potencjalne bariery zastosowania zaleceń, w tym kwestie finansowe. W wytycznych proponuje się kroki zmierzające do poprawy jakości, powiązane z wnioskami. <ol style="list-style-type: none"> Omówiono potencjalne bariery i ułatwienia w stosowaniu zaleceń. Wytyczne są uzupełnione poradami i/lub narzędziami ułatwiającymi ich stosowanie. Uwzględniono potencjalne skutki ekonomiczne postępowania zgodnego z zaleceniami.

	21. Wytyczne zawierają kryteria służące do monitorowania lub kontroli jakości.
Niezależność redakcyjna	<ul style="list-style-type: none">• Pod względem redakcyjnym wytyczne są niezależne od instytucji finansujących, a każdy potencjalny konflikt interesów został ujawniony. <p>22. Poglądy instytucji finansujących proces powstawania wytycznych nie miały wpływu na ich treść.</p> <p>23. Potencjalne konflikty interesów autorów wytycznych zostały ujawnione i podjęto w związku z nimi odpowiednie działania.</p>

Źródło: opracowanie własne na podstawie AGREE Next Steps Consortium (2017)

Aneks 8. Narzędzie Health Evidence™ – Quality Assessment Tool – Review Articles (dot. badań wtórnych)

Tabela 10. Narzędzie Health Evidence™ – Quality Assessment Tool – Review Articles

Kryteria	Tak*	Nie
1. Czy autorzy jednoznacznie sformułowali/ukierunkowali pytanie kliniczne [populacja, interwencja (strategia), punkty końcowe]?		
2. Czy zastosowano odpowiednie kryteria włączenia celem wyboru badań pierwotnych?		
3. Czy opisano przez autorów strategię wyszukiwania w sposób wyczerpujący? Zakreśl źródła użyte do skonstruowania ww. strategii: <ul style="list-style-type: none"> ○ Bazy danych dot.: <ul style="list-style-type: none"> ▪ zdrowia, ▪ psychologii, ▪ socjologii, ▪ ukierunkowane na edukację, ▪ inne. ○ Wyszukiwanie ręczne ○ Kluczowi informatorzy ○ Wykaz referencji bibliograficznych ○ Dowody nieopublikowane 		
4. Czy strategia wyszukiwania obejmuje odpowiednią liczbę lat (adekwatny czas wyszukiwania dowodów naukowych)?		
5. Czy autorzy opisują poziom dowodów naukowych w badaniach pierwotnych włączonych do przeglądu? Poziom 1 – tylko badania RCT Poziom 2 – badania nierandomizowane, kohortowe, kliniczno-kontrolne Poziom 3 – badania niekontrolowane		
6. Czy przegląd oceniał metodologię jakości badań pierwotnych, włączając: (co najmniej 4 z niżej wymienionych) <ul style="list-style-type: none"> ○ Rodzaj/projekt badania, ○ Próba badawcza, ○ Stopień uczestnictwa, ○ Źródła błędu systematycznego (czynniki zakłócające etc.) ○ Zebranie danych (miary zmiennych zależnych/niezależnych), ○ Wskaźniki rezygnacji lub follow-up ○ Analizy danych. 		
7. Czy wyniki przeglądu są przejrzyste?		
8. Czy właściwie połączono wyniki uzyskane w ramach włączonych badań?		
9. Czy zastosowano odpowiednie metody celem połączenia lub porównania wyników z poszczególnych badań?		
10. Czy uzyskane dane potwierdzają interpretację autorów?		

Źródło: opracowanie własne na podstawie <https://www.healthevidence.org/documents/our-appraisal-tools/quality-assessment-tool-dictionary-en.pdf>

* Interpretacja: dowody wysokiej jakości = od 8 do 10 zaznaczonych odpowiedzi TAK; dowody średniej jakości = od 5 do 7 zaznaczonych odpowiedzi TAK; dowody słabej jakości = wyniki poniżej 4 "TAK"

Aneks 9. Przykładowa interwencja zależna od kontekstu

Interwencja medialna promująca spożycie warzyw i owoców

Zależna od następujących czynników (kontekstu):



Dostępność i relatywna cena owoców i warzyw

Zależna od następujących czynników (kontekstu):



Czynniki geograficzne, systemy dystrybucji żywności i ceny detaliczne

Źródło: opracowanie własne

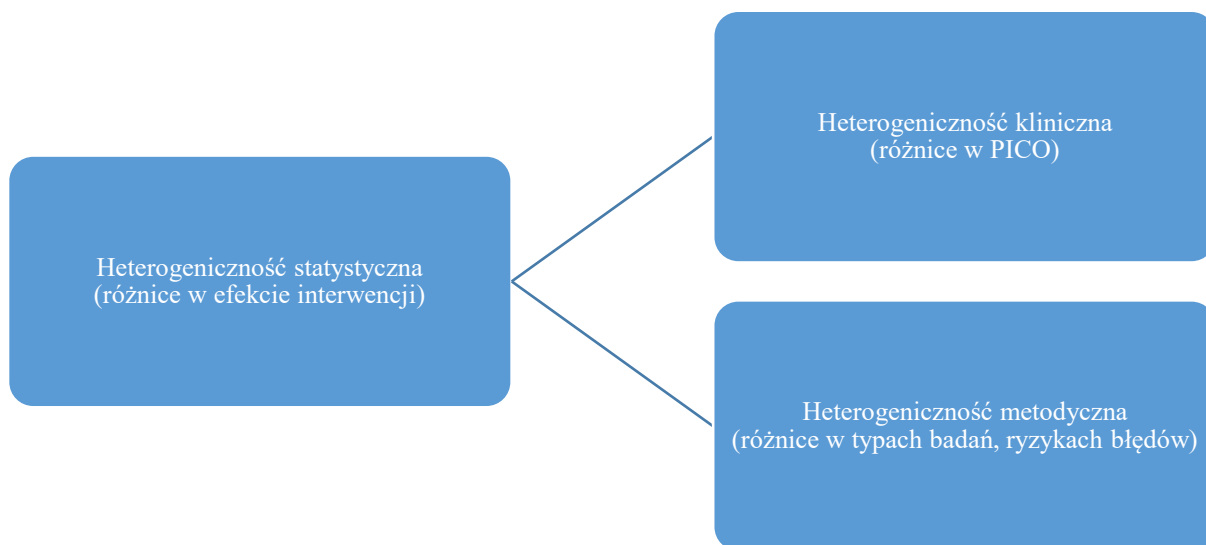
Aneks 10. Różnice pomiędzy *applicability* oraz *transferability* w zakresie stosowalności

Tabela 11. Przedstawienie różnic pomiędzy *applicability* oraz *transferability* w zakresie stosowalności

Aspekt	<i>Applicability</i>	<i>Transferability</i>
znaczenie	Czy wdrażanie (proces) interwencji może być zaimplementowany w warunkach lokalnych, bez względu na punkty końcowe.	Jeśli interwencja miałaby zostać wdrożona w warunkach lokalnych, to czy skuteczność programu byłaby podobna do poziomu uzyskanego w warunkach badania?
synonim	Wykonalność (<i>feasibility</i>)	Generalizacja (<i>generalizability</i>)
pytanie zasadnicze	Czy jest możliwe aby wdrożyć daną interwencję w warunkach lokalnych?	Jeśli interwencja ma być prowadzona w danych warunkach lokalnych, to czy może osiągnąć taką samą skuteczność jak w warunkach badania?
element poddawany ocenie	Proces interwencji	Punkt końcowy interwencji
przykład	Program cytologicznych badań przesiewowych (test Pap) lub promowania karmienia piersią w miejscach publicznych, w środowiskach z konserwatywnymi normami społecznymi dot. seksualności i płodności, może być niemożliwy do przeprowadzenia.	Jeżeli osoby prowadzące interwencję nie mają doświadczenia i posiadają niewielkie umiejętności w zakresie prowadzenia tych interwencji, to jej skuteczność w warunkach lokalnych może być niższa niż ta w warunkach badania.

Źródło: opracowanie własne

Aneks 11. Rodzaje heterogeniczności badań w metaanalizie



Źródło: opracowanie własne na podstawie Cochrane Handbook (Chapter 9.5.1 What is heterogeneity?)

Aneks 12. Potencjalna hierarchia efektów danego programu

Uczestnictwo (<i>participation</i>)	<ul style="list-style-type: none">• wielkość populacji docelowej;• charakterystyka populacji docelowej;• częstość i intensywność kontaktu/działań informacyjnych
Oddziaływanie (<i>reactions</i>)	<ul style="list-style-type: none">• stopień zainteresowania;• odczucia w stosunku do programu;• akceptacja działań programu i metod edukacyjnych
Nauka (<i>learning</i>)	<ul style="list-style-type: none">• wiedza, opinie, umiejętności, ambicje/aspiracje jako końcowe rezultaty
Działania (<i>actions</i>)	<ul style="list-style-type: none">• wzorce zachowań przejęte przez odbiorców
Zmiany systemowe i środowiskowe	<ul style="list-style-type: none">• zmiany warunków społecznych, ekonomicznych lub środowiskowych jako rezultaty rekomendacji, działań, polityki i zaimplementowanych działań
Efekty zdrowotne (<i>health outcomes</i>)	<ul style="list-style-type: none">• wskaźniki zdrowotne jako wyniki końcowe

Źródło: opracowanie własne na podstawie dokumentu CDC (2011). Introduction to program evaluation for public health programs: A self-study guide (49)

Aneks 13. Główne kryteria ewaluacji programu

Tabela 12. Przedstawienie głównych kryteriów ewaluacji PPZ

Kryterium	Pytania wyjaśniające
Użyteczność	Kto zgłasza zapotrzebowanie na wyniki ewaluacji/ kto skorzysta z wyników wspomnianej oceny? Czy proces ewaluacji udzieli stosownych informacji w odpowiednim czasie?
Wykonalność	Czy zaplanowane etapy ewaluacji są realne do wdrożenia, mając na uwadze czas, zasoby i dostępną wiedzę? Jaki jest stosunek nakładów (poświęconego czasu/wysiłku) do uzyskanych efektów?
Stosowność	Czy ewaluacja chroni prawa osób i dobro osób zaangażowanych w dany proces? Czy ewaluacja programu ma wpływ bezpośredni na populację docelową i zmiany z nią związane (tj. potencjalnych uczestników i ich otoczenie)? Czy i w jakim stopniu cele programu odpowiadają na zdiagnozowane wcześniej potrzeby?
Trafność	Czy ewaluacja programu generuje prawidłowe i wiarygodne wyniki, mając na uwadze potrzeby osób z nich korzystających?

Źródło: opracowanie własne na podstawie dokumentu CDC (2011). Introduction to program evaluation for public health programs: A self-study guide (49)

Aneks 14. Przykład wykorzystania matrycy logicznej w ewaluacji programu

W przypadku sugerowanego przez CDC programu z zakresu zdrowia publicznego dotyczącego prewencji zatrucia ołowiem w populacji dzieci, matryca logiczna przedstawiałaby następujące zagadnienia:

- wkład/dane wejściowe: środki finansowe, wyszkolony personel w zakresie badań przesiewowych, relacje z organizacjami, uprawnienia/pozwolenia prawne;
- działania:
 - pierwotne: popularyzacja programu, przesiew i identyfikacja dzieci z podwyższonym poziomem ołowiu we krwi,
 - wtórne: określenie postępowania w konkretnym przypadku, skierowanie na leczenie medyczne, ocena środowiska, rozpoczęcie procedur w zakresie zanieczyszczeń środowiska, trening rodzinny;
- produkty: liczba dzieci możliwych do zakwalifikowania, liczba dzieci poddanych przesiewowi, liczba wydanych skierowań do leczenia, liczba „ołowianych” gospodarstw domowych, liczba skierowań do oczyszczania miejsc z ołowiu;
- rezultaty krótkoterminowe: zidentyfikowane źródła/ogniska toksycznej zawartości ołowiu, wyeliminowanie wyżej wspomnianych ognisk, zastosowanie elementów prewencji w środowisku domowym przez rodziny dzieci, osoby z podwyższonym poziomem ołowiu we krwi otrzymują leczenie;
- rezultaty długoterminowe: poprawa jakości życia, zmniejszenie odsetka osób z podwyższonym poziomem ołowiu we krwi, zahamowanie rozwoju zanieczyszczeń na badanym terenie.



Rysunek 2. Matryca logiczna w ewaluacji PPZ

Źródło: opracowanie własne na podstawie dokumentu CDC (2011). Introduction to program evaluation for public health programs: A self-study guide (49)

Aneks 15. Pytania ewaluacyjne w zakresie kryteriów użyteczności i wykonalności celem określenia kierunku ewaluacji

Tabela 13. Przedstawienie pytań ewaluacyjnych w zakresie kryteriów użyteczności i wykonalności celem określenia kierunku ewaluacji

Kryterium ewaluacji	Pytania ewaluacyjne
Użyteczność	<ol style="list-style-type: none"> 1. Jaki jest cel ewaluacji? Cel odnosi się do ogólnej intencji ewaluacji. Wyraźny cel stanowi podstawę dla pytań ewaluacyjnych i metod. Przykładowe wspólne cele: <ul style="list-style-type: none"> • pozyskanie nowej wiedzy w zakresie działań realizowanych w ramach programu, • ulepszenie lub dostosowanie istniejących już procesów programu (np. strategia realizacji programu), • określenie efektów programu w odniesieniu do wkładu programu w osiągnięcie celu długoterminowego, • wpływ programu na jego uczestników jako czynnik motywujący do zmian (utrwalenia zmian) (np. poprzez edukację, szkolenia). 2. Kto będzie wykorzystywał rezultaty pochodzące z ewaluacji? Użytkownicy definiowani są jako osoby lub organizacje, które wykorzystują wyniki ewaluacji. Użytkownicy powinni być zdefiniowani na wcześniejszym etapie w procesie angażowania interesariuszy programu. 3. W jaki sposób zostaną wykorzystane wyniki ewaluacji programu? Zebrane informacje mogą mieć różne zastosowania, które powinny być szczegółowo opisane na etapie projektowania procesu ewaluacji. Poniżej przedstawiono kilka przykładów zastosowań informacji ewaluacyjnych: <ul style="list-style-type: none"> • dokumentacja powodzenia osiągniętych celów programu, • identyfikacja obszarów programu, które wymagają poprawy, • podjęcie decyzji w zakresie alokacji źródeł niezbędnych do realizacji programu, • zmobilizowanie wsparcia danej społeczności, • dystrybucja lub rozszerzenie obszaru, na którym realizowane są interwencje programu, • poprawa jakości realizowanych interwencji, • koncentracja zasobów programu na określonej populacji, • pozyskanie dodatkowych funduszy lub partnerów do realizacji programu. 4. Czego w ramach ewaluacji mogą potrzebować inni interesariusze? Czasami konieczne jest dodanie pytań ewaluacyjnych odnoszących się np. do kosztów. Jest to istotne w przypadku interesariuszy programu zainteresowanych wydatkowaniem środków finansowych (zainwestowanych przez nich).
Wykonalność	<ol style="list-style-type: none"> 1. Na jakim etapie realizacji jest program? Istnieją trzy główne etapy realizacji programu – planowanie, implementacja i utrzymanie (np. efektów programu). Każdy z tych etapów wymaga innego podejścia w zakresie ewaluacji. 2. Z jaką intensywnością realizowany jest program? Programy mogą być nakierowane na jedną główną interwencję a także wielokomponentowe, mogą w nieznacznym stopniu angażować uczestników lub wymagać od nich dużego zaangażowania. Po programach prostych lub powierzchniowych, choć potencjalnie przydatnych, nie można realistycznie oczekiwać znacznego wkładu w poprawę dalszych punktów końcowych. 3. Jakie są istotne kwestie dotyczące zasobów i logistyki w realizacji programu? Niektóre punkty końcowe są szybsze, łatwiejsze i tańsze w zmierzeniu, podczas gdy inne mogą być niemożliwe do zmierzenia. Informacje te mogą wpłynąć na decyzję dot. skierowania ewaluacji na punkty końcowe możliwe do pomiaru.

Źródło: opracowanie własne na podstawie dokumentu CDC (2011). Introduction to program evaluation for public health programs: A self-study guide (49)