

Treatment for Pediatric Gender Dysphoria: Review of Evidence and Best Practices

Leczenie dziecięcej dysforii płciowej: przegląd dowodów i najlepszych praktyk

Tłumaczenie: Magda Lewandowska & WOKiEm

Rozdział 5

Podsumowanie przeglądów systematycznych

W niniejszym rozdziale dokonano przeglądu najlepszych dostępnych informacji dotyczących ryzyka, korzyści i niepewności związanych z interwencjami powszechnie stosowanymi w leczeniu dysforii płciowej (DP) u młodzieży. Podsumowano wyniki systematycznego przeglądu (SP) przeglądów systematycznych (znanego również jako „przegląd parasolowy”) przeprowadzonego na potrzeby niniejszego raportu. W sekcji 1 zdefiniowano niektóre terminy i opisano metodologię zastosowaną do sporządzenia przeglądu. Sekcje 2-6 podsumowują wyniki przeglądu dla następujących interwencji: tranzycja społeczna; blokery dojrzewania (BP); hormony płciowe (HPP); zabiegi chirurgiczne; oraz psychoterapia. Rozdział kończy się dyskusją.

Rozdział 6 wyjaśnia ograniczenia PS oraz potrzebę – zwłaszcza w kontekście pediatricznej medycyny gender (PMG) – rozważenia innych rodzajów dowodów naukowych. Rozdział 7 zawiera przegląd tych innych rodzajów dowodów, w tym dowodów wywodzących się z podstawowych nauk medycznych, znanych mechanizmów działania leków oraz zaburzeń endokrynologicznych, których objawy są podobne do efektów medycznej tranzycji. Ta szersza baza dowodów pozwala na bardziej kompleksową ocenę potencjalnych szkód. Rozdział 8 zawiera krótkie podsumowanie implikacji tych ustaleń w celu umożliwienia podejmowania decyzji.

5.1 Metodologia

Podstawową zasadą medycyny opartej na dowodach naukowych (EBM) jest to, że optymalne podejmowanie decyzji „wymaga świadomości najlepszych dostępnych dowodów naukowych, które w idealnej sytuacji pochodzą z systematycznych podsumowań tych dowodów^[1]”.

Podstawą EBM jest systematyczny przegląd,

[który] ma na celu zebranie wszystkich dowodów empirycznych spełniających wcześniej określone kryteria kwalifikacyjne w celu udzielenia odpowiedzi na konkretne pytanie badawcze. PS wykorzystuje jasne, systematyczne metody, wybierane z myślą o zminimalizowaniu błędów

systematycznych, zapewniając w ten sposób bardziej wiarygodne wyniki, na podstawie których można wyciągać wnioski i podejmować decyzje[2].

Wnioski z każdego indywidualnego badania będą ograniczone projektem tego badania i innymi czynnikami. Zestawienie, synteza i ocena wszystkich istotnych badań dotyczących danego tematu, przy użyciu systematycznych i powtarzalnych metod w celu zminimalizowania błędów systematycznych i pomyłek, umożliwia wyciągnięcie bardziej wiarygodnych wniosków.

PS znajdują się na szczycie „hierarchii dowodów”[3]. Jednak nie wszystkie przeglądy literatury są PS, nawet jeśli tak się nazywają, dlatego ważne jest, aby ocenić proces wykorzystany do opracowania i przedstawienia wszelkich przeglądów dowodów naukowych, które twierdzą, że są PS. Na przykład „Cornell Review” jest szeroko cytowany w dziedzinie PMG w celu poparcia twierdzenia o niskim poziomie żalu, ale wyniki tego przeglądu nie potwierdzają tego wniosku[4].

Przegląd parasolowy zasadniczo stosuje metodologię PS do samych PS, czyniąc jednostką analizy nie poszczególne badania pierwotne, ale całe przeglądy systematyczne. W niniejszym przypadku przygotowano przegląd PS, bo dziedzina PTM jest już nasycona PS, z których wiele ocenia te same badania. Dzięki ocenie jakości tych PS przegląd pozwala na lepsze zrozumienie ogólnej siły, spójności i luk w bazie dowodowej.

Metodologia przeglądu była zgodna z zaleceniami dotyczącymi przeglądów PS zawartymi w Cochrane Handbook for Systematic Reviews of Interventions[5],[6]. Bazy naukowe Medline, Embase i psychologiczną bazę danych PsycINFO przeszukano pod kątem PS dotyczących skutków tranzytacji społecznej, BP, HPP, operacji chirurgicznych i psychoterapii młodzieży w wieku do 26 lat, dotkniętej DP[7]. Przeprowadzono również uzupełniające poszukiwania „szarej literatury”[8] oraz list referencyjnych kwalifikujących się przeglądów systematycznych. Dwóch recenzentów niezależnie przeanalizowało tytuły i streszczenia, a następnie pełne teksty, aby określić kwalifikowalność badań. Recenzenci rozstrzygali spory w drodze dyskusji. Do oceny uwzględnionych przeglądów systematycznych wykorzystano narzędzie do oceny ryzyka błędu systematycznego w przeglądach systematycznych (ROBIS)[9]. Wyodrębnienie danych i ocena ryzyka błędu systematycznego zostały przeprowadzone przez jednego recenzenta i sprawdzone przez drugiego. ROBIS ocenia jakość PS w czterech obszarach: 1) kryteria kwalifikowalności badań; 2) identyfikacja i wybór badań; 3) gromadzenie danych i ocena badań; oraz 4) synteza wyników. Ponadto ROBIS ocenia ogólne ryzyko błędu systematycznego w procesie PS. PS, w przypadku których stwierdzono zastrzeżenia w czterech obszarach, mogą nadal być oceniane jako charakteryzujące się niskim ryzykiem błędu systematycznego, jeśli w interpretacji dowodów odpowiednio uwzględniono te zastrzeżenia.

W pięciu kolejnych sekcjach niniejszego rozdziału podsumowano dane dotyczące wyników z przeglądów systematycznych opublikowanych w języku angielskim i sklasyfikowanych jako charakteryzujące się niskim ryzykiem błędu systematycznego. Syntezę dowodów uporządkowano według wyników będących przedmiotem zainteresowania dla każdej strategii leczenia. Wyniki te obejmowały DP, zdrowie psychiczne i samopoczucie, skutki fizjologiczne (np. supresja hormonów płciowych w przypadku BP), potrzebę dalszego leczenia lub przejście do dalszego leczenia, wyniki dotyczące bezpieczeństwa, w tym skutki uboczne i niepożądane, oraz żal. W przypadku każdego wyniku w niniejszym przeglądzie podsumowano szacunki efektów i jakość dowodów (pewność szacunków efektów) zgodnie z metodologią GRADE (Grading of Recommendations Assessment, Development, and Evaluation, [Klasyfikacja oceny, opracowania i ewaluacji zaleceń, przyp. tł.]).

W metodologii GRADE jakość dowodów jest równoznaczna z pewnością dowodów[10]. Jakość dowodów może zostać obniżona z różnych powodów: ryzyko błędu systematycznego (ograniczenia badania), niespójność (niewyjaśnione różnice w wynikach), pośredniość (dowody niemające bezpośredniego zastosowania), nieprecyzyjność (szerokie przedziały ufności lub małe próby) oraz błąd publikacji (selektywne publikowanie wyników)[11]. Z drugiej strony, jakość może zostać podwyższona, jeśli zaobserwowany jest duży efekt, wyraźna zależność dawka-odpowiedź lub jeśli czynniki zakłócające mogłyby zmniejszyć obserwowany efekt, ale efekt ten został mimo to zaobserwowany[12].

Dowody są następnie klasyfikowane do czterech kategorii według poziomu jakości: wysoka, umiarkowana, niska lub bardzo niska[13]. Dowody wysokiej jakości oznaczają, że „jesteśmy bardzo pewni, że rzeczywisty efekt jest zbliżony do oszacowanego efektu”. Dowody bardzo niskiej jakości wskazują, że „mamy bardzo małą pewność co do szacowania efektu: rzeczywisty efekt prawdopodobnie znacznie różni się od oszacowanego efektu”. W niniejszym przeglądzie podsumowano oceny GRADE z oryginalnych przeglądów systematycznych dla poszczególnych wyników, o ile były one dostępne. Wprowadzono jednak dwie modyfikacje:

1. W przypadku, gdy w PS nie przeprowadzono formalnej oceny GRADE, ale we wnioskach PS użyto sformułowań takich jak „jesteśmy bardzo niepewni” lub „nie można wyciągnąć żadnych wniosków”, uznano je za równoważne ocenie GRADE „bardzo niska jakość”[14].

2. W przypadku, gdy PS nie były zgodne co do oceny GRADE dla tego samego wyniku, w niniejszym przeglądzie rozstrzygnięto tę niezgodność poprzez ocenę de novo zgodnie z metodologią GRADE i przedstawiono uzasadnienie.

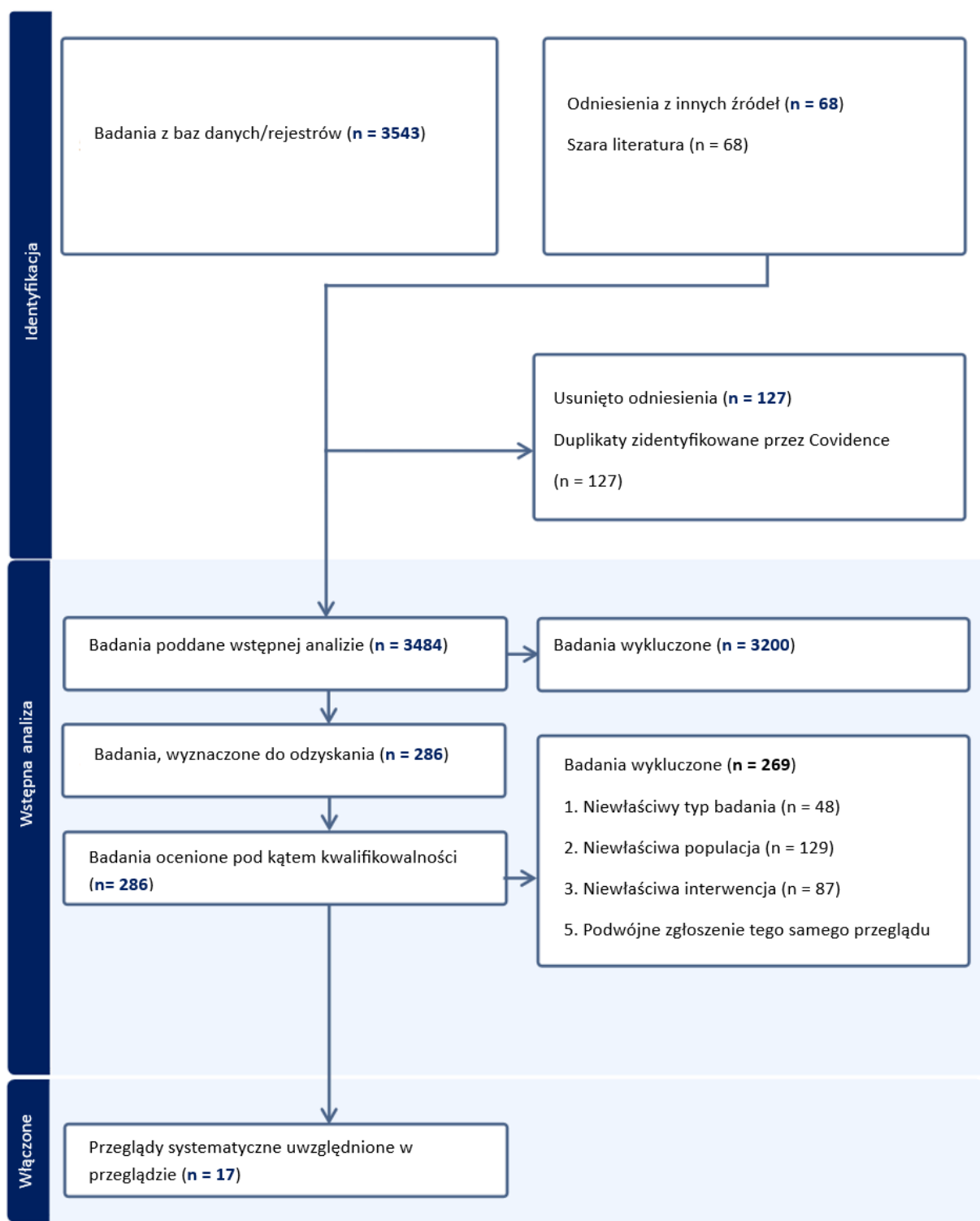
Po wstępnej analizie, 17 metaanaliz spełniło kryteria włączenia. Zbadano w nich wpływ tranżycji społecznej (n = 2)[15], BP (n = 9)[16], HPP (n = 8)[17], zabiegów chirurgicznych (n = 3)[18] i psychoterapii (n = 5)[19] u dzieci i młodzieży z DP. Spośród 17 włączonych metaanaliz dziesięć zostało ocenione jako mające ogólnie niskie ryzyko błędu systematycznego[20]. PS ocenione jako „niskie ryzyko błędu systematycznego” mogą jednak nadal mieć ograniczenia w jednej lub kilku dziedzinach. Siedem PS zostało ocenionych jako mające ogólnie wysokie ryzyko błędu systematycznego[21], często z powodu ograniczonych lub słabo zdefiniowanych pytań badawczych i kryteriów kwalifikowalności, braku oceny ryzyka błędu systematycznego oraz nieodpowiedniej syntezy wyników. Różnice w rygorystyczności metodologicznej podkreślają potrzebę ostrożnej interpretacji wyników, zwłaszcza tych pochodzących z przeglądów ocenionych jako obciążone wysokim ryzykiem błędu systematycznego.

Należy zauważyć, że w niniejszym przeglądzie pominięto dwa przeglądy systematyczne przeprowadzone przez National Institute for Health and Care Excellence (NICE) w październiku 2020 r. na potrzeby przeglądu Cass Review[22]. Zostały one pominięte, gdyż opublikowano kilka kolejnych przeglądów systematycznych, zleconych Uniwersytetowi Yorku na potrzeby przeglądu Cass Review. Te ostatnie PS są bardziej kompleksowe i aktualizują wyniki dwóch PS NICE.

W niniejszym przeglądzie pominięto również PS z 2021 r. autorstwa Baker i in.[23], ponieważ uczestnikami większości badań uwzględnionych w tym PS byli dorośli w dojrzałym wieku. Warto zauważyć, że przegląd Baker i in. został przytoczony na poparcie globalnego oświadczenia 2.1 zawartego w opracowanych przez World Professional Association for Transgender Health (Światowego Stowarzyszenia Zawodowego na rzecz Zdrowia Osób Trans) „Standardach opieki zdrowotnej dla osób trans i o różnorodnej tożsamości płciowej”, wersja 8 (SOC-8)[24]. Oświadczenie to zaleca: „Zalecamy, aby systemy opieki zdrowotnej zapewniały osobom trans i o różnorodnej tożsamości płciowej niezbędną z medycznego punktu widzenia opiekę zdrowotną

afirmującą tożsamość płciową”. Stwierdzenie to zostało następnie przytoczone na poparcie zaleceń zawartych w rozdziale SOC-8 poświęconym nastolatkom. Ze względu na sposób, w jaki PS został wykorzystany w dziedzinie PTM, Baker i in. zostali poddani oddzielnej analizie ROBIS, która wykazała wysokie ryzyko błędu systematycznego ze względu na ograniczenia w domenach ROBIS „gromadzenie danych i ocena badania” oraz „synteza i wyniki”. Ocena ROBIS dotycząca Baker i in. znajduje się w sekcji 2.3 załącznika 4, natomiast załącznik 3 zawiera wprowadzenie do PS i EBM. Rysunek 5.1 poniżej przedstawia schemat procesu wyszukiwania, selekcji i włączenia.

Rysunek 5.1 Proces wyszukiwania, selekcji i włączenia



5.2 Wyniki tranzykcji społecznej

Tranzykcja społeczna polega na zmianie jednego lub więcej aspektów wyglądu lub ekspresji danej osoby, takich jak imię, wygląd zewnętrzny lub zachowanie, w celu bycia postrzeganym i traktowanym jako osoba płci przeciwnej lub uniknięcia bycia postrzeganym i traktowanym jako osoba tej samej płci. Jak zauważono w przeglądzie Cass, nawet jeśli tranzykcja społeczna odbywa się poza placówkami opieki zdrowotnej, „ważne jest, aby postrzegać [tranzykcję społeczną] jako

aktywną interwencję, bo może ona mieć znaczący wpływ na dziecko lub młodą osobę pod względem funkcjonowania psychicznego i długoterminowych wyników” [25].

W niniejszym przeglądzie zidentyfikowano dwa przeglądy systematyczne oceniające wpływ tranzykcji społecznej. Oba zostały ocenione jako charakteryzujące się niskim ryzykiem błędu systematycznego [26]. Wyniki sugerują, że wpływ tranzykcji społecznej na długoterminową DP, wyniki psychologiczne i samopoczucie oraz przyszłe decyzje dotyczące leczenia, takie jak terapia hormonalna lub operacje, pozostaje słabo poznany. Dowody dotyczące żalu związanego z tranzycją społeczną są bardzo ograniczone. Pewność dowodów dotyczących tych wyników jest bardzo niska.

Istnieją znaczne luki w dowodach dotyczących ewaluacji tranzykcji społecznej jako interwencji u dzieci i młodzieży z DP. Większość dostępnych badań ma charakter przekrojowy i nie ma prospektywnych badań longitudinalnych z odpowiednimi grupami porównawczymi, co sprawia, że nie jest jasne czy zaobserwowane powiązania odzwierciedlają skutki tranzykcji społecznej, czy też inne czynniki leżące u jej podstaw. Ponadto opublikowane badania często nie rozróżniają skutków tranzykcji społecznej od skutków równoczesnych interwencji, takich jak psychoterapia lub leczenie farmakologiczne, co dodatkowo utrudnia interpretację.

5.3 Wyniki stosowania leków blokujących dojrzewanie

Agoniści hormonu uwalniającego gonadotropinę (GnRHa), znane jako „bloker pokwitania” (BP), są stosowane w celu zapobiegania lub zatrzymania rozwoju cech płciowych u dzieci i młodzieży z DP w okresie około-pokwitaniowym i pokwitaniowym. W niniejszym przeglądzie zidentyfikowano dziewięć metaanaliz, w których oceniano skutki leczenia BP. Wśród nich znalazły się cztery metaanalizy w języku angielskim [27], które oceniono jako obarczone niskim ryzykiem błędu systematycznego. Kryteria kwalifikacji i liczba uwzględnionych badań różniły się w tych czterech metaanalizach, a pewność dowodów dotyczących wpływu BP na DP (lub inkongruencję płciową), poprawę zdrowia psychicznego i bezpieczeństwo jest bardzo niska. Istnieją dowody o wysokim stopniu pewności, że BP wywierają działanie fizjologiczne (takie jak supresja hormonów płciowych) i, w zależności od stadium dojrzewania i płci pacjenta [28], często powodują bezpłodność, jeśli są stosowane w połączeniu z HPP. Dowody o niskim stopniu pewności sugerują, że BP mogą mieć negatywny wpływ na zdrowie kości [29]. Duża część młodzieży po zastosowaniu BP przechodzi do HPP, chociaż pewność dowodów dotyczących jakiegokolwiek przyczynowej roli BP w tym zjawisku jest bardzo niska.

W istniejącej literaturze nadal istnieją istotne luki zarówno w zakresie, jak i jakości ocenianych wyników. Wiele badań pierwotnych nie zostało odpowiednio zaprojektowanych do pomiaru lub raportowania wyników związanych z BP. Na przykład niewiele badań pierwotnych uwzględnionych w przeglądach systematycznych oceniało wpływ BP na wyniki, takie jak DP lub zdrowie psychiczne. Chociaż BP są często opisywane jako „pauza” [30], żadne badania nie zbadały systematycznie ich roli w procesie podejmowania decyzji ani wyników pacjentów, którzy przerwali leczenie.

Podczas oceny wyników skupiono się głównie na krótkotrwałych zmianach psychologicznych lub fizjologicznych lub wynikach zastępczych, takich jak myśli samobójcze lub gęstość mineralna kości, a nie na punktach końcowych, takich jak samobójstwa lub złamanie. Dowody są szczególnie ograniczone w odniesieniu do długoterminowych wyników związanych z płodnością, wzrostem i rozwojem neurokognitywnym.

Chociaż badania sugerują, że znaczna część młodzieży po zahamowaniu dojrzewania przechodzi na HPP[31], istnieją minimalne dowody na wpływ ścieżki BP/HPP w porównaniu z samymi HPP lub na to, jak ta ścieżka wpływa na długoterminowe wyniki, takie jak poddanie się operacji. Ponadto nie przeprowadzono żadnych badań, które jasno zbadałyby fizyczne lub psychospołeczne trajektorie niewielkiej mniejszości osób, które przyjmowały BP, ale nie przeszły na HPP. W rezultacie istnieje niewiele danych na temat tego, co dzieje się po zakończeniu leczenia, a założenie, że skutki BP są nieodwracalne, pozostaje w dużej mierze niesprawdzone. Innym ważnym ograniczeniem jest to, że w większości badań pierwotnych nie rozróżniono skutków BP u pacjentów płci żeńskiej i męskiej. Ogólnie rzecz biorąc, brak długoterminowych, wysoce wiarygodnych dowodów dotyczących najważniejszych wyników (w tym samobójstw, złamań, płodności, wzrostu i rozwoju neurokognitywnego) pozostawia znaczną niepewność co do efektów BP.

5.4 Wyniki stosowania hormonów płciowych

Estrogen dla mężczyzn i testosteron dla kobiet są stosowane poza wskazaniami rejestracyjnymi w celu wywołania zmian fizycznych cech płciowych młodzieży z DP. W niniejszym przeglądzie zidentyfikowano osiem metaanaliz dotyczących skutków leczenia HPP, z których cztery były w języku angielskim i oceniono je jako „niskie ryzyko błędu systematycznego”[32]. Pewność dowodów jest bardzo niska w odniesieniu do wpływu na DP lub inkongruencję, poprawę zdrowia psychicznego oraz bezpieczeństwo, w tym płodność i zdrowie kośćca. Istnieją dowody o wysokim stopniu pewności, że HPP wywiera skutki fizjologiczne.

Podobnie jak w przypadku BP, istnieją istotne luki w dowodach dotyczących HPP. Wiele badań nie zostało specjalnie zaprojektowanych w celu uchwycenia pełnego zakresu długoterminowych wyników i koncentrowało się głównie na krótkoterminowych zmianach psychologicznych lub fizjologicznych. Kluczowe wyniki, takie jak wpływ na DP, inne wyniki dotyczące zdrowia psychicznego i jakości życia, były mierzone w sposób niespójny, a gdy były zgłaszane, często pochodziły z niewielkich badań obserwacyjnych z ograniczoną długością obserwacji. Niezbyt dobrze poznane pozostają niezwykle ważne długoterminowe wyniki. Zaburzenia funkcji seksualnych, pomimo tego, że są bardzo istotne dla długoterminowego dobrostanu, były rzadko oceniane. Chociaż w kilku badaniach odnotowano częstość występowania zdarzeń sercowo-naczyniowych, do oceny skumulowanego ryzyka konieczna jest długoterminowa obserwacja[33].

Dowody dotyczące płodności są skąpe, a dane na temat tego, czy skutki dla płodności różnią się w zależności od wieku w momencie rozpoczęcia leczenia lub czy skutki te są odwracalne, są niewielkie. Ponadto nie zbadano w wystarczającym stopniu złożonych skutków sekwencyjnego (a czasami równoczesnego, zwłaszcza u pacjentów płci męskiej) leczenia BP, a następnie HPP, takich jak ich wpływ na regenerację kości, ostateczny wzrost w wieku dorosłym lub prawdopodobieństwo konieczności interwencji chirurgicznej. Podobnie jak w przypadku BP, brak rygorystycznych metodologicznie, długoterminowych badań znacznie ogranicza naszą wiedzę na temat szerszych skutków zdrowotnych stosowania HPP w okresie dojrzewania. Brak badań pierwotnych, w których oddzielnie oceniono wpływ estrogeny u mężczyzn i testosteronu u kobiet, stanowi istotne ograniczenie, ponieważ interakcje między hormonami płciowymi mogą mieć znaczący wpływ na wyniki leczenia.

5.5 Wyniki zabiegów chirurgicznych

Maskulinizująca mastektomia u nastoletnich pacjentek jest najczęściej wykonywaną operacją u młodzieży z DP. W niniejszym przeglądzie zidentyfikowano trzy przeglądy systematyczne

dotyczące operacji u dzieci lub nastolatków z DP, z których większość dotyczyła wyłącznie mastektomii[34]. Dwa PS oceniono jako charakteryzujące się niskim ryzykiem błędu systematycznego[35]. Istnieją bardzo pewne dowody na to, że mastektomia wiąże się z przewidywalnymi powikłaniami chirurgicznymi, takimi jak martwica i blizny. Pewność dowodów jest bardzo niska w odniesieniu do wpływu tej operacji na DP lub inkongruencję płciową, poprawę zdrowia psychicznego, w tym skłonności samobójcze i depresję, oraz długoterminowe wyniki, takie jak funkcje seksualne, jakość życia i żal.

Istnieją znaczne luki w dowodach dotyczących operacji u nastolatków z DP. Większość badań to serie przypadków lub małe badania obserwacyjne, z ograniczonymi grupami kontrolnymi lub bez takich grup, a zatem nie są one w stanie wyodrębnić skutków operacji od wcześniejszych interwencji medycznych lub psychospołecznych. Wyniki takie jak wpływ na DP, zdrowie psychiczne i jakość życia, są zgłaszane w sposób niespójny i często brakuje im walidacji. Wyniki długoterminowe, w tym trwałość korzyści psychologicznych, funkcje seksualne, potrzeba ponownych operacji, a także satysfakcja i dobre samopoczucie w wieku dorosłym, pozostają słabo scharakteryzowane. Ponadto, chociaż żal jest częstym tematem dyskusji publicznej, istniejące badania nie dostarczają solidnych danych na temat czynników wpływających na żal. Istnieją również luki w wiedzy na temat innych rodzajów operacji, w tym operacji narządów płciowych.

5.6 Wyniki psychoterapii

Dyskusje na temat roli psychoterapii w leczeniu młodzieży z DP charakteryzują się wewnętrznymi niespójnościami medycyny gender, gdzie psychoterapia jest zarówno uznawana za ważne narzędzie, jak i stygmatyzowana, jeśli jej celem jest rozwiązanie problemu DP (rozdział 14). W przeglądzie zidentyfikowano łącznie pięć metaanaliz oceniających psychoterapię u dzieci i młodzieży z DP. Tylko dwie z nich zostały ocenione jako posiadające niskie ryzyko błędu systematycznego[36].

Dowody dotyczące efektów psychoterapii są ograniczone. Przegląd przeprowadzony w 2024 r. przez Dopp i in. obejmował jedno randomizowane badanie kontrolowane (RCT [od ang. randomised controlled trial, przyp. tł.]) oceniające wpływ „krótkich filmów wideo ukierunkowanych na radzenie sobie” na skłonności samobójcze, które nie dostarczyło bezpośrednich dowodów na skuteczność psychoterapii u młodzieży z DP. Nie ma innych RCT. W obu przeglądach systematycznych stwierdzono, że interwencje psychoterapeutyczne były realizowane w różnych formatach, co wskazuje na znaczną heterogeniczność w zakresie implementacji tej interwencji. Formaty interwencji bezpośrednich i internetowych były bardzo zróżnicowane i obejmowały podejście indywidualne, grupowe, rodzinne i łączone. Interwencje bezpośrednio odbywały się w różnych środowiskach, takich jak specjalistyczne kliniki gender, programy społecznościowe, społeczne kliniki zdrowia psychicznego, weekendowe rekolekcje i program intensywnego leczenia stacjonarnego. Ta heterogeniczność ogranicza możliwość uogólnienia dowodów. W przypadku wyników dotyczących zdrowia psychicznego pewność dowodów była bardzo niska. Nie zgłoszono jednak żadnych szkód.

W przeglądzie nie znaleziono dowodów na wpływ psychoterapii na samą DP. Ta luka w dowodach może wynikać z powiązania psychoterapii z „terapią konwersyjną”. Jak zauważono w przeglądzie Cass, „rola terapii psychologicznych we wspieraniu dzieci i młodzieży z inkongruencją płciową lub dystresem została przyćmiona przez niekonstruktywną, spolaryzowaną debatę na temat praktyk konwersyjnych”[37]. Istnieją bardzo ograniczone dowody na rolę psychoterapii w leczeniu problemów zdrowia psychicznego współwystępujących

z DP u dzieci i młodzieży[38], w tym depresji, stanów lękowych, zaburzeń odżywiania, samookaleczania i skłonności samobójczych. Wczesne doniesienia często nie oddzielały efektów psychoterapii od efektów innych równoczesnych interwencji, takich jak tranzycja społeczna, BP lub HPP, co utrudniało przypisanie obserwowanych wyników wyłącznie psychoterapii. Niewiele wiadomo na temat tego, które podejścia terapeutyczne mogą być bardziej lub mniej skuteczne w przypadku poszczególnych podgrup. Brak solidnych badań został potwierdzony w raporcie Cass, w którym zauważono, że „nie podjęto systematycznych rozważań na temat tego, w jaki sposób należy stosować interwencje psychospołeczne, ani nie zbadano ich skuteczności”[39].

Warto zauważyć, że chociaż bezpośrednie dowody dotyczące psychoterapii u dzieci i młodzieży z DP są ograniczone, istnieją dowody potwierdzające rolę psychoterapii w leczeniu dzieci i młodzieży z innymi problemami zdrowia psychicznego, w tym depresją, stanami lękowymi, zaburzeniami odżywiania, samookaleczaniem i skłonnościami samobójczymi (zob. sekcja 14.5.1).

5.7 Dyskusja

5.7.1 Ustalenia niniejszego przeglądu

Niniejszy przegląd przeglądów systematycznych obejmuje 17 przeglądów, oceniających skutki interwencji u dzieci i młodzieży z GD. W przypadku tranzycji społecznej pewność korzyści i szkód jest bardzo niska ze względu na problemy związane z tym, jak badania zostały zaprojektowane. Chociaż badania sugerują, że wczesna tranzycja społeczna wiąże się z wysokim odsetkiem utrzymywania się DP i >90% odsetkiem kontynuacji BP i HPP[40], z metodologicznego punktu widzenia istnieje niepewność ze względu na bardzo niską jakość dowodów wyjaśniających przyczynowość/ścieżki przyczynowe.

W odniesieniu do tranzycji medycznej, niniejszy przegląd podkreśla spójny wzorzec występujący we wszystkich interwencjach dotyczących dzieci i młodzieży z DP: chociaż BP, HPP i zabiegi chirurgiczne niezawodnie powodują oczekiwane zmiany fizjologiczne, nadal istnieje znaczna niepewność co do ich wpływu na zdrowie psychiczne i długoterminowe skutki zdrowotne. Dowody wysokiej jakości potwierdzają, że BP hamują rozwój w okresie pokwitania, a HPP powodują zmiany cech płciowych. Istnieją dowody niskiej jakości sugerujące, że BP mają negatywny wpływ na zdrowie kości[41]. Jednak jakość dowodów dotyczących takich wyników, jak DP, zdrowie psychiczne, jakość życia i żal, pozostaje bardzo niska we wszystkich rodzajach interwencji. Do interpretacji wielu wyników przeglądów systematycznych należy podchodzić z ostrożnością, zwłaszcza biorąc pod uwagę źródła niepewności, solidność wniosków niniejszego przeglądu oraz ograniczenia metodologiczne przeglądów systematycznych.

5.7.2 Źródła niepewności dowodów naukowych

We wszystkich uwzględnionych przeglądach systematycznych jakość (pewność) dowodów jest ograniczona głównie przez brak rygorystycznych metodologicznie badań. Na przykład we wszystkich badaniach pierwotnych uwzględnionych w dziewięciu przeglądach systemowych dotyczących BP nie przeprowadzono żadnych uznanych randomizowanych badań kontrolowanych.

Brak odpowiednich grup kontrolnych jest kolejnym powszechnym ograniczeniem metodologicznym badań. W niniejszym przeglądzie stwierdzono, że spośród wszystkich badań pierwotnych w dziewięciu przeglądach systematycznych dotyczących BP tylko sześć było

badaniami obserwacyjnymi z równoległymi grupami kontrolnymi, w których porównano stosowanie BP z brakiem stosowania BP u dzieci i młodzieży z DP[42]. Jednak na ogólną pewność dowodów wpłynęły nie tylko niedociągnięcia metodologiczne, ale także takie czynniki, jak mała wielkość próby oraz zaobserwowanie niewielkich lub braku efektów, co w terminologii GRADE oznacza nieprecyzyjność. Na przykład wielkość próby w sześciu porównawczych badaniach obserwacyjnych dotyczących BP była niewielka, a trzy badania obejmowały mniej niż 100 uczestników. W jednym PS obniżono jakość dowodów dotyczących wpływu BP na ogólne funkcjonowanie ze względu na nieprecyzyjność, zauważając, że dowody „sugerują zarówno możliwość korzyści, jak i szkody badanych interwencji[43]”.

Niespójne wyniki badań są kolejną często wymienianą kwestią obniżającą jakość dowodów. Na przykład w jednym PS dotyczącym HPP stwierdzono „ograniczone lub niespójne dowody dotyczące DP, satysfakcji z wyglądu ciała, wyników psychospołecznych i neurokognitywnych, płodności, wysokości/wzrostu, zdrowia kości i efektów sercowo-metabolicznych[44]”. Podobne ograniczenia zaobserwowano w badaniach oceniających inne interwencje medyczne i chirurgiczne u młodzieży z DP.

Dlatego ważne jest, aby wyjaśnić, że brak dowodów wysokiej lub nawet umiarkowanej jakości nie wynika wyłącznie z braku randomizowanych badań kontrolowanych (RCT). Jakość dowodów można by poprawić poprzez dobrze przeprowadzone badania obserwacyjne[45], które mogą dostarczyć dowodów umiarkowanej, a nawet wysokiej jakości, jeśli zaobserwowane efekty są duże, spójne i wyraźnie związane z interwencją.

5.7.3 Solidność wniosków niniejszego przeglądu

Chociaż niniejszy przegląd skupiał się na angielskojęzycznych PS ocenionych jako mające niskie ryzyko błędu systematycznego, inny PS o niskim ryzyku błędu systematycznego, opublikowany w języku niemieckim, zawierał podobne wnioski, potwierdzające ogólne wzorce zidentyfikowane w niniejszym przeglądzie. Jak wspomniano wcześniej, niniejszy przegląd nie obejmował PS dotyczących BP i HPP zleconych przez NHS w 2020 r., ponieważ zostały one zastąpione przez PS opublikowane w 2024 r. na potrzeby przeglądu Cass. Jednak dwa wcześniejsze przeglądy systematyczne również stwierdziły, że jakość dowodów była bardzo niska[46], a ich uwzględnienie nie zmieniłoby wniosków niniejszego przeglądu.

Ponadto należy również zauważyć, że chociaż niniejszy przegląd skupiał się na przeglądach systematycznych ocenionych jako charakteryzujące się niskim ryzykiem błędu systematycznego, nie oznacza to, że przeglądy te były pozbawione ograniczeń metodologicznych. „Niskie ryzyko błędu systematycznego w przeglądzie systematycznym” w ROBIS pozwala na pewną elastyczność. Uwzględnione przeglądy systematyczne nadal miały pewne niedociągnięcia, takie jak ograniczone strategie wyszukiwania, nieaktualne wyszukiwanie literatury[47] lub brak precyzji w ocenie dowodów. Natomiast przeglądy systematyczne ocenione jako obciążone wysokim ryzykiem błędu systematycznego wykazywały poważne wady metodologiczne, w tym źle zdefiniowane kryteria kwalifikowalności, brak oceny ryzyka błędu systematycznego dla uwzględnionych badań oraz niesystematyczne podejście do syntezy. Przeglądy te uznano za zbyt stronnicze, aby można było na ich podstawie sformułować wiarygodne wnioski. Wyniki przedstawione w niniejszym przeglądzie odzwierciedlają zatem celowo ostrożne podejście, opierające się wyłącznie na najbardziej wiarygodnych źródłach dostępnych w chwili obecnej.

Wreszcie, zamiast wykraczać poza zakres dowodów, niniejszy przegląd ogranicza się do podsumowania wniosków zawartych w przeglądach systematycznych. W rezultacie może on nie

obejmować niektórych najnowszych badań (ze względu na termin przeprowadzenia wyszukiwania literatury w ramach przeglądów systematycznych). Jednak ukierunkowane wyszukiwanie[48] ostatnio opublikowanych badań nie ujawniło żadnych opublikowanych lub trwających badań, które mogłyby znacząco zmienić wnioski, zwłaszcza te dotyczące korzyści. Wynika to z utrzymujących się problemów, takich jak brak grup kontrolnych, nieodpowiednia wielkość próby i ograniczona obserwacja. Jeśli planowane brytyjskie badanie RCT dotyczące BP[49] otrzyma zgodę komisji etycznej, może ono poprawić bazę dowodową dotyczącą krótkoterminowych (24-miesięcznych) skutków BP w odniesieniu do niektórych miar wyników. Jednakże, w oparciu o dotychczas ujawnione publicznie elementy projektu badania, może ono nie dostarczyć informacji na temat skutków leczenia, gdy po BP stosuje się HPP – co wydaje się mieć miejsce w ponad 90% przypadków[50]. Taka łączna terapia miałaby inny profil ryzyka i korzyści niż stosowanie samych BP, zarówno pod względem potencjalnych korzyści, jak i potencjalnych szkód. Wreszcie, przewidywana krótkoterminowa obserwacja nie pozwoli na odpowiednie zbadanie kluczowej kwestii, jaką są długoterminowe wyniki.

Bez znaczącej zmiany kierunku ku bardziej solidnym projektom badawczym — takim jak randomizowane badania kontrolowane, w których bada się każdą interwencję w jej aktualnej postaci (tj. sekwencyjne leczenie BP, a następnie HPP), prospektywne kohorty z odpowiednimi grupami porównawczymi lub eksperymenty naturalne — jakość dowodów prawdopodobnie pozostanie niska[51]. Przeprowadzenie w najbliższej przyszłości nowych przeglądów systematycznych w celu uzyskania odpowiedzi na te same pytania badawcze prawdopodobnie nie przyniesie nowych wniosków. Ten brak solidnych badań i dowodów ma bezpośredni wpływ na wytyczne kliniczne i kształtowanie protokołów, gdzie oczekiwane korzyści muszą być rozważane w świetle utrzymującej się niepewności dowodów udokumentowanych w literaturze.

5.7.4 Ograniczenia i mocne strony niniejszego przeglądu

Wszystkie przeglądy systematyczne i przeglądy przeglądów systematycznych mają ograniczenia. Przeglądy systematyczne mają na celu uchwycenie wszystkich istotnych badań, które spełniają z góry określone kryteria kwalifikacyjne. Cel ten może jednak zostać podważony przez błąd systematyczny wynikający z raportowania (lub braku raportowania) — formę błędu systematycznego, która pojawia się, gdy na decyzję o opublikowaniu, opóźnieniu lub wybiórczym przedstawieniu wyników badania ma wpływ istotność statystyczna, wielkość lub kierunek wyników[52].

Jak omówiono w rozdziale 6 niniejszego przeglądu, badania w dziedzinie PMG charakteryzują się znacznym błędem publikacji i raportowania. Skala tego problemu jest trudna do oszacowania ze względu na powszechne stosowanie słabej metodologii i brak spójności w raportowaniu wielkości efektów[53]. Co ważne, stroniczość publikacji nie została uznana za powód obniżenia pewności dowodów w uwzględnionych przeglądach systematycznych.

Kolejnym ograniczeniem jest to, że celem przeglądów systematycznych jest udzielenia odpowiedzi na dobrze zdefiniowane pytania badawcze, które mają z góry określone wyniki, ale PS nie mogą wykrywać sygnałów ani generować hipotez. Wiele badań w zakresie PMG koncentruje się na wynikach psychologicznych, ale tylko w sposób niespójny śledzi inne wyniki, które mogą być sygnałami niepożądanych skutków interwencji. W przypadku gdy wyniki nie są zgłaszane przy użyciu rygorystycznych standardów metodologicznych – takich jak wyraźne wymienienie w specjalnej tabeli wyników – a zamiast tego są jedynie wspomniane w części poświęconej omówieniu, metodologia PS prawdopodobnie nie oznaczy badania jako istotnego dla danego wyniku podczas procesu selekcji. W rezultacie ważne niekorzystne skutki mogą

pozostać niezauważone. Kwestia ta została omówiona dość szczegółowo w następnym rozdziale.

Kluczową cechą przeglądów systematycznych i przeglądów przeglądów systematycznych jest nacisk na dane epidemiologiczne na poziomie populacji[54], często przy ograniczonym uwzględnieniu dowodów mechanistycznych lub rozważań dotyczących biologicznej wiarygodności. Ten wąski zakres ogranicza możliwość wyjaśnienia w niniejszym przeglądzie potencjalnych ścieżek przyczynowych lub pełnej interpretacji zaobserwowanych powiązań – lub ich braku.

Niniejszy przegląd PS ma również inne ograniczenia. Podobnie jak wszystkie przeglądy, ogranicza się on do syntezy opublikowanych PS i nie zawiera ponownej analizy badań pierwotnych. Zakres niniejszego przeglądu został celowo zawężony, koncentrując się na młodych pacjentach i wykluczając pośrednie dowody pochodzące z populacji dorosłych, które mogłyby dostarczyć informacji na temat długoterminowych skutków.

Do mocnych stron zalicza się systematyczną ocenę ryzyka błędu systematycznego w przeglądach przy użyciu narzędzia ROBIS. W ten sposób niniejszy przegląd oferuje przejrzystą ocenę wiarygodności wniosków każdego PS. Umożliwia również porównanie PS z badaniami podobnych interwencji, ale różniącymi się zakresem, pomagając w identyfikacji spójnych wyników, luk metodologicznych oraz obszarów zgodności lub rozbieżności.

5.7.5 Wnioski

W niniejszym przeglądzie zsyntetyzowano najlepsze dostępne dowody kliniczne pochodzące z danych populacyjnych, podkreślając spójny wzorzec interwencji stosowanych u dzieci i młodzieży z DP. Korzyści i szkody wynikające z tranzycji społecznej pozostają nieznane; BP, HPP i zabiegi chirurgiczne konsekwentnie powodują pewne skutki fizyczne i fizjologiczne; istnieje również znaczna niepewność co do ich skutków psychologicznych i długoterminowych skutków zdrowotnych.

Tłumaczenie: Magda Lewandowska

[1] G. Guyatt i in. (2015, s. 10).

[2] Higgins i in. (2019, s. xxiii).

[3] Zob. załącznik 3.

[4] Cornell University & What We Know Project (2017). W przeglądzie Cornell stwierdzono, że „żał wynika najczęściej z braku wsparcia społecznego po tranzycji lub złych wyników operacji przeprowadzonych przy użyciu starszych technik”. W przeglądzie tym nie dokonano jednak właściwej syntezy dowodów pochodzących z uwzględnionych badań ani nie oceniono jakości dowodów. Wyniki przeglądu należy zatem interpretować z ostrożnością. Przegląd Cornell nie został uwzględniony w niniejszym przeglądzie, ponieważ populacją docelową były osoby dorosłe, a zatem nie ma on wyraźnego zastosowania do dzieci i młodzieży z DP (zob. sekcja 4.3.1.1).

[5] Higgins i in. (2019).

[6] Szczegółowa metodologia i wyniki niniejszego przeglądu znajdują się w załączniku 4.

[7] Przeglądy systematyczne w tej dziedzinie zazwyczaj dotyczyły pacjentów poniżej 18 lub 26 roku życia. Aby uwzględnić osoby w wieku 18 lat, w przeglądzie podjęto decyzję o wyszukaniu przeglądów systematycznych dotyczących populacji pacjentów poniżej 26 roku życia. Pozwala to zachować równowagę między kompleksowością a bezpośredniością (możliwością zastosowania) przeglądu dowodów naukowych. Chociaż istnieją obawy dotyczące bezpośredniości dowodów, dowody dotyczące pacjentów poniżej 18 roku życia i pacjentów powyżej 18 roku życia nie wykazały żadnych istotnych różnic w wynikach.

[8] Określenie „szara literatura” odnosi się do badań i informacji powstałych poza tradycyjnymi kanałami publikacji naukowych, a zatem takich, które mogą nie być formalnie recenzowane ani indeksowane w głównych bazach danych.

[9] Whiting i in. (2016).

[10] „Bardzo niska jakość dowodów” jest równoznaczna z „bardzo niską pewnością dowodów” (Balslem i in., 2011; G. H. Guyatt i in., 2008).

[11] Balslem i in. (2011); G. H. Guyatt i in. (2008); G. H. Guyatt, Oxman, Kunz, Brozek i in. (2011); G. H. Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Falck-Ytter i in. (2011); G. H. Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Glasziou i in. (2011); G. H. Guyatt, Oxman, Montori i in. (2011); G. H. Guyatt, Oxman, Vist i in. (2011).

[12] G. H. Guyatt, Oxman, Sultan i in. (2011).

[13] Balslem i in. (2011).

[14] Balslem i in. (2011); Santesso i in. (2020).

[15] Dopp i in. (2024); Hall i in. (2024).

[16] Chew i in. (2018); Dopp i in. (2024); Ludvigsson i in. (2023); Miroshnychenko, Roldan i in. (2025); Ramos i in. (2021); Rew i in. (2021); Taylor, Mitchell, Hall, Heathcote i in. (2024); Thompson i in. (2023); Zepf, König i in. (2024).

[17] Chew i in. (2018); Dopp i in. (2024); Karalexi i in. (2020); Ludvigsson i in. (2023); Miroshnychenko, Ibrahim i in. (2025); Taylor, Mitchell, Hall, Langton i in. (2024); Thompson i in. (2023); Zepf, König i in. (2024).

[18] Dopp i in. (2024); Miroshnychenko i in. (2024); Thompson i in. (2023).

[19] Dopp i in. (2024); Expósito-Campos i in. (2023); Heathcote i in. (2024); Malpas i in. (2022); Thompson i in. (2023).

[20] Dopp i in. (2024); Hall i in. (2024); Heathcote i in. (2024); Ludvigsson i in. (2023); Miroshnychenko i in. (2024); Miroshnychenko, Ibrahim i in. (2025); Miroshnychenko, Roldan i in. (2025); Taylor, Mitchell, Hall, Heathcote i in. (2024); Taylor, Mitchell, Hall, Langton i in. (2024); Zepf, König i in. (2024). PS przeprowadzone przez tę samą grupę badawczą były zasadniczo zgodne z tą samą metodologią.

[21] Chew i in. (2018); Expósito-Campos i in. (2023); Karalexi i in. (2020); Malpas i in. (2022); Ramos i in. (2021); Rew i in. (2021); Thompson i in. (2023).

[22] National Institute for Health and Care Excellence (2020b, 2020a).

[23] Baker i in. (2021).

[24] Coleman i in. (2022).

[25] Cass (2024, s. 158).

[26] Dopp i in. (2024); Hall i in. (2024).

[27] Dopp i in. (2024); Ludvigsson i in. (2023); Miroshnychenko, Roldan i in. (2025); Taylor, Mitchell, Hall, Heathcote i in. (2024).

[28] Zob. dalsza dyskusja na temat płodności w rozdziale 7.

[29] Zob. dalsza dyskusja na temat fizjologicznych mechanizmów, poprzez które indukowany hipogonadotropowy hipogonadyzm negatywnie wpływa na rozwój kości.

[30] Zob. np. ulotkę informacyjną dotyczącą blokerów dojrzewania opracowaną przez Cambridge Health Alliance/Harvard Medical School Teaching Hospital (Cambridge Health Alliance, b.d.).

[31] W przeglądach systematycznych, które zostały ocenione jako posiadające niskie ryzyko błędu systematycznego, tylko dwa badania Carmichael i in. (2021); Karakılıç Özturan i in. (2023) podały dane dotyczące tego wyniku, z których każde wykazało wskaźnik progresji od BP do HPP wynoszący >95%. Inne badania, które nie zostały uwzględnione w niniejszym przeglądzie, cytowane w innych częściach niniejszego przeglądu, również odnotowały >90% wskaźnik progresji z BP do HPP. Brik i in. (2020); van der Loos i in. (2023);

Wiepjes i in. (2018).

[32] Dopp i in. (2024); Ludvigsson i in. (2023); Miroshnychenko, Ibrahim i in. (2025); Taylor, Mitchell, Hall, Langton i in. (2024).

[33] Dotyczy to w szczególności tranzykcji medycznej u dzieci (ze względu na rozpoczęcie HPP w okresie dojrzewania czas trwania leczenia będzie dłuższy niż w przypadku wcześniej badanych populacji dorosłych).

[34] Dopp i in. (2024) uwzględnili 18 badań; ponad połowa z nich dotyczyła mastektomii. Miroshnychenko i in. (2024) uwzględnili 39 badań i ocenili wyłącznie skutki mastektomii. Inny PS obejmował 25 uczestników poddanych zabiegom chirurgicznym, z których 24 przeszło mastektomię. Zob. Dopp i in. (2024); Miroshnychenko i in. (2024); Thompson i in. (2023).

[35] Dopp i in. (2024); Miroshnychenko i in. (2024).

[36] Dopp i in. (2024); Heathcote i in. (2024).

[37] Cass (2024a, s. 150).

[38] Zob. rozdział 14.

[39] Cass (2024a, s. 155).

[40] R. Hall i in. (2024).

[41] Fizjologiczny mechanizm wpływu PB na mineralizację kości zostanie omówiony w rozdziale 7.

[42] Becker-Hebly i in. (2021); Costa i in. (2015); Grimstad i in. (2021); Jensen i in. (2019); McGregor i in. (2024); Tordoff i in. (2022).

[43] Miroshnychenko, Roldan, et al. (2025).

[44] Taylor, Mitchell, Hall, Langton, et al. (2024).

[45] Sposoby poprawy badań obserwacyjnych w tej dziedzinie można znaleźć na liście kontrolnej w publikacji Ludvigsson et al. (2023).

[46] National Institute for Health and Care Excellence (2020b, 2020a).

[47] Zgodnie z Cochrane Handbook przeglądy powinny być jak najbardziej aktualne. Jednakże, ponieważ aktualność nie jest jedynym czynnikiem brany pod uwagę, przeglądy systematyczne z nieaktualnymi strategiami wyszukiwania mogą nadal być uznane za charakteryzujące się niskim ryzykiem błędu systematycznego (J. P. T. Higgins i in., 2019).

[48] Ukierunkowane wyszukiwanie w tej dziedzinie pozwoliło zidentyfikować najnowsze publikacje oparte na badaniach obserwacyjnych przeprowadzonych w Stanach Zjednoczonych (Hidalgo i in., 2024; Olson-Kennedy i in., 2025; Wittlin i in., 2025), Kanadzie (Lawson i in., 2024) Holandii (Van Der Meulen i in., 2025) oraz Wielkiej Brytanii (Morandini i in., 2023). Badania te mają podobne ograniczenia (w tym projekt badania, mała wielkość próby, krótkotrwała obserwacja itp.) jak badania pierwotnie uwzględnione w przeglądach systematycznych.

[49] National Institute for Health and Care Research (2024).

[50] Brik i in. (2020); Carmichael i in. (2021); van der Loos, Klink i in. (2023); Wiepjes i in. (2018).

[51] Trwa debata na temat tego, czy należy kontynuować wysokiej jakości badania nad skutkami PB/HPP u nieletnich. W debacie tej uwzględniono kwestie etyczne. Niektórzy twierdzą, że badania powinny koncentrować się na pacjentach, którzy już otrzymali PB, HPP lub przeszli operację (Ryan, 2024). W przeglądzie Cass wezwano do przeprowadzenia bardziej rzetelnych badań nad psychoterapeutycznymi metodami leczenia DP, które są obiecującym podejściem terapeutycznym, dotychczas zaniedbanym w badaniach (Cass, 2024a).

[52] G. H. Guyatt, Oxman, Montori, Vist i in. (2011).

[53] Stronniczość publikacji jest niezwykle trudna do zmierzenia, a istniejące narzędzia statystyczne mają znaczne ograniczenia (Afonso i in., 2024).

[54] G. Guyatt i in. (1992, 2015).