

NEUROLOGIA

Czy na podstawie porównań pośrednich można wnioskować o skuteczności terapii w rdzeniowym zaniku mięśni (SMA)?

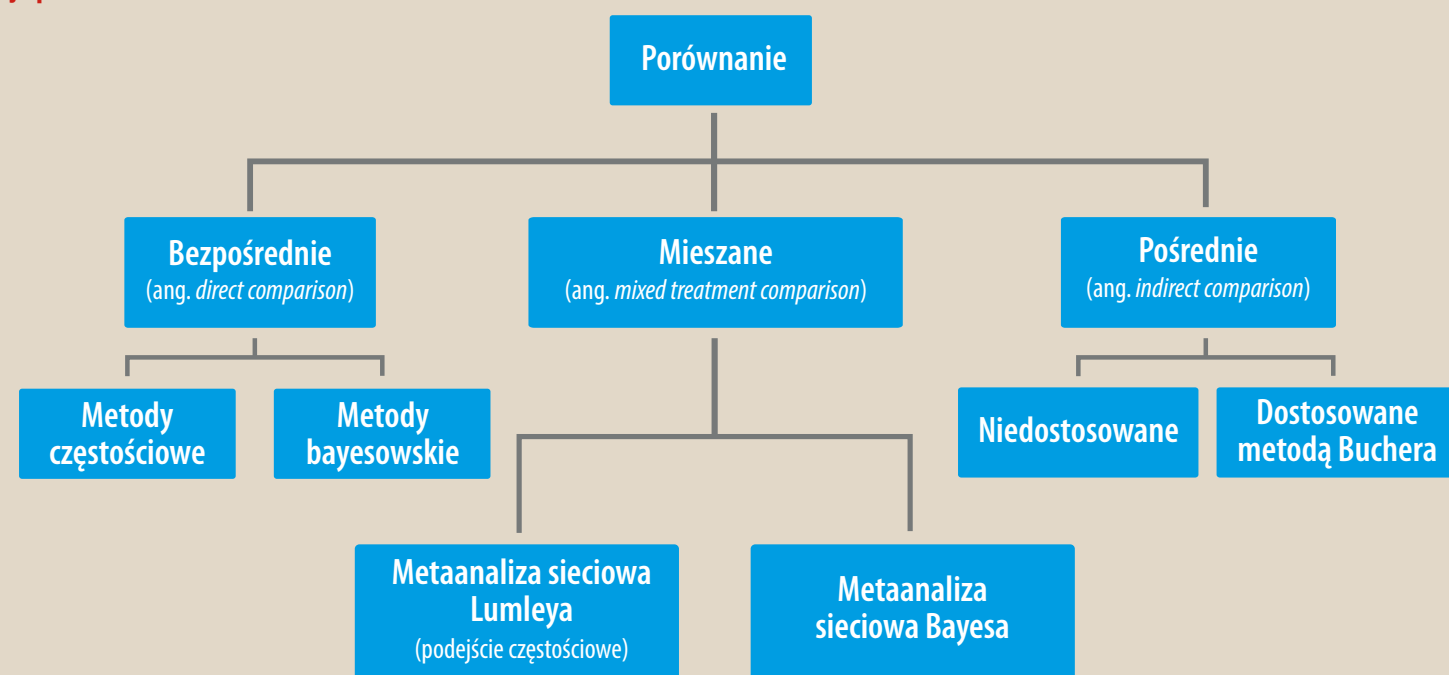
Esencją nauki – a więc i medycyny – są dowody, których jakość różni się w zależności od zastosowanej metodologii. Najwyższą wiarygodnością charakteryzują się przeglądy systematyczne randomizowanych badań klinicznych z metaanalizą. W wielu przypadkach brakuje jednak wysokiej jakości danych, a jeśli nawet są dostępne, wyciąganie wiarygodnych wniosków porównawczych jest ograniczone, gdyż najczęściej nie dotyczą porównań z aktywnym komparatorem, tylko z placebo. W takich sytuacjach wykonywane są porównania pośrednie.

Warunkiem przeprowadzenia porównania pośredniego jest dokonanie oceny heterogeniczności metodycznej i klinicznej badań włączonych do analizy. Wytyczne zalecają dokładną analizę badanej populacji, interwencji stanowiącej ramię referencyjne i punktów końcowych, a następnie przedstawienie w formie tabelarycznej zidentyfikowanych różnic.

Jeżeli heterogeniczność badań jest zbyt duża, należy poprzestać na jakościowym zestawieniu wyników. W przypadku braku możliwości przeprowadzenia porównania przez ramię referencyjne można rozważyć zastosowanie jednej z metod wskazanych przez wytyczne HTA (*health technology assessment*): proste zestawienie badań bez dostosowania (*naive comparison*), porównanie

z danymi historycznymi (*benchmarking with historical controls*), porównanie wyników badań po korekcie o różnice w charakterystyce populacji (*matching adjusted indirect comparison* – MAIC) (ryc. 1). Wytyczne HTA podkreślają, że interpretacja wyników porównań pośrednich powinna być bardzo ostrożna. Należy również wskazać ograniczenia

Rycina 1. Rodzaje porównań



Coyte PK, Shang S, Xiao Z, et al. Matching-adjusted comparisons demonstrate better clinical outcomes with SC peginterferon beta-1a every two weeks than with SC interferon beta-1a three times per week. *Mult Scler Relat Disord* 2018; 22: 134-138



Kanadyjska Agencja Oceny Leków i Technologii w Ochronie Zdrowia (Canadian Agency for Drugs and Technologies in Health)

„Ważne ograniczenia, takie jak niskie prawdopodobieństwo uwzględnienia czynników modyfikujących i prognostycznych oraz niska wiarygodność statystyczna danych dla wielu porównań, oznaczają, że rzeczywisty obserwowany efekt przypisywany rzydysplamowi jest niepewny”.

<https://cadth.ca/sites/default/files/DRR/2021/SR0661%20Evrydsi%20Recommendation%20Final.pdf>



Niemiecki Instytut ds. Jakości i Efektywności w Opiece Zdrowotnej (German Institute for Quality and Efficiency in Health Care)

„Firma nie przedstawiła wystarczającego uzasadnienia wyboru zmiennych w swojej analizie MAIC, dlatego nie można wykluczyć selektywnego raportowania. Na podstawie kryteriów wykluczenia z badań FIREFISH i ENDEAR można stwierdzić, że badana populacja badania ENDEAR miała mniej korzystne rokowania w odniesieniu do funkcji oddechowych. Miało to szczególny wpływ na interpretację wyników dotyczących wentylacji w porównaniu poszczególnych ramion różnych badań przedstawionych przez firmę”.

https://www.iqwig.de/download/a21-50_risdiplam_extract-of-dossier-assessment_v1-0.pdf



Szkockie Konsorcjum Medyczne (Scottish Medicines Consortium)

„Brakuje bezpośrednich danych porównawczych dotyczących rzydysplamu. Firma przedstawiła porównania pośrednie dotyczące pierwszego zarejestrowanego leku nusinersen w dwóch populacjach pacjentów (SMA typu 1 i typu 2/3). Kluczową wątpliwość stanowi jakość dowodów pośrednich. Wiarygodność porównania pośredniego była ograniczona przez małą liczbę pacjentów (dalsze zmniejszenie po dopasowaniu do analizy MAIC), ryzyko błędu ze względu na różnice w pewnych początkowych cechach, których nie można było dopasować, brak wspólnego ramienia kontrolnego do przeprowadzenia analizy u pacjentów z SMA typu 1, a także brak możliwości wykorzystania podstawowych (pierwszorzędowych) wyników badań dla rzydysplamu (nie oceniono ich dla nusinersenu) oraz niekompletnych (nieodrzalnych) danych dotyczących przeżycia wolnego od zdarzeń i przeżycia całkowitego”.

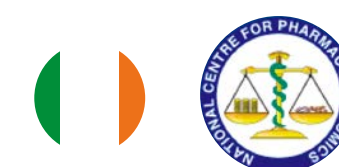
<https://www.scottishmedicines.org.uk/media/6976/risdiplam-evrydsi-final-jan-2022-amended-040722-for-website.pdf>



Duńska Agencja Leków (Danish Medicines Agency)

„Duńska Agencja Leków zdecydowała o braku akceptacji dla wyników analizy MAIC, ponieważ prowadzi ona do znacznego ograniczenia oceny i tak już bardzo małej populacji oraz istnieje co najmniej jeden istotny parametr (wsparcie funkcji układu oddechowego), który nie został wzięty pod uwagę przy tworzeniu porównania”.

https://medicinraadet.dk/media/omnif0tv/medicin%C3%A5dets-vurdering-af-risdiplam-til-spinal-muskelatrofi-vers-1-0-x_adlegacy.pdf



Narodowe Centrum Farmakoekonomiczne w Irlandii (National Centre for Pharmacoeconomics – NCPe Ireland)

„Względne efekty leczenia SMA typu 1 uzyskano z porównania interwencji między grupami pacjentów, które nie miały wspólnego ramienia kontrolnego. Porównania bez odniesienia są problematyczne ze względu na utratę randomizacji i dlatego dają wyniki, które prawdopodobnie podlegają stroniczości (błędnej interpretacji) [...]. Rezultaty są wysoce niepewne ze względu na odmienną porównywanych populacji. W związku z tym nie można określić korzyści w porównaniu do leku Spinraza lub Zolgensma w SMA typu 1”.

<https://www.ncpe.ie/wp-content/uploads/2021/03/Technical-summary-Risdiplam-HTA-ID-21006.pdf>



Norweska Agencja Leków (Statens legemiddelverk)

„Ocena wyników w skalach HINE-2 oraz CHOP-INTEND przeprowadzona przez Norweską Agencję Oceny Technologii wskazuje, że ze względu na niepewność w analizach i małe zbiory danych z krótkim czasem obserwacji nie ma podstaw do twierdzenia o statystycznie istotnym, dodatkowym działaniu rzydysplamu w porównaniu z nusinersenem.

Norweską Agencję Leków, na podstawie badań FIREFISH i ENDEAR dotyczących SMA typu 1, uważa, że rzydysplam prawdopodobnie nie ma gorszego działania niż nusinersen w krótkim okresie. Względny efekt długoterminowy nie jest znany. Nie mamy wystarczających danych, aby móc oszacować wiarygodną względną wielkość efektu lub określić, czy efekt jest równy, czy też jeden lek daje lepszy efekt niż drugi”.

[https://nyemetoder.no/Documents/Rapporter/ID2021_088_Risdiplam_Evrydsi_%20behandling%20av%205q%20spinal%20muskelatrofi%20\(SMA\)%20barn%20fra%202022%20m%20ned%20subgruppe_metodevurdering_kun%20offentlig%20versjon.pdf](https://nyemetoder.no/Documents/Rapporter/ID2021_088_Risdiplam_Evrydsi_%20behandling%20av%205q%20spinal%20muskelatrofi%20(SMA)%20barn%20fra%202022%20m%20ned%20subgruppe_metodevurdering_kun%20offentlig%20versjon.pdf)

oraz przeprowadzić analizę wrażliwości pozwalającą na ocenę wpływu poszczególnych badań (zwłaszcza tych odbiegających metodologicznie) na wynik porównania [1].

Niedostosowane porównanie pośrednie

Wśród metod niedostosowanych wyróżnia się pierwotne porównanie oszacowań punktowych i wyników dla pojedynczych ramion z niezależnych badań, tak jakby pochodziły z jednego badania. Metoda ta nie jest zalecana, ponieważ ignoruje randomizowaną naturę materiałów źródłowych. Kolejną wadą takiego podejścia jest ignorowanie mocy statystycznej oszacowań i wykluczenie wnioskowania na podstawie przedziałów ufności [1].

Dostosowanie metodą Buchera

Metodą dostosowanego porównania pośredniego jest metoda Buchera. Pierwotnie została wprowadzona w celu obliczania ilorazu szans (*odds ratio* – OR) w sytuacji, gdy chcemy porównać A vs C, podczas gdy mamy dostępne porównania bezpośrednie A vs B i B vs C. Metoda Buchera może być zastosowana tylko do badań dwuramiennych. Zakłada ona również, że względna skuteczność ocenianej interwencji jest taka sama we wszystkich badaniach włączonych do analizy, dlatego istotne jest, żeby uwzględnione badania były zbliżone pod kątem metodyki, punktów końcowych, stosowanych interwencji i charakterystyki populacji. W celu oszacowania efektu z taką samą precyzją jak w przypadku porównania bezpośredniego należy wykorzystać 4-krotnie więcej danych, co powoduje, że z matematycznego punktu widzenia jest to metoda mniej efektywna niż porównanie bezpośrednie [1].

Porównanie pośrednie z dopasowaniem populacji

Analiza MAIC jest metodą porównania pośredniego, której celem jest zniesienie heterogeniczności wynikającej z odmiennej charakterystyki populacji w badaniach włączonych do porównania pośredniego, aby uzyskać wiarygodniejsze porównanie danych interwencji. W przeciwieństwie do klasycznego porównania pośredniego wykorzystującego dane zagregowane metoda MAIC wymaga dostępu do danych poszczególnych pacjentów (*individual patient data* – IPD) z przynajmniej jednej interwencji. Na tej podstawie wykonuje się ważenie efektu, aby określić, jaki wynik powinien być uzyskany w populacji badania porównawczego [1].

Metoda MAIC może być stosowana do badań jednoramiennych lub badań bez wspólnego komparatora. W takich sytuacjach, jak interwencje chirurgiczne, choroby rzadkie, choroby nowotworowe o złym rokowaniu, zastosowanie IPD może być jedynym sposobem dostosowania różnic między próbami i powinno być preferowane w stosunku do porównań nieskorygowanych. Brak wspólnego ramienia porównawczego należy jednak odnotować jako istotne ograniczenie, ponieważ walidacja dopasowania lub zastosowanie względnych miar efektu nie będą możliwe [2].

W związku z powyższym wykonanie wiarygodnych porównań pośrednich stanowi wyzwanie i zawsze należy traktować ich wyniki z ostrożnością.



Fot. iStockphoto

Przykładem zastosowania metody MAIC jest porównanie terapii stosowanych w rdzeniowym zaniku mięśni przedstawione w publikacji Ribero i wsp. z 2022 r. [3]. Jednym z warunków właściwego porównania pośredniego jest kompatybilność porównywanych badań. Nie jest on spełniony w przypadku analizy Ribero i wsp. Autorzy porównują badanie fazy 2/3 FIREFISH (otwarte, jednoramienne) z badaniem fazy 3 ENDEAR (randomizowane, podwójnie zaślepione, kontrolowane placebo). Popelniono istotny błąd dotyczący czasu obserwacji, zestawiając przedwcześnie zakończone po 9 miesiącach badanie ENDEAR (wysokie istotne wyniki w pierwszych miesiącach leczenia) z ponad 12-miesięcznym FIREFISH i nie rozważając porównania tego ostatniego z otwartym przedłużeniem ENDEAR – badaniem SHINE. Ribero i wsp. w sposób niezgodny z przyjętymi wytycznymi zestawiają badanie wyjściowe z zawężoną populacją (FIREFISH) z badaniem włączającym szerszą populację na podstawie danych zagregowanych (ENDEAR). Czynniki, które miałyby zrównoważyć porównywane populacje, zostały ograniczone do trzech: średniego wieku przyjęcia pierwszej dawki, średniego czasu trwania choroby oraz średniego wyniku w skali CHOP-INTEND. Pomimo uwzględnienia danych jednostkowych z badania FIREFISH i próby dopasowania ich do populacji z badania ENDEAR nie uwzględniono kluczowych różnic wynikających z innego sposobu opisu populacji włączonej i kryteriów wykluczenia.

W badaniu ENDEAR 51 proc. badanych pacjentów wymagało wsparcia żywieniowego, podczas gdy w publikacji Ribero i wsp. uwzględniono jedynie 9-procentową subpopulację chorych żywionych przez zgłębnik gastrostomijny. Ponadto podczas analiz charakterystyk populacji założono

jednorodność w zakresie funkcji układu oddechowego. Tymczasem szczegółowa analiza kryteriów wykluczenia z badania FIREFISH wskazuje, że z badania zostali wykluczeni chorzy hospitalizowani w ciągu ostatnich 2 miesięcy z powodu zaburzeń oddychania, stosujący wentylację nieinwazyjną w czasie czuwania, z hipoksemią w czasie czuwania, z niewydolnością oddechową w wywiadzie lub ciężkim zapaleniem płuc i brakiem pełni sprawności wentylacyjnej w czasie badania przesiewowego. Warto podkreślić, że z badania ENDEAR takich chorych nie wykluczono. Błędne dopasowanie populacji z tych dwóch badań pod względem tak istotnych parametrów skłania ku temu, by porównania MAIC przedstawione we wspomnianej publikacji uznać za mało wiarygodne.

Powyżej analizowane różnice w błędnie dopasowanych populacjach oraz jakość analizy MAIC zostały również skomentowane przez agencje oceny technologii medycznych na całym świecie.

Opracowanie: *Damian Matusiak, Małgorzata Solarczyk*

Piśmiennictwo

1. Szwarz N, Kaczorek-Juskiewicz A, Kalbarczyk A. Porównania pośrednie w analizach klinicznych ocenianych przez AOTMiT – przegląd i opis zastosowanych metod. Agencja Oceny Technologii Medycznych i Taryfikacji, Warszawa 2019.
2. Signorovitch JE, Sikirica V, Erder HM i wsp. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012; 15: 940-947.
3. Ribero VA, Daigl M, Mart Y i wsp. How does risdiplam compare with other treatments for types 1–3 spinal muscular atrophy: a systematic literature review and indirect treatment comparison. *J Comp Eff Res* 2022; 11: 347-370.