

Stosowanie zasad medycyny opartej na dowodach naukowych w neurologii

Janis M. Miyasaki, MD, MEd, FRCPC, FAAN

The Morton and Gloria Shulman
Movement Disorders Centre,
Toronto Western Hospital,
University of Toronto, Toronto,
Kanada

Adres do korespondencji:
The Morton and Gloria Shulman
Movement Disorders Centre,
Toronto Western Hospital,
University of Toronto, 399
Bathurst Street, 7 McL, Toronto,
ON M5T 2S8, Canada

e-mail: miyasaki@uhnresearch.ca

Neurol Clin 28 (2010) 489–503

Neurologia po Dyplomie
2012; 7 (2): 48-59

SŁOWA KLUCZOWE: neurologia, medycyna oparta na dowodach naukowych, krytyczna ocena, wytyczne

Tło historyczne

Według najnowszego sondażu przeprowadzonego przez organizację Kaiser Health dla polityków kandydujących na urząd prezydencki od czasu wojny w Iraku Amerykanie zaczęli się bardziej interesować problemem opieki zdrowotnej.¹ W tej dziedzinie głosujący wyrazili potrzebę rozszerzenia zakresu ubezpieczenia zdrowotnego i zmniejszenia kosztów opieki zdrowotnej. Może to sugerować, że Amerykanie nie chcą zwiększenia ilości, ale raczej poprawy jakości w tym zakresie. Według raportu Agency for Healthcare Quality and Research jakość i sprawne funkcjonowanie opieki zdrowotnej oznacza „zapewnienie właściwej opieki odpowiedniemu pacjentowi we właściwym czasie i w odpowiedni sposób za pierwszym razem.”² Ponadto zgodnie z deklaracją misją agencji jest prowadzenie i wspieranie badań usług zdrowotnych w celu „zmniejszenia ryzyka doznania szkody na skutek działań ochrony zdrowia przez wykorzystywanie badań i technologii opartych na dowodach, aby zapewnić możliwie najlepszą opiekę, zastosowanie wyników badań naukowych w praktyce do uzyskania lepszego dostępu do skutecznych usług zdrowotnych i zmniejszenia niepotrzebnych kosztów związanych z opieką zdrowotną, a także poprawa wyników opieki zdrowotnej przez zachęcanie dostawców, odbiorców i pacjentów do wykorzystywania informacji opartych na dowodach przy podejmowaniu świadomych wyborów lub decyzji terapeutycznych.”

Medycyna oparta na dowodach naukowych (evidence based medicine, EBM) jest starym pojęciem, które obecnie zyskało na aktualności. Skuteczność w opiece zdrowotnej stanowi długo pielęgnowany cel, którego początki datuje się na wiek XIX. W ostatnich czasach głęboki wpływ na badania naukowe w dziedzinie medycyny wywarł Archibald Cochrane, publikując w 1971 roku książkę pod tytułem „Skuteczność i sprawność: Uwagi na temat służby zdrowia”.³ Cochrane wyszedł z założenia, że zasoby zawsze są niewystarczające, dlatego przy zapewnianiu opieki zdrowotnej należy wykorzystywać najpewniejsze dowody. W szczególności zalecał stosowanie wyników randomizowanych badań kontrolowanych (randomized controlled trials, RCT) jako prawdopodobnie najmniej obciążonych błędem wnioskowania.

Głównym nieporozumieniem w związku z zasadami medycyny opartej na dowodach naukowych jest to, że traktują one randomizowane badania kontrolowane jako jedyne wiarygodne źródło dowodu. Jest to być może podtrzymywane przez fakt że w swoich raportach organizacja Cochrane Collaborations opiera się wyłącznie na wynikach RCT. Jednak Guyatt i wsp.,⁴ którzy jako pierwsi użyli terminu medycyna oparta na dowodach naukowych, zdefiniowali ją jako

sumienne, jawne i rozważne wykorzystywanie najlepiej dostępnych dowodów przy podejmowaniu decyzji dotyczących opieki nad pacjentami. W związku z tym inne rodzaje dowodów również stanowią wskazówkę w praktyce.

Na początek należy podkreślić, że EBM nie jest książką kucharską medycyny, nie usiłuje ograniczać autonomii lekarza, nie stanowi narzędzia ataku na sali rozpraw sądowych (podobnie jak narzędzia obrony), nie ogranicza się do wyników RCT i nie lekceważy doświadczenia klinicznego.

Krytycy drwią z założeń EBM jako książki kucharskiej medycyny – wysiłku mającego na celu ujednoczenie leczenia i odejścia od sztuki medycznej. Niechęć ta bierze początek z wczesnej historii Ameryki, gdy doświadczeni lekarze indywidualizowali opiekę medyczną, a podróżujący handlowcy dostarczali mikstury na wszystkie choroby. Ci ostatni byli sprzedawcami panaceum, podczas gdy pierwsi byli lekarzami dżentelmenami. Współczesna kultura również docenia indywidualność, zwłaszcza niezależność. Zasady EBM kwestionują postępowanie tych bohaterów, ujawniając, porządkując i demokratyzując wiedzę. Zamiast utrzymywać wiedzę w rękach nielicznej elitarnej grupy ekspertów, zwolennicy zasad EBM umożliwiają nawet nowicjuszom dostęp do najlepszych dowodów. Dowód zazwyczaj odnosi się do uczestników badań klinicznych – idealnych pacjentów z objawami małej grupy chorób, które mogą być tylko nieznacznie podobne do stwierdzanych u pacjentów zgłaszających się do gabinetów lekarskich. Nawet wytyczne uwzględniające wiele informacji nie zapewniają prostych algorytmów w każdej sytuacji. Większość wytycznych podkreśla, że zalecenia nie powinny zastępować oceny klinicznej i podejmowania decyzji przez klinicystów dotyczących konkretnych pacjentów. Dlatego w opiece nad pacjentem ocena kliniczna pozostaje istotną częścią podejmowania decyzji.

Wytyczne, główne narzędzie EBM, mogą być traktowane jako standard w praktyce lekarskiej. W sądzie uważa się jednak, że jest to tylko jeden standard postępowania praktycznego – mogą istnieć różne wytyczne dotyczące danego tematu i stąd różne standardy. Innym standardem jest praktyka kliniczna. Kolejnym ograniczeniem jest fakt, że większość wytycznych przestaje być aktualna w ciągu dwóch lat od publikacji. Dlatego najnowsze informacje mogą nie zostać uwzględnione w wytycznych. Ignorowanie wytycznych nie jest rozsądne, ale bezkrytyczne postępowanie według nich również nie służy pacjentom. Najlepszą obroną w praktyce klinicznej jest świadomość istnienia i znajomość uaktualnianych danych oraz uwzględnianie preferencji pacjenta.

Czym więc jest medycyna oparta na dowodach naukowych? Jest to praktyka rozpoznawania potrzeb w zakresie informacji i formułowania pytań, na które daje się odpowiedzieć, przeszukiwania piśmiennictwa w celu znalezienia najlepszego dowodu, krytycznej oceny wiarygodności dowodów, ich zastosowania i oceny wyników. Wymaga to wielu umiejętności: formułowania dobrych pytań (co jest trudniejsze niż się wydaje), przeszukiwania piśmiennictwa i jego krytycznej

Ramka 1. Kryteria jakości artykułu diagnostycznego według American Academy of Neurology

Klasa I: Dowód przedstawiony na podstawie wyników badania prospektywnego z udziałem szerokiego spektrum osób z podejrzeniem danej choroby, stosujących referencyjny (złoty) standard dla definicji przypadku, z zastosowaniem zaślepionej ewaluacji testu, umożliwiającej ocenę odpowiednich testów pod kątem dokładności diagnostycznej. Wszyscy pacjenci poddani testowi diagnostycznemu są lub nie są chorzy.

Klasa II: Dowód przedstawiony na podstawie wyników badania prospektywnego z udziałem wąskiego spektrum osób z podejrzeniem danej choroby lub z wykorzystaniem dobrze zaprojektowanego badania retrospektywnego z udziałem szerokiego spektrum osób z daną chorobą (rozpoznaną za pomocą metody będącej złotym standardem), porównanych z szerokim spektrum osób z grupy kontrolnej z zastosowaniem zaślepionej ewaluacji testu umożliwiającej ocenę odpowiednich testów pod kątem dokładności diagnostycznej.

Klasa III: Dowód przedstawiony na podstawie wyników badania retrospektywnego z udziałem wąskiego spektrum osób z daną chorobą lub z grupy kontrolnej, w którym standard referencyjny przy braku jego obiektywności stosowany jest przez inną osobę niż ta, która wykonała test.

Klasa IV: Jakikolwiek projekt badania, w którym nie zastosowano niezależnej oceny testu lub dowód przedstawiony został jedynie na podstawie opinii eksperta lub opisowego badania serii przypadków (bez grupy kontrolnej).

Edlund W, Gronseth G, So Y, Franklin G. AAN Clinical Practice Guideline Process Manual. St. Paul: American Academy of Neurology. 2005; za zgodą.

oceny, wydobywania przydatnych informacji oraz analizowania wyników po zastosowaniu tych informacji w praktyce. Dlatego biorąc pod uwagę wymagany czas i wysiłek, najbardziej odpowiednie dla tej metody są główne kwestie kliniczne.

Przykład roboczy

W artykule opisano hipotetyczny przypadek pacjentki, na przykładzie którego przedstawiono kolejne kroki postępowania według zasad medycyny opartej na dowodach naukowych. Doktor Jones przyjmuje 55-letnią panią Smith z chorobą Parkinsona. Od około roku pacjentka zauważyła drżenie spoczynkowe lewej ręki. Badanie neurologiczne ujawnia wzmoczone napięcie mięśni o typie koła zębatego w lewej kończynie górnej i bradykinezę. Zwraca również uwagę brak balansowania lewą kończyną górną bez innych nieprawidłowości. Dr Jones potwierdza rozpoznanie parkinsonizmu, najprawdopodobniej w przebiegu choroby Parkinsona. Pani Smith pyta lekarza, jak bardzo jest pewny swojej diagnozy.

W celu uzyskania pomocy dr Jones sięga po piśmiennictwo i zadaje pytanie: „jak często rozpoznanie choroby Parkinsona u pacjenta (populacja pacjenta) zostaje potwierdzone (trafność diagnozy) w badaniu autopsyjnym (złoty standard)?” Dr Jones chce sprecyzować swoje pytanie przez dodatkowe wprowadzenie ograniczeń poszukiwania. Jakie kryteria dotyczące projektu badania powinien narzucić?

Dr Jones przeszukuje stronę internetową Amerykańskiej Akademii Neurologii (American Academy of Neurology, AAN) w celu znalezienia Podręcznika Wytycznych Dotyczących Praktyki Klinicznej (Clinical Practice Guideline Process Manual) i znajduje pewne kryteria (wymienione w ramce 1).⁵

Kwestia trafności rozpoznania jest złożona, ponieważ w przypadku choroby Parkinsona nie istnieje odpowiedni test diagnostyczny, a złotym standardem nadal jest badanie autopsyjne. Dr Jones wykorzystuje kryteria rozpoznania choroby Parkinsona, które do postawienia diagnozy, wymagają stwierdzenia dwóch z trzech objawów: drżenie, bradykinezja i sztywność. Znajduje badanie retrospektywne (klasa III), w którym histopatologiczne potwierdzenie rozpoznania stanowiło złoty standard diagnostyczny, a kryteria kliniczne odpowiadają kryteriom przez niego stosowanym.⁶ Wyniki badania wskazują na 99% czułość, ale tylko 8% swoistość tych kryteriów. Dodatnia wartość predykcyjna wynosi 77%, a ujemna – 67%.

$$\text{Czułość} = \frac{\text{Liczba wyników prawdziwie dodatnich}}{\text{Liczba wyników prawdziwie dodatnich} + \text{liczba wyników fałszywie dodatnich}}$$

Czułość testu odpowiada na pytanie: „Zakładając obecność choroby, jakie jest prawdopodobieństwo dodatniego wyniku testu?” Swoistość określa zdolność do wykluczenia choroby i udziela odpowiedzi na pytanie: „Zakładając brak choroby, jakie jest prawdopodobieństwo, że wynik testu będzie ujemny?” Zwykle uważa się, że kryteria diagnostyczne powinny charakteryzować się dużą swoistością. W diagnozowaniu choroby Parkinsona większość chorób należy do grupy zespołów parkinsonowskich, dlatego u pacjentów z tymi rozpoznaniami należy podjąć próbę leczenia lewodopą. Kryteria diagnostyczne mogą być w tym przypadku bardziej czułe, ponieważ zastosowanie leczenia jest właściwym postępowaniem mimo braku potwierdzenia choroby Parkinsona w badaniu autopsyjnym.

Następnie dr Jones sprawdza dodatnią wartość predykcyjną, ponieważ wskaźnik ten ocenia prawdopodobieństwo choroby przy dodatnim wyniku testu. Z kolei ujemna wartość predykcyjna wskazuje, jakie jest prawdopodobieństwo braku choroby przy ujemnym wyniku testu. W kategoriach liczbowych:

$$\text{Dodatnia wartość predykcyjna} = \frac{\text{Liczba wyników prawdziwie dodatnich}}{\text{Liczba wyników prawdziwie dodatnich} + \text{liczba wyników fałszywie dodatnich}}$$

$$\text{Ujemna wartość predykcyjna} = \frac{\text{Liczba wyników prawdziwie ujemnych}}{\text{Liczba wyników prawdziwie ujemnych} + \text{liczba wyników fałszywie ujemnych}}$$

Gdyby dr Jones miał dostęp do oryginalnych danych, chciałby znać trafność diagnostyczną (skuteczność) testu lub odsetek prawidłowych wyników, prawdziwie dodatnich i prawdziwie ujemnych.

$$\text{Trafność diagnostyczna (skuteczność)} = \frac{\text{Liczba wyników prawdziwie dodatnich} + \text{liczba wyników prawdziwie ujemnych}}{\text{Liczba wszystkich wyników testu}}$$

Ci sami badacze zastosowali kryteria rozpoznania choroby Parkinsona (wymagające obecności wszystkich trzech objawów [drżenie, bradykinezja i sztywność]), wykazując ich 65% czułość i 71% swoistość, przy 88% dodatniej i 40% ujemnej wartości predykcyjnej.

Dr Jones preferuje stosowanie kryteriów o dużej czułości i swoistości. W dalszej części artykułu omówiono zastosowanie kryteriów dotyczących asymetrii objawów i braku cech atypowych. Przy uwzględnieniu tych parametrów zarówno czułość, jak i swoistość rosną do 75%, podczas gdy dodatnia wartość predykcyjna wynosi 93%, a ujemna – 49%. Na podstawie tych danych dr Jones wnioskuje, że jest na 93% pewny, że pacjentka cierpi na chorobę Parkinsona. Niewłaściwe rozpoznania dotyczą najczęściej zaniku wieloukładowego, postępującego porażenia nadjądrowego, otępienia z ciałami Lewy’ego i innych zespołów parkinsonowskich.

Na kolejnej wizycie pani Smith zgłasza zaniepokojenie o swoją przyszłość. Pyta lekarza: „Jakie jest ryzyko wystąpienia u mnie otępienia związanego z chorobą Parkinsona? Jak szybko choroba będzie postępować?”

Dr Jones przeszukuje piśmiennictwo w celu uzyskania odpowiedzi na pytanie: „Jakie jest ryzyko rozwoju otępienia u pacjenta z drżeniem w przebiegu choroby Parkinsona?” To pokazuje, jak formułować pytanie przy przeszukiwaniu piśmiennictwa. Pomocny jest przy tym akronim PICO: dana populacja pacjentów (Patient), interwencja (Intervention) (w tym przypadku czynnik prognostyczny), kontrola (Control) i wynik leczenia (Outcome). Dr Jones pamięta, że najbardziej wiarygodna (najmniej obciążona błędem wnioskowania) informacja dotycząca terapii pochodzi z RCT z obiektywnie mierzonymi skutkami zdrowotnymi lub z podwójnie ślepa próbą. Zastanawia się nad zastosowaniem porównywalnych kryteriów dla dobrego artykułu prognostycznego. Próbuje zminimalizować ryzyko obciążenia błędem wnioskowania, dr Jones stwierdza, że artykuł prognostyczny powinien dotyczyć szerokiego spektrum pacjentów, poddanych obserwacji pod względem danego skutku zdrowotnego przez z góry określony czas. Skutek zdrowotny

Ramka 2. Klasyfikacja dowodów prognostycznych według American Academy of Neurology

Klasa I: Dowód przedstawiony na podstawie wyników badania prospektywnego z udziałem szerokiego spektrum osób potencjalnie zagrożonych wystąpieniem danej choroby. Badanie ocenia zdolność prognostyczną przy zastosowaniu niezależnego złotego standardu dla definicji przypadku. Czynniki predykcyjne jest oceniany przy zaślepieniu pod względem objawów klinicznych, a wynik jest oceniany przy zaślepieniu na obecność czynnika predykcyjnego. U wszystkich pacjentów ocenia się obecność czynnika predykcyjnego i zmienne dotyczące wyniku.

Klasa II: Dowód przedstawiony na podstawie wyników badania prospektywnego z udziałem wąskiego spektrum osób zagrożonych wystąpieniem danej choroby lub na podstawie wyników badania retrospektywnego z udziałem szerokiego spektrum osób z daną chorobą porównanych z szerokim spektrum osób z grupy kontrolnej. Badanie ocenia dokładność prognostyczną czynnika ryzyka przy wykorzystaniu akceptowalnego niezależnego złotego standardu dla definicji przypadku. Czynniki ryzyka jest oceniany przy zaślepieniu na wyniki.

Klasa III: Dowód przedstawiony z wykorzystaniem wyników badania retrospektywnego z udziałem wąskiego spektrum osób z daną chorobą lub z grupy kontrolnej. Badanie ocenia zdolność prognostyczną przy wykorzystaniu akceptowalnego niezależnego złotego standardu dla definicji przypadku. Wynik, jeśli nie jest obiektywny, jest określany przez inną osobę niż ta, która oceniała czynnik ryzyka.

Erdlund W, Gronseth G, So Y, Franklin G. AAN Clinical Practice Guideline Process Manual. St. Paul: American Academy of Neurology. 2005; za zgodą.

powinien być obiektywny lub przynajmniej mierzony przez badacza zaślepionego pod kątem obecności lub braku czynnika rokowniczego.

W celu udzielenia pani Smith odpowiedzi na temat rokowania dr Jones poszukuje badań, które dotyczą wszystkich pacjentów z chorobą Parkinsona z dominującym drżeniem lub sztywnością akinytyczną bez drżenia, oceniane za pomocą skali UPDRS (Unified Parkinson's Disease Rating Scale). Progresa zaburzeń ruchowych powinna być oceniana przez badacza zaślepionego pod kątem przewagi drżenia lub sztywności kinetycznej. Mógłby również zaakceptować wyniki badania prospektywnego z udziałem pacjentów z dominującym drżeniem, u których ocena progresji choroby w czasie była prowadzona przez zaślepionego badacza. Na ostatnim spotkaniu z zarządcami szpitala dowiedział się o metodzie stopniowania dowodów według czterech A: dostępny (accessible), dokładny (accurate), przydatny (applicable) i dający wyrazić się w działaniu (actionable).⁷ Chociaż łatwa do zapamiętania, metoda ta nie pomaga lekarzowi w stopniowaniu jakości dowodu na tyle, aby mógł on na tej podstawie wydać zalecenia swojej pacjentce.

Dr Jones sprawdza kryteria dla badań prognostycznych umieszczone na stronie internetowej ANN (ramka 2).⁵ Stwierdza, że wyszukane przez niego artykuły charakteryzują się niską jakością. Brakuje w nich kryteriów włączenia, danych na temat czasu obserwacji, zaślepienia badacza w odniesieniu do istotnych czynników prognostycznych, a także przejrzystych wykresów, które umożliwiłyby obliczenia.

Istotne z punktu widzenia dr. Jonesa jest pytanie: „Biorąc pod uwagę określoną sytuację, jakie jest ryzyko progresji zaburzeń chodu z powodu zaburzeń równowagi?” Może on obliczyć bezwzględne zmniejszenie (lub wzrost) ryzyka osiągnięcia np. w ciągu 5 lat 3 stadium niesprawności w skali Hoehn-Yahra (obecność obustronnych objawów i niestabilności posturalnej) przez pacjenta z drżeniem w porównaniu z chorym ze sztywnością akinytyczną bez drżenia. Chciałby także przedstawić pani Smith bezwzględne zmniejszenie (lub wzrost) ryzyka ośpienia w przebiegu choroby u pacjentów z dominującym drżeniem w porównaniu z chorymi ze sztywnością akinytyczną bez drżenia.

$$\text{Bezwzględne zmniejszenie ryzyka (\%)} = \frac{\text{Współczynnik zdarzeń w grupie kontrolnej} - \text{współczynnik zdarzeń w grupie leczonej}}{\text{Współczynnik zdarzeń w grupie kontrolnej}}$$

$$\text{Względne zmniejszenie ryzyka (\%)} = \frac{\text{Współczynnik zdarzeń w grupie kontrolnej} - \text{współczynnik zdarzeń w grupie leczonej}}{\text{Współczynnik zdarzeń w grupie kontrolnej}}$$

Ponieważ względne zmniejszenie ryzyka odnosi się do grupy kontrolnej, może ono wydawać się imponujące w stosunku do niewielkiego zysku. Na przykład w badaniu z udziałem 500 pacjentów z omdleniami leczenie okazało się istotnie lepsze niż stosowanie placebo. W grupie osób otrzymujących placebo częstość występowania omdleń wynosiła 4%, a w grupie leczonej 3%. Bezwzględne zmniejszenie ryzyka wystąpienia omdlenia u osób z grupy leczonej wyniosło 1%, ale względne zmniejszenie ryzyka – 25% – liczba bardziej imponująca, ale myląca.

Na przykład dla osób z omdleniem:

$$\text{NNT} = \frac{1}{\text{Bezwzględne zmniejszenie ryzyka}} = \frac{1}{0,04 - 0,03} = 100$$

Wysoka NNT (number needed to treat – liczba pacjentów, których trzeba poddać danej interwencji przez określony czas, aby zapobiec jednemu niekorzystnemu punktowi końcowemu – przyp. tłum.) mówi dr. Jonesowi, że wynik leczenia omdlenia jest niewielki.

Dr Jones przypomina sobie, że *p* (istotność) również może być myląca. Zdaje sobie sprawę, że *p* równe 0,05 oznacza, że prawdopodobieństwo, iż uzyskane wyniki są dziełem

przypadku, a nie rzeczywistą różnicą między brakiem leczenia a wynikiem eksperymentalnej terapii wynosi 1 na 20. Mniejsze p nie oznacza, że wyniki są bardziej znaczące – p odzwierciedla ryzyko, że uzyskane wyniki lub różnica między grupą kontrolną a eksperymentalną są wynikiem przypadku, a nie rzeczywistym skutkiem leczenia. Mniejsze p oznacza mniejsze ryzyko przypadkowości wyników i większe prawdopodobieństwo związku z terapią.

Dr Jones zastanawia się, jaki wskaźnik lepiej odzwierciedla dokładność wyników badania. Analizuje więc definicję przedziału ufności (confidence interval, CI). Dowiaduje się, że CI określa dolną i górną granicę przedziału, w którym rzeczywisty wynik leczenia znajduje się z 95% pewnością. Innymi słowy CI ocenia ilościowo wynik, a także niepewność związaną z tym wynikiem.⁸ Ponadto dr Jones dowiaduje się, że badania obciążone błędami systematycznymi i charakteryzujące się dużą liczbą uczestników mogą mieć wąski CI. Przed określeniem CI dr Jones musi najpierw ocenić jakość badania.

Przy obliczaniu przedziału ufności użyte dane i równania zależą od danych analizowanych. Do obliczenia CI czytelnik może użyć kalkulatorów statystycznych dostępnych na stronach internetowych.

Dr Jones zauważa, że CI zależy głównie od liczby uczestników badania. Na przykład, jeśli u 25% pacjentów z drżeniem w ciągu 5 lat dochodzi na postępu choroby i zaburzeń równowagi, a liczba obserwowanych pacjentów z drżeniem jako objawem dominującym wynosi 200, wówczas 95% CI wynosi 19-31. Oznacza to, że istnieje 95% pewność, że odsetek pacjentów z dominującym drżeniem, u których w ciągu 5 lat wystąpią zaburzenia równowagi znajduje się w przedziale od 19 do 31%. Stąd wniosek, że drżenie jest istotnym objawem klinicznym, a niepewność co do tego jest mała.

Jeśli liczba uczestników byłby dużo mniejsza – powiedzmy 20 – 95% CI wyniosłoby 8-42%, co oznacza, że wynik jest mniej precyzyjny.

Mniejsza liczba uczestników próby oznacza, że 95% CI znajduje się w przedziale od 15 do 35, co sprawia, że wynik jest mniej dokładny.

Dr Jones ponownie analizuje dostępne dowody na temat drżenia i progresji choroby. Z aktualnego przeglądu dowiaduje się, że istnieje silny dowód na obecność dodatniej zależności między późniejszym wiekiem zachorowania (>57 roku życia) a większą niesprawnością, jak również, że pacjenci z dominującymi objawami osiowymi, ocenianymi według złożonej skali punktowej zaburzeń chodu i równowagi, charakteryzują się większą niesprawnością w porównaniu z pacjentami, u których dominuje drżenie. Ze względu na brak danych, lekarz nie jest w stanie udzielić pani Smith dalszych informacji.⁹ Dr Jones mówi pani Smith, że u pacjentów z dominującym drżeniem oraz tych, którzy zachorowali w młodszym wieku, rokowanie jest lepsze niż u chorych bez drżenia lub osób, u których pierwsze objawy wystąpiły po 57 roku życia.

Po 2 latach pani Smith ponownie zgłasza się na wizytę i skarży się na nasilenie niesprawności. Wcześniej do

złagodzenia objawów stosowała ćwiczenia, teraz jednak prosi o przepisanie leku na poprawę funkcji ruchowych. Pani Smith twierdzi, że pisanie na tablicy stało się dla niej uciążliwe i wymaga więcej wysiłku. Ocenianie prac studentów zajmuje jej dużo czasu, musi więc wyjeżdżać do pracy o godzinę wcześniej, żeby móc się do niej przygotować.

Dr Jones przypomina sobie badania porównujące skuteczność lewodopy i agonistów receptora dopaminowego w leczeniu ruchowych objawów choroby Parkinsona. Stawia następujące pytanie: „Który z leków, lewodopa czy agonista receptora dopaminowego, jest skuteczniejszy w leczeniu niesprawności ruchowej u pacjentów z chorobą Parkinsona, dotychczas nieleczonych?” Przeszukując piśmiennictwo znajduje kilka artykułów i wytyczne dotyczące rozpoczęcia leczenia w chorobie Parkinsona. Analizując wytyczne, zauważa, że uwzględniono w nich wyniki prawie wszystkich artykułów dotyczących porównania skuteczności lewodopy z agonistami receptora dla dopaminy. Sprawdzając datę publikacji wytycznych, zastanawia się nad ich wartością, ponieważ zostały one opublikowane w 2002 roku.¹⁰ Dr Jones przegląda stronę internetową AAN i stwierdza, że wytyczne te zostały ponownie zatwierdzone 15 października 2005 roku.¹¹

Analizując metodologię opracowywania wytycznych, dr Jones stwierdza, że charakteryzują się one różną dokładnością. Co więcej, okazuje się, że istnieją różne kryteria oceny jakości wytycznych. Analizując kryteria obowiązujące w bazie danych National Guideline Clearinghouse. Wszystkie przedstawione na tej stronie internetowej wytyczne muszą spełniać cztery kryteria: (1) muszą być poparte przez systematycznie opracowywane stanowisko, obejmujące zalecenia, strategie postępowania lub informację, (2) muszą być opracowane przez medyczne stowarzyszenie specjalistyczne, grupę specjalistów, publiczną lub prywatną organizację, agencję rządową na poziomie federalnym, stanowym lub lokalnym, bądź organizację ochrony zdrowia oraz (3) muszą umożliwić stworzenie dokumentacji i weryfikację, czy dokonano przeglądu systematycznego piśmiennictwa oraz istniejących dowodów naukowych, opublikowanych w recenzowanych czasopiśmie i (4) muszą być przedstawione w pełnej wersji tekstowej, dostępnej na życzenie w formie drukowanej lub elektronicznej, w języku angielskim oraz opracowanej i zweryfikowanej w ciągu ostatnich 5 lat.

Dr Jones odnosi wrażenie, że kryteria te nie są wystarczająco rygorystyczne, aby na ich podstawie nie zalecić pacjentce leczenie. Zapoznaje się z instrumentem AGREE (Appraisal of Guidelines for Research and Evaluation), opracowanym przez AGREE Collaboration.¹² Instrument AGREE składa się z 23 pozycji zgrupowanych w sześciu domenach: zakres i cel, udział zainteresowanych stron, rzetelność procesu opracowywania wytycznych, przejrzystość i sposób przedstawienia, możliwość praktycznego zastosowania i niezależność redakcyjna. Każda pozycja stopniowana jest w czteropunktowej skali (ramka 3). Dr Jones stwierdza, że stosowanie tego instrumentu więcej niż raz jest zbyt pracochłonne. Czyta

Ramka 3. Instrument AGREE

Zakres i cel

1. Ogólne cele, pytania kliniczne i populacja pacjentów, do której odnoszą się wytyczne, są szczegółowo opisane.

Udział osób zainteresowanych

2. Zespół opracowujący wytyczne obejmuje przedstawicieli wszystkich grup zawodowych, określa jasno docelowych użytkowników, umożliwia wypróbowanie wytycznych przez przyszłych użytkowników i uwzględni punkt widzenia i preferencje pacjentów.

Rzetelność opracowywania wytycznych

3. 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.

4. Istnieje jasna metoda uaktualniania wytycznych.

Przejrzystość i sposób przedstawienia

5. Zalecenia są szczegółowe i jednoznaczne, przedstawiają też różne opcje leczenia. Udostępnione są narzędzia ułatwiające korzystanie z wytycznych.

Użyteczność

6. Omówiono potencjalne bariery zastosowania zaleceń, w tym kwestie finansowe.
7. W wytycznych proponuje się kroki zmierzające do poprawy jakości, powiązane z wnioskami.

Niezależność redakcyjna

8. Pod względem redakcyjnym wytyczne są niezależne od instytucji finansujących, a każdy potencjalny konflikt interesów został ujawniony.

Na podstawie: *The Agree Collaboration. Appraisal of guidelines for research and evaluation: AGREE instrument. Dostępne na stronie: www.agreecollaboration.org.*

Ramka 4. Krytyka grupy roboczej GRADE dotycząca istniejących systemów

1. Czy metodę (procedurę) można zastosować w różnych sytuacjach klinicznych? Skuteczność, działania niepożądane, rozpoznanie, rokowanie.
2. Czy produkt końcowy jest przydatny dla pacjenta, lekarzy i decydentów?
3. Czy metoda jest jasna i logiczna?
4. Jaki jest stopień subiektywności analizy lub opracowania?
5. Czy został sprecyzowany poziom dowodu lub siła zaleceń?
6. Czy zostały pominięte ważne aspekty?
7. Czy system jest skuteczny w rozróżnianiu dużego i małego poziomu dowodu lub dużej i małej siły zaleceń?
8. Czy ocena jest odtwarzalna?

Na podstawie: *Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Services Research 2004; 4: 38-45.*

Dr Jones potrzebuje bardziej szczegółowych informacji. Chce zrozumieć stopień zwiększenia korzyści lub redukcji ryzyka przy zastosowaniu różnych leków. Dr Jones znajduje jeden z oryginalnych artykułów.¹⁴ Przedstawione w nim wyniki badania wykazały, że ryzyko rozwoju dyskinez po 5 latach u pacjentów, u których leczenie rozpoczynano od włączenia lewodopy (lek kontrolny) wynosiło 31%. U osób, które rozpoczęły terapię pramipeksolem ryzyko to wynosiło 10%. Dr Jones przypomina sobie, że NNT oznacza liczbę pacjentów, których należałoby leczyć agonistą receptora dopaminowego, aby zapobiec wystąpieniu dyskinez u jednego z nich. Liczbę tę oblicza się według wzoru:

$$NNT = \frac{1}{\text{Bezwzględne zmniejszenie ryzyka}} = \frac{1}{\% \text{ osób z dyskinezami leczonych lewodopą} - \% \text{ osób z dyskinezami leczonych pramipeksolem}} = \frac{1}{31\% - 10\%} = 5$$

W związku z tym dr Jones mówi pani Smith, że bezwzględne zmniejszenie ryzyka przy zastosowaniu pramipeksolu wyniesie 21%, czyli prawdopodobieństwo rozwoju u niej dyskinez przy rozpoczęciu leczenia od pramipeksolu, wynosi 10%, a przy lewodopie - 31%. Dr Jones musiałby rozpocząć leczenie od pramipeksolu u pięciu pacjentów, aby uniknąć rozwinięcia się dyskinez w ciągu 5 lat u jednego z nich.

artykuł dotyczący systemów stopniowania wytycznych¹³ i znajduje listę pytań (ramka 4), służących do tego celu.

Dr Jones stwierdza, że wytyczne opracowane przez AAN spełniają większość tych kryteriów, co wskazuje na to, że procedura ta ma logiczne uzasadnienie, a wnioski są przekonujące.

Analizując wnioski, dr Jones upewnia się o wysokiej jakości dowodu uzasadniającego stosowanie lewodopy lub agonistów receptora dopaminowego jako początkowej terapii niesprawności ruchowej. Wybór zależy od indywidualnego podejścia do pacjenta, ponieważ jej stosowanie przynosi większe korzyści pod względem funkcji ruchowych, ale wiąże się z większym ryzykiem wystąpienia dyskinez, a w przypadku agonistów receptora dopaminowego ryzyko to jest znacznie mniejsze.

Po zapoznaniu się z tą informacją pacjentka pyta o leczenie neuroprotektoryjne stosowane w chorobie Parkinsona. Przeszukując bazę danych PubMed, dr Jones dowiaduje się o kilku badaniach, dotyczących neuroprotekcji w chorobie Parkinsona i jednych wytycznych, opublikowanych znów przez AAN.¹⁵ Już wcześniej dokonał oceny wytycznych AAN i wie, że spełniają one większość kryteriów metody dobrej jakości. Wykorzystuje więc te wytyczne jako podstawową informację i formułuje następujące pytania: „Czy stosowanie selegiliny lub rasagiliny zwalnia postęp choroby Parkinsona?” „Czy istnieją inne metody leczenia, zwalniające postęp choroby Parkinsona?”

Analizując wytyczne, dr Jones wyciąga wniosek, że lewodopa wywiera prawdopodobnie wpływ neuroprotektoryjny przez 9 miesięcy, z kolei stosowanie rasagiliny może przynieść korzyści związane z działaniem neuroprotektoryjnym, ale w celu dokładnej oceny tego wpływu konieczne jest przeprowadzenie długoterminowych badań. Na podstawie wcześniejszej analizy wywnioskowano, że selegilina przynosi korzyści kliniczne, ale nie działa neuroprotektoryjnie.¹⁰ Dr Jones rozmawia o tym z panią Smith. Biorąc pod uwagę prawdopodobieństwo uzyskania korzyści w zakresie objawów ruchowych i mniej działań niepożądanych, pani Smith decyduje się na leczenie lewodopą. Dr Jones przypomina jej, że chociaż korzyści związane z zaburzeniami ruchowymi są większe przy zastosowaniu lewodopy, to ryzyko wystąpienia powikłań ruchowych w ciągu 5 lat wynosi blisko 31%, w porównaniu z 10% ryzykiem lub 21% zmniejszeniem ryzyka wystąpienia dyskinez u pacjentów leczonych na początku choroby pramipeksolem.

Cztery lata później pani Smith zgłasza występowanie dyskinez, znacznie zaburzących jej sprawność i występujących przez większość dnia mimo licznych modyfikacji dawkowania leku. Przeszukując piśmiennictwo dr Jones formułuje następujące pytanie: „Czy amantadyna, klozapina lub agonści receptora dopaminowego zmniejszają nasilenie dyskinez u pacjentów z chorobą Parkinsona, podtrzymując jednocześnie funkcje ruchowe pacjenta?” Dr Jones analizuje wyniki badań dotyczących serii przypadków, donoszących o korzyściach wynikających ze stosowania amantadyny, klozapiny i agonistów dopaminy w leczeniu dyskinez. Ponieważ serie przypadków są obciążone większym ryzykiem popełnienia błędów wnioskowania (zwykle badanie prowadzone w jednym ośrodku, z udziałem jednego lekarza przy wąskim spektrum pacjentów) wyszukuje badania kliniczne, dotyczące leczenia dyskinez. Znajduje jedno badanie przeprowadzone z podwójnie ślepą próbą i kontrolowane placebo, oceniające skuteczność amantadyny w zmniejszaniu dyskinez oraz jedno dotyczące leczenia tych powikłań ruchowych za pomocą klozapiny.^{16,17} Przed podjęciem decyzji terapeutycznej na podstawie wyników tych badań dr Jones dokona ich klasyfikacji pod względem jakości.

Dr Jones zdaje sobie sprawę, że istnieje wiele systemów klasyfikacji dowodów. Zapoznaje się więc z systemem GRADE

Ramka 5. Ocena dowodów na podstawie badania pilotażowego GRADE

- Jakość dowodu w różnych badaniach dla każdego wyniku
 - Jakość dowodu jest duża w badaniach typu RCT
 - Jakość dowodu jest umiarkowana w badaniach Quasi-RCT (dalej nie sprecyzowane)
 - Jakość dowodu jest mała w badaniach obserwacyjnych
 - Jakość dowodu pochodzącego z jakiegokolwiek innego źródła jest bardzo mała
 - 1 stopień: poważne błędy, istotna niezgodność lub pewna niejasność co do możliwości zastosowania w rozważanej kwestii
 - 2 stopnie: bardzo poważne błędy, duża niejasność dotycząca możliwości zastosowania w rozważanej kwestii
 - +1 stopień, jeśli: istnieje silna zależność i nie ma poważnych zmiennych zakłócających, zależność jest stała i bezpośrednia
 - +2 stopnie, jeśli: istnieje wybitnie silna zależność i nie ma poważnych zmiennych zakłócających, zależność jest stała i bezpośrednia
- Względna istotność skutków zdrowotnych
 - Dla każdego badanego skutku zdrowotnego, w celu podjęcia decyzji odnośnie do danej sytuacji klinicznej, należy ocenić jego względną wartość. Na przykład korzyść wynikająca z leczenia może być duża, ale nie określono ryzyka zgonu. Ryzyko może przekraczać jakkolwiek korzyść wynikającą z leczenia. Nie zmienia to opinii na temat ogólnej jakości dowodu.
- Całkowita jakość dowodu wobec istotnych skutków zdrowotnych
 - Całkowita jakość dowodu wobec istotnych skutków zdrowotnych powinna być oparta na wartości najniższej dotyczącej tych skutków zdrowotnych, które są kluczowe przy podejmowaniu decyzji
- Bilans korzyści i ryzyka
 - W przypadku, gdy punkty końcowe są poważne (jak śmiertelność), przewidywana korzyść powinna być oceniona względem prawdopodobnego lub możliwego ryzyka.
- Zalecenia
 - Powinny być oparte na jakości dowodu, możliwości zastosowania zaleceń w praktyce i przy uwzględnieniu bilansu ryzyka i prawdopodobnej korzyści
 - Kategorie zaleceń:
 - Jeśli 90-100% osób zgodziłoby się na zastosowanie zaleceń = zrób to
 - Jeśli 60-90% osób zgodziłoby się na zastosowanie zaleceń = prawdopodobnie zrób to
 - Jeśli 40-60% osób zgodziłoby się na zastosowanie zaleceń = może zrób to
 - Jeśli 30-40% osób zgodziłoby się na zastosowanie zaleceń = prawdopodobnie nie rób tego
 - Jeśli 0-30% osób zgodziłoby się na zastosowanie zaleceń = nie rób tego

Na podstawie: Atkins D, Briss PA, Eccles M, et al. Systems for grading the quality of evidence and the strength of recommendations II: Pilot study of a new system. *BMC Health Services Research* 2005; 5: 25-37.

Ramka 6. Kryteria oceny metody leczenia według American Academy of Neurology

Klasa I: Prospektywne badanie typu RCT z zaślepieniem dotyczącym skutku zdrowotnego, przeprowadzone na reprezentatywnej populacji. Wymagania:

- Jasno zdefiniowany pierwotny wynik końcowy
- Jasno zdefiniowane kryteria wyłączenia i włączenia
- Właściwe odnotowywanie wykluczonych uczestników badania i osób, które zmieniły leczenie w czasie badania, przy czym ich liczba musi być wystarczająco mała dla utrzymania minimalnego obciążenia błędem wnioskowania
- Przedstawienie odpowiedniej charakterystyki wyjściowej, w dużym stopniu równorzędnej w poszczególnych grupach terapeutycznych lub odpowiednie statystyczne korygowanie różnic

Klasa II: Prospektywne badanie kohortowe z parowaniem, obejmujące reprezentatywną populację, w którym ocena skutku zdrowotnego została zaślepiena, które spełnia kryteria wymienione w punktach a-d lub badanie typu RCT obejmujące reprezentatywną populację, które nie spełnia jednego kryterium wymienionego w punktach a-d.

Klasa III: Wszystkie inne badania kontrolne (obejmujące grupę kontrolną o dobrze zdefiniowanym naturalnym przebiegu choroby lub pacjentów będących swoją własną grupą kontrolną), obejmujące reprezentatywną populację, w którym wyniki są oceniane niezależnie bądź niezależnie uzyskiwane poprzez obiektywny pomiar.

Klasa IV: Dowód przedstawiony na podstawie wyników badań niekontrolowanych, opisu serii przypadków, pojedynczych przypadków lub opinii eksperta.

Eldlund W, Gronseth G, So Y, Franklin G. AAN Clinical Practice Guideline Process Manual. St. Paul: American Academy of Neurology. 2005; za zgodą.

(Grading of Recommendations Assessment, Development and Evaluation Working Group) i zauważa, że kryteria pozwalają na ocenę jakości badania, nie może jednak zlokalizować żadnego źródła wyjaśniającego pojęcie „poważne ograniczenie” (ramka 5).¹⁸ Zwraca się do metody opracowanej przez ANN i stwierdza, że w odniesieniu do leczenia odpowiednie są kryteria wymienione w ramce 6.⁵

Stosując te dwa systemy, dr Jones klasyfikuje pierwsze badanie, dotyczące skuteczności amantadyny w leczeniu dyskinez. Badanie to było randomizowane, kontrolowane placebo i ukończyło je 92% chorych (klasa I wg kryteriów leczenia ANN), ale wszyscy pochodzili z jednego ośrodka (wąskie spektrum chorych, co było powodem degradacji do klasy II). Dlatego stosując kryteria ANN, dr Jones wysuwa wniosek, że amantadyna jest prawdopodobnie skutecznym lekiem w zmniejszaniu dyskinez w chorobie Parkinsona i można rozważyć jej stosowanie w tym celu. Stosując system klasyfikacji

Ramka 7. System klasyfikacji dowodów i zaleceń według GRADE

Stopień 1A: Silna rekomendacja (stosowana u większości pacjentów w większości przypadków bez zastrzeżeń), dowód o wysokiej jakości

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

Stopień 1B: Silna rekomendacja (może być zastosowana u większości pacjentów w większości przypadków bez zastrzeżeń), dowód o średniej jakości

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

Stopień 1C: Silna rekomendacja (ale może ulec zmianie, gdy będzie dostępny dowód o wysokiej jakości), dowód o niskiej lub bardzo niskiej jakości

Korzyści wyraźnie przewyższają ryzyko i obciążenia

Dowód o niskiej lub bardzo niskiej jakości pochodzący z badań obserwacyjnych lub opisu serii przypadków

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

Korzyści równoważą ryzyko i obciążenia

Badania typu RCT bez poważnych ograniczeń lub silny dowód pochodzący z badań obserwacyjnych

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

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

Stopień 2C: Bardzo słaba rekomendacja (zastosowanie innych metod alternatywnych może być równie uzasadnione), dowód o niskiej lub bardzo niskiej jakości

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

Guyatt G, Gutterman D, Baumann MH, et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians Task Force. Chest 2006;129:174-81; za zgodą.

GRADE, ograniczenia związane z możliwym błędem wnioskowania wynikającym z udziału jednego ośrodka i krótkiego czasu trwania badania oznaczają, że jakość dowodu spada z wysokiej do niskiej. To przekłada się na stwierdzenie, że „dalsze badania będą miały prawdopodobnie istotny wpływ na pewność oceny efektu i prawdopodobnie zmienią tę ocenę.”¹⁹ Zalecenia odpowiadające temu poziomowi dowodu zależą od bilansu ryzyka i korzyści wynikających ze stosowania amantadyny w leczeniu dyskinez. Według doniesień badaczy w żadnej grupie pacjentów nie doszło do zdarzeń niepożądanych, ale badanie trwało zaledwie 3 tygodnie. Dlatego bilans ryzyka i korzyści jest niejasny. Na podstawie kryteriów GRADE amantadyna otrzymała zalecenia stopnia 2C (ramka 7).²⁰ To znaczy, że na podstawie dowodów niskiej jakości i przy niepewnym bilansie korzyści i ryzyka, może być zalecana bez przekonania w leczeniu dyskinez. Zalecenia uwzględniające kryteria opracowane przez ANN i kryteria GRADE są podobne.

Porównując te dwa badania, dr Jones podsumowuje uzyskane informacje: amantadyna powoduje 24% zmniejszenie dyskinez, mierzonych w skali Goetza, skraca czas trwania fazy „on” z dyskinezami i poprawia funkcje ruchowe pod względem występowania fazy „off”. Badanie z zastosowaniem klozapiny, w którym oceniano czas wystąpienia dyskinez, wykazało skrócenie czasu ich trwania o 1,7, bez zmian w zakresie występowania fazy „on” lub „off”. U 3 z 25 badanych leczonych klozapiną doszło do rozwoju eozynofilii (zwiąstan agranulocytozy), która szybko ustępowała. Więcej pacjentów wycofało się z badania w grupie otrzymującej placebo niż w grupie otrzymującej klozapinę. Względne ryzyko odchylenia w badaniu krwi przy stosowaniu klozapiny jest obliczone według wzoru

$$\text{Ryzyko wystąpienia eozynofilii} = \frac{\text{Liczba pacjentów z eozynofilią}}{\text{Liczba pacjentów przyjmujących klozapinę}} = \frac{3}{25} = 12\%$$

W obu badaniach oceną objęto małą liczbę pacjentów, a czas obserwacji był krótki (3 tygodnie dla amantadyny w schemacie naprzemiennym i 10 tygodni dla klozapiny). Dr Jones przekazuje te informacje pani Smith i przekonuje, że ponieważ stosowanie klozapiny może wywołać potencjalne działania niepożądane (agranulocytoza, konieczność cotygodniowego badania krwi), a badanie dotyczące skuteczności amantadyny przyniosło znaczące rezultaty (zmniejszenie nasilenia dyskinez i poprawa pod względem występowania fazy „off”), sugeruje włączenie amantadyny. Pani Smith zgadza się z tym planem leczenia.

Przez kolejne 6 miesięcy pacjentka cieszy się dobrą kontrolą dyskinez, jednak po tym czasie zgłasza ponowne ich nasilenie, przy czym utrzymują się one przez większość dnia i znacznie ograniczają jej sprawność. Dr Jones postanawia

poszukać danych na temat opcji leczenia chirurgicznego, jakie może być zastosowane u pani Smith, która ma obecnie 59 lat. Jest świadomy znacznego rozwoju metod leczenia chirurgicznego od czasów, kiedy rozpoczął naukę, a pallidotomia była zabiegiem z wyboru. Ogranicza zakres przeszukiwania bazy danych PubMed do prac opublikowanych od 2003 roku, w języku angielskim, badań typu RCT oraz wytycznych lub metaanalizy. Jego pytanie brzmi: „Czy u pacjenta z dyskinezami powodującymi inwalidztwo, stymulacja jądra niskowzgórzowego przynosi korzyści?”. Z przeszukanych streszczeń 33 spełniły kryteria wyszukiwania (22 lipca 2008 roku). Dr Jones wybiera jeden artykuł, aby zapoznać się z nim w całości.²¹ Badacze przedstawiają w nim wyniki leczenia 178 osób przydzielonych losowo do leczenia chirurgicznego lub optymalnego leczenia farmakologicznego. Wyniki badania wykazały znacznie większą poprawę w zakresie funkcji ruchowych u chorych poddanych stymulacji jądra niskowzgórzowego w porównaniu z osobami leczonymi farmakologicznie, a ocena jakości życia w obu grupach badanych wykazała podobną poprawę. Dr Jones zauważa, że podczas gdy chorych przydzielano losowo do leczenia chirurgicznego lub optymalnej farmakoterapii, oceniający wyniki nie byli zaślepieni pod względem przydzielenia do poszczególnych grup. Dlatego wynikające z tego ryzyko błędu wnioskowania jest znaczne. Metaanaliza badań dotyczących tego tematu wykryła to samo niedociągnięcie – w objętych nią badaniach skutki zdrowotne były oceniane przez niezślepienych badaczy, co wiązało się z ryzykiem obciążenia wyników błędem wnioskowania.²²

Ponieważ leczenie chirurgiczne jest inwazyjne dr Jones oblicza NNH (number needed to harm – liczba pacjentów, których poddanie określonej interwencji w określonym czasie wiąże się z wystąpieniem jednego dodatkowego niekorzystnego punktu końcowego – przy. tłum.). Z artykułu Deuschla i wsp.²¹ dowiaduje się, że w grupie pacjentów leczonych chirurgicznie u 12,8% doszło do poważnych zdarzeń niepożądanych, a w grupie lezonej farmakologicznie – u 3,8% chorych.

$$\text{NNH} = \frac{1}{\text{Bezwzględny wzrost ryzyka}} = \frac{1}{(\% \text{ badanych, poddanych głębokiej stymulacji jądra niskowzgórzowego, u których doszło do poważnego zdarzenia niepożądanego} - \% \text{ pacjentów z grupy kontrolnej, u których doszło do poważnego zdarzenia niepożądanego})} = \frac{1}{(12,8\% - 3,8\%)} = \frac{1}{9\%} = 11$$

Zatem na 11 pacjentów poddanych głębokiej stymulacji jądra niskowzgórzowego, u jednego dojdzie do wystąpienia poważnego zdarzenia niepożądanego.

Badacze przedstawili wyniki sparowane (każdy pacjent leczony chirurgicznie był sparowany z pacjentem leczonym najskuteczniejszymi metodami farmakologicznymi), wykazując, że w 55 z 78 par pacjentów chory leczony chirurgicznie osiągnął większą poprawę. Na podstawie tych danych dr Jones nie był w stanie powiedzieć pani Smith, jakie by było prawdopodobieństwo ustąpienia u niej dyskinez, gdyby poddała się leczeniu chirurgicznemu (bezwzględne zmniejszenie ryzyka). Aby móc obliczyć to prawdopodobieństwo, dr Jones musiałby wiedzieć, co w jej przypadku oznacza dobra odpowiedź na leczenie (pełne ustąpienie dyskinez, ustąpienie dyskinez powodujących inwalidzstwo czy zmniejszenie ich nasilenia). Mógłby wówczas znaleźć odpowiednią liczbę spełniających to kryterium pacjentów leczonych chirurgicznie, porównanych z odpowiednią liczbą chorych leczonych najskuteczniejszymi środkami farmakologicznymi, również spełniających to kryterium.

Pani Smith uwzględniła tę informację i postanowiła zastosować się do zaleceń lekarza. Odchodzi na emeryturę i decyduje się na kontynuację leczenia farmakologicznego. Blisko 10 lat po postawieniu rozpoznania pani Smith w towarzystwie męża ponownie zgłasza się na wizytę do dr. Jonesa. Mąż pacjentki relacjonuje występowanie u niej epizodów dezorientacji i zaburzeń pamięci. Państwo Smith zastanawiają się, czy u pacjentki doszło do rozwoju otępienia i czy istnieje leczenie poprawiające funkcje poznawcze.

Aby odpowiedzieć na to pytanie, dr Jones musi najpierw postawić właściwą diagnozę. Szuka informacji na temat testów przesiewowych w kierunku otępienia u osób z chorobą Parkinsona. Znajduje jedno badanie, w którym wykorzystano wyczerpujące metody testowe (Cambridge Cognitive Examination [Cam-Cog] w porównaniu z Mini-Mental Status Examination [MMSE]).²³ Badanie objęło pacjentów z chorobą Parkinsona i otępieniem, rozpoznany na podstawie kryteriów Diagnostic and Statistical Manual of Mental Disorders (czwarta edycja) (złoty standard kliniczny) oraz chorych bez otępienia. W związku z tym testy zostały zastosowane w szerokiej populacji pacjentów, którymi dr Jones również jest zainteresowany. Przy wykrywaniu otępienia testy dają podobne wyniki. Dr Jones koncentruje się na wynikach uzyskanych przez pacjentów w skali MMSE, ponieważ uważa, że może je wykorzystać w swojej praktyce (spełnienie kryterium zastosowalności). Wyniki badania wskazują, że czułość testu w wykrywaniu otępienia dla wartości odciętej 23 na 30 punktów wynosi 98%, a swoistość 77%.

Dr Jones przypomina sobie, że czułość stanowi odsetek wyników prawdziwie dodatnich dla testu, a swoistość – odsetek wyników prawdziwie ujemnych lub według wzoru matematycznego

$$\text{Czułość} = \frac{\text{Liczba wyników prawdziwie dodatnich}}{\text{Liczba wyników prawdziwie dodatnich} + \text{liczba wyników fałszywie dodatnich}}$$

Ramka 8.

<http://www.cche.net/usersguides/main.asp>

<http://www.poems.msu.edu/InfoMastery/> (kurs korzystania z EBM z określeniem czasu na kompletne moduły)

<http://www.jr2.ox.ac.uk/bandolier/index.html> (baza definicji i dyskusja na temat EBM)

http://www.cmaj.ca/cgi/collection/evidence_based_medicine_series?page52 (źródłowe strony EBM)

$$\text{Swoistość} = \frac{\text{Liczba wyników prawdziwie ujemnych}}{\text{Liczba wyników prawdziwie ujemnych} + \text{liczba wyników fałszywie ujemnych}}$$

Dlatego, aby zidentyfikować pacjentów dotkniętych chorobą, należy użyć testu przesiewowego o dużej czułości, a następnie, w celu identyfikacji rzeczywiście chorych osób, należy zastosować metodę o większej swoistości. Test diagnostyczny powinien cechować się dużą swoistością, ponieważ jest stosowany do postawienia ostatecznego rozpoznania u osób kwalifikowanych do leczenia. W przypadku otępienia w przebiegu choroby Parkinsona skutecznym testem przesiewowym i diagnostycznym jest wydłużony test Cambridge Cognitive Examination (czułość 95% i swoistość 94%).

Wzbogacony w tę wiedzę dr Jones wykonuje test MMSE. Pani Smith uzyskuje 21 na 30 punktów. Wnioskuje więc z uzasadnioną pewnością, że pani Smith cierpi na zaburzenia funkcji poznawczych i udziela małżonkom dalszych porad.

Podsumowanie

Historia pani Smith pokazuje wiele możliwości zastosowania zasad medycyny opartej na dowodach naukowych w praktyce. EBM nie wykorzystuje badań typu RCT, algorytmów czy praktyki komitetów naukowych. EBM stanowi zbiór umiejętności, w sposób przejrzysty wykorzystywanych w celu ułatwienia podejmowania decyzji terapeutycznych. Wykorzystywanie dostępnych dowodów naukowych w celu dostarczenia właściwych i rzetelnych informacji klinicydom i pacjentom jest ważnym elementem stosowania zasad EBM. Więcej informacji o EBM znajduje się na stronach internetowych wymienionych w tabeli 8.

Przed pracownikami ochrony zdrowia stoi większe niż kiedykolwiek wyzwanie przedstawienia dowodów na słuszność podejmowanych przez siebie decyzji. Zgodnie z wyraźnym stanowiskiem Agency for Healthcare Research and Quality, EBM odgrywa i będzie odgrywać kluczową rolę w poprawianiu jakości opieki zdrowotnej. W celu zapewnienia pacjentom możliwie najlepszej opieki każdy neurolog powinien opanować umiejętności zgodne z założeniami EBM.

Artykuł z *Neurologic Clinics of North America* (Volume 28, Issue 2, Pages 489-503, May 2010, Janis M. Miyasaki, MD, MEd, FRCPC, FAAN) jest publikowany za zgodą Elsevier Inc., New York, New York, USA.

Tłumaczenie Medical Tribune Polska. Ani autorzy, licencjonodawca, Elsevier Inc., i wydawca, Medical Tribune Polska, nie gwarantują ani nie odnoszą się do jakości i wartości reklamowanych produktów i usług, ani stanowiska reprezentowanego przez reklamodawców.

PIŚMIENNICTWO

1. Available at: <http://www.kff.org/kaiserpolls/upload/7625.pdf>. Accessed July 30, 2008.
2. Available at: <http://www.ahrq.gov/about/highlt07.htm>. Accessed July 30, 2008.
3. Cochrane AL. Effectiveness and efficiency: random reflections on health services. 2nd edition. London: Nuffield Provincial Hospitals Trust; 1971.
4. Guyatt G, Cairns J, Churchill D, et al. Evidence-based medicine working group evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
5. American Academy of Neurology. Clinical practice guideline process manual. 2004 edition. Available at: <http://www.aan.com/globals/axon/assets/3749.pdf>. Accessed July 30, 2008.
6. Hughes AJ, Daniel SE, Kilford L, et al. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatr* 1992;55:181-4.
7. Rundall TG, Martelli PF, Arroyo L, et al. The informed decisions toolbox: tools for knowledge transfer and performance improvement. *J Healthc Manag* 2007;52: 325-42.
8. Sackett DL, Richardson WS, Rosenberg W, et al. Evidence based medicine. London: Churchill Livingstone; 1998.
9. Post B, Maruschka P, Merkus P, et al. Prognostic factors for the progression of Parkinson's disease: a systematic review. *Mov Disord* 2007;22:1839-51.
10. Miyasaki JM, Martin W, Suchowersky O, et al. Practice parameter: initiation of treatment for Parkinson's disease: an evidence-based review. *Neurology* 2002; 58:11-7.
11. Available at: <http://www.aan.com/index.cfm?axon5redirect&&path5/go/practice/guidelines>. Accessed July 30, 2008.
12. Available at: www.agreecollaboration.org. Accessed July 30, 2008.
13. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004;4:38-45.
14. Parkinson Study Group. Pramipexole versus levodopa as initial treatment for Parkinson disease. *JAMA* 2000;284:1931-8.
15. Suchowersky O, Gronseth G, Perlmutter J, et al. Practice parameter: neuroprotective strategies and alternative therapies for Parkinson disease and evidence-based review. *Neurology* 2006;66:976-82.
16. Snow BJ, Macdonald L, McAuley D, et al. The effect of amantadine on levodopa-induced dyskinesias in Parkinson's disease: a double-blind, placebo-controlled study. *Clin Neuropharmacol* 2000;23:82-5.
17. Durif F, Debilly B, Galitzky M, et al. Clozapine improves dyskinesias in Parkinson disease: a double-blind, placebo-controlled study. *Neurology* 2004;62:381-8.
18. Atkins D, Briss PA, Eccles M, et al. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005;5:25-37.
19. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336:924-6.
20. Guyatt GH, Gutterman D, Baumann MH, et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians Task Force. *Chest* 2006;129: 174-81.
21. Deuschl G, Schade-Brittinger C, Krack P, et al. A randomized trial of deep-brain stimulation for Parkinson's disease. *N Engl J Med* 2006;355:896-908.
22. Kleiner-Fisman G, Herzog J, Fisman DN, et al. Subthalamic nucleus deep brain stimulation: summary and meta-analysis of outcomes. *Mov Disord* 2006;21: 290-304.
23. Hobson P, Meara J. The detection of dementia and cognitive impairment in a community population of elderly people with Parkinson's disease by use of the CAMCOG neuropsychological test. *Age Ageing* 1999;28:39-42.

Komentarz

Dr n. med. Katarzyna Szamotulska
Zakład Epidemiologii, Instytut Matki i Dziecka,
Warszawa

Komentowany artykuł przedstawia zastosowanie zasad medycyny opartej na dowodach naukowych (EBM) w codziennej praktyce lekarskiej, co jest zadaniem niełatwym. Ogólnie, medycyna oparta na dowodach polega na tym, aby czerpać inspirację do własnej pracy klinicznej z wyników badań naukowych dobrej jakości prowadzonych na świecie. Wyboru dowodów dokonuje się na podstawie samodzielnej krytycznej oceny jakości publikacji naukowych.

Medycyna oparta na dowodach naukowych stanowi swego rodzaju rewolucję myślenia w medycynie. Przeciwstawia się subiektywnemu, nierzadko autorytarnemu podejmowaniu decyzji opartemu wyłącznie na doświadczeniu i własnych lub narzuconych przekonaniach, bez odwoływania się do obiektywnych ustaleń nauki. „Postępuję tak, bo tak uważam”. „Postępuję tak, bo tego wymaga ode mnie przełożony”. „Postępuję tak, bo tak nauczyłem/am się kiedyś na studiach”. „Postępuję tak, bo przekonała mnie prezentacja leku przedstawiona przez producenta”.

Niedawno usłyszałam od pracownika naukowego jednej z klinik, przygotowującego się do wygłoszenia wykładu, że pewien problem zdrowotny, którym zajmuje się ta klinika, dotyczy 30% populacji i dotyka coraz większej liczby osób. Tymczasem z wielu badań zarówno krajowych, jak i zagranicznych wynika, że częstość występowania tego problemu wynosi mniej niż 10% i częstość ta nie zmienia się w czasie

i jest porównywalna między krajami. Medycyna coraz lepiej radzi sobie z tym problemem zdrowotnym i wzrasta zainteresowanie nim, ale nie jego częstota. Przykład ten ilustruje myślenie nieoparte na dowodach. A jak często podobne wypowiedzi dotyczą skuteczności postępowania, przydatności niektórych metod diagnostycznych czy rokowania?

Początkowo wyznawcy medycyny opartej na dowodach naukowych wyobrażali sobie, że lekarz w sytuacji, gdy jego wiedza na jakiś temat będzie niepełna, błyskawicznie wyszuka w komputerze, np. w bazie PubMed, artykuły potrzebne do uzyskania odpowiedzi na nurtujące pytanie, szybko oceni, które publikacje mają największą wartość naukową, natychmiast zinterpretuje wyniki z tych publikacji i na ich podstawie udzieli odpowiedzi sobie lub pacjentowi. Praktyka wskazuje jednak, że w codziennej praktyce nie jest to możliwe w tej formie.

Po pierwsze, trzeba mieć czas na wyszukiwanie i czytanie artykułów. Po drugie, trzeba mieć dostęp do komputera i do bazy artykułów naukowych w formie pełnych tekstów, a nie tylko abstraktów. Po trzecie, trzeba znać język angielski. Po czwarte, trzeba umieć ocenić wartość naukową znalezionych publikacji. Po piąte, trzeba posiadać znajomość podstawowych metod statystycznych stosowanych w badaniach w medycynie oraz umieć interpretować nierzadko skomplikowane tabele i ryciny z wynikami. Po szóste, trzeba umieć przełożyć te wyniki na praktykę.

Nie są to warunki możliwe do spełnienia w sytuacji, gdy naszym zadaniem jest przede wszystkim szybkie i skuteczne udzielenie pomocy pacjentowi. Niemniej, gdy chcemy uporządkować lub uaktualnić swoją wiedzę na konkretny temat, przygotowujemy zalecenia dotyczące postępowania, szukujemy poważne badanie naukowe, prowadzimy kształcenie, zasady medycyny opartej na dowodach są bardzo przydatne, uczą bowiem krytycznej oceny prac naukowych i pomagają zobiektywizować wiedzę.

O ile nad kwestią organizacji czasu i dobrą znajomością języka angielskiego trudno tu dyskutować, to przynajmniej w medycznych ośrodkach akademickich mamy coraz lepszą dostępność do baz artykułów naukowych, jest także sporo czasopism publikujących tłumaczone na polski artykuły naukowe. Oceny wartości naukowej artykułów, a także podstaw statystyki trzeba się nauczyć.

Warto dowiedzieć się, na czym polega dobrze zaplanowane badanie, po co jest potrzebna grupa kontrolna i jak

powinna być dobrana, dlaczego wyniki z badania eksperymentalnego mają silniejszą moc dowodu niż wyniki uzyskane z badania opisowego, dlaczego opisowe badanie kohortowe jest cenniejsze niż badanie kliniczno-kontrolne, jak powinno być skonstruowane badanie diagnostyczne. Ważne, aby zdać sobie sprawę z tego, jak istotne jest stosowanie najbardziej obiektywnych pomiarów przez badaczy oraz zapewnienie kompletności danych: zarówno pod względem liczby osób zbadanych w stosunku do liczby osób, które zamierzano zbadać, jak i pod względem liczby pomiarów wykonanych u tych osób w stosunku do liczby pomiarów, które miały być wykonane. Warto zrozumieć, na czym polega wnioskowanie statystyczne, co to jest istotność statystyczna, a co to jest przedział ufności, jakie parametry są najbardziej interesujące w oszacowanych modelach statystycznych dla wielu zmiennych, na czym polega analiza przeżycia. Warto również rozważyć, czy i jakie znaczenie dla wnioskowania przyczynowo-skutkowego ma liczba p , a jakie obecność zmiennych zakłócających. Istotne jest również, aby uświadomić sobie, jak ważna dla możliwości stosowania wyników prac naukowych w praktyce jest reprezentatywność osób uczestniczących w badaniu względem osób, na które chcielibyśmy przenosić wnioski płynące z badania.

W pewnym sensie artykuł Miyasaki pomaga nam w takiej nauce. Należy przy okazji dodać, że autor pracuje w szpitalu oddalonym o około 100 km od McMaster University w Kanadzie – miejsca, gdzie powstał termin medycyna oparta na dowodach naukowych. Co prawda koncepcja jak najszerszego stosowania wyników badań naukowych (zwłaszcza randomizowanych badań eksperymentalnych) w praktyce związana jest z nazwiskiem Archie Cochrane'a z Wielkiej Brytanii, to jednak McMaster University jest obecnie mekką EBM.

Stosowanie zasad medycyny opartej na dowodach ma wzbogacić, a nie zastąpić wiedzę lekarza, która – o ile jest gruntowna – ma priorytetowe znaczenie. Jest prawdą, że w większości przypadków, mimo ogromnego rozwoju naukowej medycyny, nie ma wystarczających dowodów: wyniki są sprzeczne lub jest mało badań, lub badania prowadzone były w niewielkich grupach pacjentów, dając niepewne wyniki. Między innymi dlatego medycyna pozostaje wciąż w dużej mierze sztuką. Niemniej, nie ma powodów, aby nie stosować w praktyce klinicznej metod o udowodnionej skuteczności, a także by stosować metody, które czynią więcej szkody niż pożytku.