



FORUM:

Czego dowiadujemy się z pragmatycznych badań leków psychotropowych?

What are we learning from pragmatic trials of psychotropic drugs?

Efektywność jako miara wyniku w badaniach lekowych w psychiatrii¹

Effectiveness as an outcome measure for treatment trials in psychiatry

W. WOLFGANG FLEISCHHACKER¹, GUY M. GOODWIN²

¹ Department of Psychiatry and Psychotherapy, Biological Psychiatry Division, University of Innsbruck, Innsbruck, Austria

² University Department of Psychiatry, Warneford Hospital, Oxford, OX3 7JX, UK

STRESZCZENIE

Panuje obecnie pewna dezorientacja co do względnej wartości badań klinicznych prowadzonych w celu określenia skuteczności w odróżnieniu od tych, w których bada się efektywność leków². Stanowi to szczególne wyzwanie wówczas, gdy badania, przeprowadzane jako eksperyment przez firmy farmaceutyczne dla instytucji regulujących rynek farmaceutyczny [regulators], służą do kształtowania i informowania praktyki klinicznej, zwłaszcza jeśli inne badania wykonane w warunkach bardziej zbliżonych do rzeczywistości nie potwierdzają przewidywanego korzystnego działania leku. Dokonałiśmy w tej dziedzinie przeglądu poświęconego przede wszystkim nowym lekom przeciwpsychotycznym. Krótko omówiliśmy też inne wskazania, m.in. zaburzenia nastroju, na które zwracano dotychczas znacznie mniej uwagi, lecz te same trudności są prawdopodobne również w ich przypadku. Doszliśmy do wniosku, że wtedy, gdy wyniki badania skuteczności są pozytywne, a badania efektywności negatywne, niekoniecznie trzeba preferować badanie efektywności – może się po prostu nie udało. Wówczas, gdy badania skuteczności i badania efektywności prowadzą do podobnych wniosków, ich wyniki wzajemnie się potwierdzają.

SUMMARY

There is at present some confusion about the relative value of clinical trials performed to investigate efficacy vs. those designed to investigate effectiveness. This is particularly challenging when studies performed as experiments for regulators by companies are used to shape and inform clinical practice, especially if studies conducted under more real life conditions fail to support predicted benefits. We review the field in relation to the new antipsychotics, in particular. Other indications, including mood disorders, which are also briefly touched upon, have so far received less definitive attention, but are likely to encounter the same difficulties. We conclude that, where the results of efficacy trials are positive and an effectiveness trial is negative, one should not necessarily prefer the effectiveness trial – it may simply have failed. Where efficacy trials and effectiveness trials point to similar conclusions, then the findings are mutually supportive.

Słowa kluczowe: badania kliniczne / metodologia / schizofrenia / zaburzenia nastroju / zaburzenie dwubiegunowe / depresja / leki przeciwpsychotyczne / leki przeciwdepresyjne / stabilizatory nastroju / skuteczność / efektywność / pragmatyczne badania lekowe

Key words: clinical trials / methodology / schizophrenia / mood disorders / bipolar disorder / depression / antipsychotics / antidepressants / mood stabilizers / efficacy / effectiveness / pragmatic trials

W ostatnich latach w naszej dziedzinie coraz trudniejsze staje się przełożenie wyników randomizowanych kontrolowanych badań klinicznych (RCTs, *randomized controlled clinical trials*) na praktykę kliniczną, czyli na informacje

o klinicznej użyteczności nowych farmaceutyków do leczenia schizofrenii i zaburzeń nastroju. Trudności te podkreśla fakt, że wnioski z meta-analiz i systematycznych przeglądów często bywają rozbieżne. Na przykład, Leucht i in. [1]

¹ Przedruk z *World Psychiatry*. 2009;8(1):23–36 – za życzliwą zgodą wydawcy. *World Psychiatry* jest organem Światowego Stowarzyszenia Psychiatrycznego (*World Psychiatric Association*).

² Przyp. tłum.: We wszystkich poniższych artykułach powtarzają się dwa terminy: *efficacy* [skuteczność] i *effectiveness* [efektywność], definiowane w literaturze dotyczącej badań lekowych następująco. (1) *Efficacy* – celem eksploracyjnego badania jest ustalenie związku przyczynowego, czyli odpowiedź na pytanie, czy dany lek działa w idealnych, optymalnych warunkach. Próba pacjentów jest wysoce wyselekcjonowana, bez współwystępujących dolegliwości i leczenia skojarzonego innymi lekami. Dany lek porównuje się z placebo lub innym lekiem o znanej skuteczności. Wskaźnikiem wyniku mogą być nawet stosunkowo niewielkie efekty leczenia. Wyklucza się tendencyjność ze strony badacza i pacjenta [podwójnie ślepa próba – ani pacjent, ani badacz nie wie, który z porównywanych leków otrzymuje pacjent]. Pacjenci są losowo przydzielani do podgrup leczonych porównywanymi metodami. (2) *Effectiveness* – poszukiwanie odpowiedzi na pytanie pragmatyczne, czy lek działa w rutynowych warunkach praktyki klinicznej. Kryteria doboru pacjentów nie są tak ostre, mogą być leczeni równocześnie innymi lekami. Informacja o rodzaju leku jest jawna, pacjenci mogą decydować o przyjmowaniu leku [zgodzić się na leczenie lub je przerwać, jak w życiu]. (Na podstawie: Pitter M.H., White A.R. [1999]. Efficacy and effectiveness. *Focus on Alternative and Complementary Therapies*, 4, 109–110).

po meta-analizie RCTs, w których porównywano leki przeciwpsychotyczne pierwszej generacji z lekami drugiej generacji doszli do wniosku, że „risperidon i olanzapina wykazują bardziej efektywne niż haloperidol oddziaływanie na ogólną symptomatologię i objawy negatywne” oraz że wszystkie testowane leki przeciwpsychotyczne drugiej generacji powodują mniej objawów pozapiramidowych i w mniejszym stopniu prowadzą do używania leków przeciwocholinergicznyc. Geddes i in. [2] analizując mniej więcej ten sam zestaw danych, wnioskuje jednak, że „nie ma jasnych dowodów na to, że atypowe leki przeciwpsychotyczne są bardziej efektywne czy lepiej tolerowane niż konwencjonalne leki przeciwpsychotyczne”. Davis i in. [3] po przeanalizowaniu danych ze 142 badań podają, że tylko niektóre leki przeciwpsychotyczne drugiej generacji wykazują większą skuteczność niż leki tradycyjne, natomiast Tandon i Fleischhacker [4] na podstawie jakościowego przeglądu dostępnych danych empirycznych doszli do wniosku, że „meta-analityczne badania porównawcze skuteczności leków przeciwpsychotycznych drugiej generacji innego typu niż klopazyna nie dostarczają niepodważalnych dowodów świadczących o ich innej skuteczności”.

Jeśli chodzi o zaburzenia nastroju, kontrowersja dotyczy przede wszystkim różnic między lekiem a placebo w badaniach skuteczności leków przeciwdepresyjnych [5], przy czym w mediach nagłaśniano wniosek, że w przypadku depresji jednobiegunowej leki te są tak samo skuteczne jak pastylki cukru. Problemem wspólnym dla tej kontrowersji i debaty nad lekami przeciwpsychotycznymi jest ekstrapolowanie na codzienną praktykę wyników badań przeprowadzonych w dość sztucznych warunkach, po to, by uzyskać zezwolenie na rejestrację leku.

Niewątpliwie wnioski z tych publikacji można interpretować jako wzajemnie sprzeczne. Przed naszą dziedziną stanęło zatem wyzwanie – trzeba znaleźć powody tych rozbieżności oraz rozwiązania, które zwiększyłyby przydatność badań klinicznych dla codziennej praktyki.

Stwierdzono, że dobór pacjentów jest jednym z głównych czynników odpowiedzialnych za rozbieżność wyników badań. Badania kliniczne leków przeciwpsychotycznych u pacjentów ze schizofrenią prowadzono na bardzo wyselekcjonowanych populacjach pacjentów [6–8], niereprezentatywnych dla tych chorych, dla których przeznaczone byłyby owe leki w codziennej praktyce. Coraz większy procent pacjentów wypada z RCTs, co niekiedy wiąże się ze specyficzną metodologią badania [9]. Budzi to wątpliwości co do analiz, w których tak czy inaczej trzeba jakoś zastąpić brakujące dane, oraz podważa proste wnioski – na przykład takie, że jakiś jeden lek prawdopodobnie będzie efektywny w leczeniu danego zaburzenia. Trudno twierdzić coś takiego, kiedy prawie połowa pacjentów z manią z grupy otrzymującej aktywny preparat nie kończy trzytygodniowego badania lekowego. Co więcej, poddano w wątpliwość, czy tradycyjne miary wyników leczenia, takie, jak poprawa globalnych wyników na skalach szacunkowych do oceny objawów psychopatologicznych, mają trafność ekologiczną, czyli mierzą rzeczywiste wyniki leczenia u pacjentów [10, 11].

Te same problemy są jeszcze bardziej widoczne w przypadku badań lekowych dotyczących depresji. Wielu pacjentów uczestniczących w RCTs leków przeciwdepresyjnych

zgłasza się z ogłoszenia i może otrzymywać wynagrodzenie za udział w badaniu. Dzieje się tak niewątpliwie w Stanach Zjednoczonych, gdzie przeprowadzono wiele takich badań. Co więcej, powszechnie uważa się, że aby spełnić kryteria kwalifikujące do udziału w badaniu zawyża się punktację kandydata w skalach do oceny depresji, co maskuje późniejsze wyniki leczenia przypisywane albo aktywnemu lekowi, albo placebo [5].

Owo niezadowolenie spowodowało, że wprowadzono pojęcie efektywności. W badaniach efektywności dąży się do uzyskania nieselekcjonowanej lub mniej wyselekcjonowanej grupy pacjentów, stosując szerokie kryteria włączenia do badań oraz niewiele powodów do wykluczenia kandydatów. Aby utrzymać niski odsetek osób wypadających z badania, można posłużyć się prostą metodologią badania. Zamiast mierzyć efekty interwencji terapeutycznych w kategoriach dosyć specyficznych zmian w zakresie psychopatologii, w badaniach nad efektywnością dąży się do mierzenia czegoś bardziej konkretnego. W dużych badaniach dotyczących chorób sercowo-naczyniowych za wskaźnik wyniku leczenia często przyjmuje się zgon. W psychiatrii zgon zdarza się zbyt rzadko, żeby go brać pod uwagę jako wskaźnik wyniku, ale przyjęcie do szpitala lub przerwanie leczenia uważa się za wskaźniki klinicznie istotne. W ramach lekkiego sprzeciwu wobec tendencji do mierzenia „twardych” wyników leczenia, równoległe poszukuje się wskaźników związanych z doświadczeniem pacjenta – często grupowanych pod ogólnym i oklepanym hasłem „jakości życia”. Ponadto są naciski, aby uwzględnić ocenę ekonomiczną wybranych metod leczenia. Wszystko to ma na celu uzyskanie rezultatów, które da się przełożyć na codzienną praktykę kliniczną, ale to także brzmi pozornie prosto.

W dalszym ciągu artykułu podamy kilka przykładów dużych pragmatycznych badań klinicznych z udziałem pacjentów cierpiących na schizofrenię i zaburzenia nastroju, a następnie omówimy zalety i wady badań efektywności w porównaniu z tradycyjnymi RCTs.

BADANIA EFEKTYWNOŚCI LEKÓW W SCHIZOFRENII

W ciągu ostatniego dziesięciolecia wykonano różne badania efektywności leków w schizofrenii – skoncentrujemy się na tych, które przeprowadzono na dużych próbach pacjentów. W przeglądzie uwzględniono badania zarówno metodą ślepej próby, jak i otwarte, o ile tylko stosowano w nich losowy przydział pacjentów do różnych metod leczenia. Ów losowy przydział, z odpowiednią kontrolą i utajeniem procesu przydzielania do grup, uważamy za kluczową właściwość umożliwiającą uczciwe porównanie dwóch metod leczenia.

CATIE: Clinical Antipsychotic Trials in Intervention Effectiveness (Kliniczne badania efektywności interwencji za pomocą leków przeciwpsychotycznych)

Program CATIE był badaniem klinicznym sponsorowanym przez amerykański National Institute of Mental Health (NIMH – Narodowy Instytut Zdrowia Psychicznego) po przetargu na kontrakt badawczy. Te duże badania pragmatyczne składały się z trzech faz. W pierwszej porównywano

pięć leków przeciwpsychotycznych nowej generacji z perfenazyną, lekiem pierwszej generacji. Po fazie I pacjenci mieli możliwość przejścia do jednej z dwóch różnych podgrup fazy II. W jednej zaplanowano początkowo porównanie klozapiny z innymi lekami przeciwpsychotycznymi nowej generacji u pacjentów, u których stwierdzono lekooporność w fazie I, druga zaś miała obejmować pacjentów z problemami tolerancji na lek. Przydział do różnych metod leczenia w fazie I i II był losowy i podwójnie ślepy, z wyjątkiem grupy otrzymującej klozapinę. Po fazie II pacjentów można było przesunąć do otwartych badań lekowych, w których badano różne starsze lub nowsze leki przeciwpsychotyczne. Przerwanie leczenia z dowolnej przyczyny stanowiło podstawowy wskaźnik wyniku leczenia [12].

W pewnym sensie badanie to reprezentuje metodologię hybrydową, ponieważ kryteria włączania i wskaźniki wyniku leczenia były zgodne z zasadą efektywności, podczas gdy reszta schematu badania była taka, jak w tradycyjnym RCTs. Co więcej, tego typu schemat obejmujący kolejne stadia może zachęcać do wczesnego przerywania leczenia w fazie I, ponieważ umożliwia przejście do drugiej fazy badania.

Opublikowano kilka artykułów przedstawiających wyniki fazy I i II oraz bardziej specyficzne efekty leczenia [13–15]. Poza tym, że częstość przerywania leczenia ze wszystkich możliwych powodów była istotnie mniejsza przy olanzapinie, nie wykazano większej skuteczności leków przeciwpsychotycznych drugiej generacji w porównaniu z perfenazyną w żadnej z opublikowanych dotychczas analiz. Perfenazynę wybrano z powodów pragmatycznych, aby zwiększyć poczucie równego traktowania. Oceniono, że lek bardziej typowy, taki, jak haloperidol, nie nadaje się ze względu na z góry przyjęte opinie pacjentów i badaczy. Skomentowano, że wybrana została perfenazyna „ponieważ ma słabsze działanie i umiarkowany profil skutków ubocznych”. Kwestią otwartą pozostaje, czy perfenazyna należycie reprezentuje grupę klasycznych leków przeciwpsychotycznych.

CUtLASS: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (Badanie kosztów i użyteczności najnowszych leków przeciwpsychotycznych w schizofrenii)

W badaniu tym, sponsorowanym przez brytyjską National Health Services (Państwową Służbę Zdrowia) również podjęto próbę porównania efektywności nowszych i starszych leków przeciwpsychotycznych. Klinicyści, którzy chcieli włączyć pacjenta do tego badania, musieli najpierw zdecydować, czy wykazywał on lekooporność na poprzednio stosowane leczenie (w takich przypadkach pacjenci byli włączani do podgrup, w których porównywano klozapinę z innymi lekami nowej generacji), albo też czy zmiana leczenia jest wskazana z innych przyczyn (w takich przypadkach pacjenci byli losowo przydzielani do podgrup otrzymujących jakiś lek przeciwpsychotyczny albo pierwszej, albo drugiej generacji). W tych dwóch ostatnich podgrupach klinicyści sami wybierali lek według swego uznania. Jakość życia przyjęto za główny wskaźnik wyniku leczenia.

Na koniec tego jednorocznego otwartego badania okazało się, że klozapina daje lepsze wyniki niż inne leki drugiej generacji u pacjentów lekoopornych, natomiast nie stwierdzono przewagi leków przeciwpsychotycznych drugiej generacji (46% pacjentów leczono olanzapiną) w porównaniu

ze starszymi preparatami (49% pacjentów w tej grupie otrzymywało sulpiryd) [16,17]. Z farmakologicznego punktu widzenia włączenie i nadreprezentacja sulpirydu jako leku „pierwszej generacji” było nieprzydatne. Sulpiryd jest farmakologicznie bardzo zbliżony do amisulpirydu, który stosowano w grupie otrzymującej leki drugiej generacji. Decyzja ta częściowo może odzwierciedlać trudności z rekrutacją pacjentów, i rzeczywiście w tych badaniach nie osiągnięto zakładanych początkowo celów dotyczących naboru do grup, co prawdopodobnie wynikało z nierównego traktowania. Klinicyści mogli spostrzegać leki „atypowe” jako lepsze i trudno ich było przekonać do stosowania starszych ([i bardziej „typowych”) leków przeciwpsychotycznych.

Niewykrycie różnic między lekami pierwszej i drugiej generacji staje się zatem wątpliwe. Co więcej, włączano pacjentów do badania dlatego, że trzeba im było zmienić sposób leczenia, a zatem potencjalnie wybierano pacjentów, którzy albo słabiej reagowali na lek [18] lub gorzej tolerowali leczenie (albo i jedno, i drugie).

CAFE: Comparison of Atypicals in First Episode psychosis (Porównanie leków atypowych w leczeniu pierwszego epizodu psychozy)

Przerwanie leczenia z dowolnej przyczyny było główną miarą wyników leczenia w tym badaniu klinicznym prowadzonym metodą podwójnie ślepej próby, w którym porównywano kwetiapinę z risperidonem i olanzapiną [19]. Częstość przerywania leczenia była wysoka we wszystkich trzech grupach, lecz ich nie różniła. To samo odnosi się też do zmian w wynikach skali objawów pozytywnych i negatywnych (PANSS). Podobnie jak w programie CATIE, olanzapina prowadziła do większego rozpowszechnienia przyrostu wagi.

EUFEST: European First Episode Study in Schizophrenia (Europejskie badania nad pierwszym epizodem schizofrenii)

W tym jednorocznym, randomizowanym, lecz nieprowadzonym metodą ślepej próby badaniu zrealizowanym w 13 krajach europejskich i w Izraelu, zajmowano się efektywnością leków przeciwpsychotycznych nowej generacji: amisulpirydu, kwetiapiny, olanzapiny i ziprasidonu, w porównaniu z małymi dawkami haloperidolu u pacjentów z pierwszym epizodem schizofrenii [20]. Głównym wskaźnikiem wyniku terapii było przerwanie leczenia lekiem, do którego pacjenci byli losowo przydzieleni na początku badania. Wszystkie leki nowej generacji wykazały lepsze działanie niż haloperidol. Ponadto, haloperidol nawet w małych dawkach wywoływał więcej skutków ubocznych, czyli objawów pozapiramidowych niż nowsze farmaceutyki. Globalne wyniki w skali PANSS, jeden z drugorzędnych wskaźników wyniku leczenia, nie różnicowały grup [21]. Wyniki w skali PANSS mierzono jednak rzadziej niż inne wskaźniki.

Dane uzyskane w programie EUFEST są sprzeczne z wnioskami, które często formułuje się na podstawie badań CATIE i CUtLASS – że leki atypowe nie wykazują istotnej przewagi nad starszymi preparatami. Haloperidol w małych dawkach był gorzej tolerowany niż leki drugiej generacji i przekładał się na istotnie krótszy okres utrzymywania się w leczeniu u pacjentów z pierwszym epizodem. Leki atypowe zarówno w programie CATIE, jak i EUFEST wykazywały odmienne działanie i nie wydaje się, aby

w stosowanych dawkach były równoważne. Porównanie z perfenazyną w CATIE, oraz jednej heterogennej grupy farmaceutyków z inną w programie CUtLASS, ogranicza wnioski, jakie można wyciągnąć z tych badań.

Trzeba jasno powiedzieć, że naturalistyczne badania kliniczne odzwierciedlają również naturalistyczne postępowanie lecznicze w praktyce, może nie zawsze zgodne z ogólnie przyjętymi dowodami empirycznymi i wytycznymi. Na przykład, w CATIE tylko około 40% wszystkich pacjentów w fazie I przyjmowało maksymalne dozwolone dawki. Z drugiej strony, badania pragmatyczne, w których daje się badaczom większą swobodę manewru w kwalifikowaniu pacjentów do badań i modyfikowaniu leczenia, mają tę zaletę, że większy procent pacjentów utrzymuje się w leczeniu, czego przykładem są programy CUtLASS i EUFEST. Metoda ślepej próby także ma wpływ na częstość przerywania leczenia: ogólnie rzecz biorąc, wskaźniki częstości przedwczesnego przerywania leczenia są wyższe w badaniach metodą podwójnie ślepej próby, takich jak CATIE i CAFE.

BADANIA EFEKTYWNOŚCI W ZABURZENIACH NASTROJU

STAR*D: *Sequenced Treatment Alternatives to Relieve Depression* (Sekwencyjne alternatywne strategie leczenia depresji)

W pierwszym etapie programu STAR*D nie zajmowano się skutecznością jakiegoś leku przeciwdepresyjnego w porównaniu z jego odpowiednikiem, natomiast wszystkim uczestnikom podawano jeden lek – citalopram, selektywny inhibitor zwrotnego wychwytu serotoniny. Systematycznie mierzono wyniki leczenia u ponad 2000 pacjentów z jednobiegunowym zaburzeniem afektywnym z epizodem dużej depresji.

Ogólnie biorąc, wskaźniki częstości remisji były prawdopodobnie niższe niż oczekiwano, zaś obciążenie skutkami ubocznymi większe. Trzydzieści procent badanych uzyskało remisję, co wymagało leczenia przez ponad 8 tygodni. Analiza podgrup wykazała szczególną skuteczność u kobiet z dużymi osiągnięciami życiowymi. Najśłabsze wyniki leczenia stwierdzono u pacjentów z dłuższymi epizodami depresji, większą liczbą współwystępujących zaburzeń psychicznych (zwłaszcza lękowych oraz z nadużywaniem substancji psychoaktywnych), większą liczbą ogólnych dolegliwości somatycznych, oraz niższym wyjściowym poziomem funkcjonowania i jakości życia [18].

W programie STAR*D pierwotnie zamierzano porównać różne strategie leczenia wprowadzane wówczas, gdy monoterapia citalopramem zostanie uznana za niewystarczającą. Niestety, wskutek zbyt permissywnego traktowania wyborów dokonywanych przez pacjentów częstość naprawdę losowego przydziału do konkurencyjnych metod leczenia nie spełniała oczekiwań. Po nieudanym leczeniu citalopramem, w drugim etapie STAR*D można było zmienić lek na inny (bupropion, sertralinę, lub wenlafaksynę) lub na terapię poznawczą, albo też wybrać wspomaganie citalopramu bupropionem, buspironem lub terapią poznawczą. Tylko 21 spośród 1439 pacjentów zgodziło się na losowe przydzielenie do któregoś z tych wariantów. Przytłaczająca większość miała swoje preferencje, na które

pozwał schemat badania. Wobec tego, chociaż porównanie strategii zmiany ze strategią wspomaganie leczenia było bardzo interesujące klinicznie, jednak spełzło na niczym, ponieważ umożliwiono pacjentom wybór jednego z tych podejść. U około 30% pacjentów we wszystkich grupach otrzymujących farmakoterapię nastąpiła remisja po zmianie leczenia, niezależnie od rodzaju zmiany [22, 23]. Częstość remisji u pacjentów otrzymujących terapię poznawczą była znacznie niższa [lecz różnica nie była istotna statystycznie z powodu braku mocy] [24]. Dalsze etapy w algorytmie leczenia ucierpiały z powodu malejącej liczby pacjentów, a między większością wyników leczenia nie było statystycznie istotnych różnic.

Możliwe są wzajemnie sprzeczne interpretacje programu STAR*D. Nihilisci powiedzą, że nie dowiedzieliśmy się z niego niczego. Optymiści twierdzą, że badane strategie leczenia wykazały przyzwoitą ogólną częstość remisji, jeśli postępowano zgodnie z algorytmem. Nie można rozstrzygnąć, czy reprezentuje to poprawę w leczeniu w rzeczywistych warunkach. Mocną stroną tego badania były: wielkość próby oraz niektóre wstępne dane farmakogenetyczne.

STEP-BD: *Systematic Treatment Enhancement Program for Bipolar Disorder* (Systematyczne wspomaganie leczenia dwubiegunowego zaburzenia afektywnego)

W programie STEP, porównywalnym ze STAR*D, podjęto duży wysiłek, by zbadać różne ścieżki leczenia pacjentów z dwubiegunowym zaburzeniem afektywnym. Spośród pięciu proponowanych badań tylko to, w którym uczestniczyli depresyjni pacjenci z zaburzeniem dwubiegunowym przyniosło użyteczne, randomizowane dane. W tym badaniu wzięło udział 366 pacjentów w stanie ostrej depresji. Analizowano ich odpowiedź na dodanie leków przeciwdepresyjnych lub placebo do już trwającego leczenia stabilizatorami nastroju (w praktyce interpretowano to bardzo liberalnie i dopuszczano równoczesne podawanie niemal dowolnego leku nie-przeciwdepresyjnego). Wyniki były negatywne, tzn. nie stwierdzono większej liczby przypadków remisji (lub przejścia w stan manii) w podgrupie otrzymującej leki przeciwdepresyjne [25].

Wynik ten można interpretować albo jako badanie negatywne (leki przeciwdepresyjne nie działają w depresji dwubiegunowej), albo jako badanie nieudane (nie wiemy, czy leki przeciwdepresyjne działają w depresji dwubiegunowej). Przy braku pozytywnego leczenia kontrolnego pozostaje to kwestią do dyskusji. Wniosek, jaki płynie z badania leków przeciwdepresyjnych w programie STEP jest być może taki: nie należy łączyć nieścisłości prawdziwego życia z nierzeczywistością (placebo) badania prowadzonego w celu uzyskania rejestracji leku.

BALANCE: *Bipolar Affective disorder – Lithium/Anticonvulsant Evaluation* (Dwubiegunowe zaburzenie afektywne – ewaluacja litu/leków przeciwdrgawkowych)

W programie BALANCE zakończonym w połowie roku 2008 porównywano skojarzone leczenie litem i kwasem walproinowym [np. Depakote] z monoterapią litem lub kwasem walproinowym. W badaniu poszukiwano odpowiedzi na pytanie o wyższość terapii skojarzonej nad monoterapią w długoterminowym leczeniu choroby afektywnej dwubiegunowej. Uważano, że będzie to miało ogólną

(*generic*) wartość w odniesieniu do zaburzenia dwubiegunowego, ponieważ w długotrwałej terapii łączenie różnych leków stało się bardzo powszechne, chociaż niemal nie ma niezależnych dowodów empirycznych potwierdzających korzyści z takiego postępowania. Schemat badania umożliwiał ponadto porównanie litu z kwasem walproinowym w monoterapii.

W badaniu zamierzano początkowo oceniać liczbę ponownych hospitalizacji jako podstawowy wskaźnik wyników leczenia, lecz wymagana wielkość próby byłaby wówczas bardzo duża (ponad 1000 uczestników) i przy braku odpowiednich funduszy nabór pacjentów do badania prawdopodobnie trwałby zbyt długo. W rzeczywistości, biorąc pod uwagę udane kontrolowane badania lamotryginy z użyciem placebo [25, 26] oraz nauzkę, jaka płynie z nieudanych badań nad kwasem walproinowym [27] (w których także planowano posłużenie się liczbą ponownych hospitalizacji jako podstawowym wskaźnikiem końcowym), przemyślano sprawę i za podstawowy wskaźnik przyjęto czas do momentu interwencji z powodu nowego epizodu zaburzeń nastroju. Ostatecznie próba liczyła ponad 400 osób, z czego 330 udało się zrandomizować tzn. losowo przydzielić do różnych podgrup. Analizę wyników badania przeprowadzono w połowie 2008 roku. Częstość mierzonych zdarzeń wskazuje na adekwatną czułość próby.

Podobnie jak EUFEST, program BALANCE był randomizowanym badaniem otwartym. Miał więc główną zaletę wszelkich RCTs: losowy i niejawni przydział do różnych metod leczenia. Badanie mogło być jednak skażone przez tendencyjność (pozytywne lub negatywne nastawienie) klinicysty lub pacjenta. W praktyce, znaczna długość okresu poprzedzającego terapię skojarzoną pomogła uniknąć w badaniu słabego przestrzegania zaleceń i do pewnego stopnia złagodziła tendencyjność, czyli pozytywne lub negatywne nastawienie do poszczególnych metod leczenia. Tym niemniej, na sposób leczenia mogło częściowo wpływać tendencyjne nastawienie, zwłaszcza w odniesieniu do wczesnych interwencji. Czynniki te ograniczają wartość wyników badania, lecz ich nie unieważniają, ponieważ brak zaślepienia próby jest oczywiście bardziej zbliżony do rzeczywistej praktyki klinicznej. Przede wszystkim, nie jesteśmy przekonani, że wśród uczestniczących w badaniu klinicystów lub pacjentów dałoby się wykryć jakieś jedno dominujące nastawienie (pozytywne lub negatywne) do którejkolwiek z badanych metod leczenia.

DYSKUSJA

Porównując mocne strony badań nas skutecznością i efektywnością trzeba będzie wyważyć zalety: z jednej strony – badania dobrze zdefiniowanych, jednorodnych grup pacjentów przy użyciu najnowszej metodologii i podwójnie ślepej próby, a z drugiej – uzyskiwania danych bliższych codziennej praktyce klinicznej, co oznacza rekrutowanie bardziej reprezentatywnych grup pacjentów i posługiwanie się potencjalnie bardziej użytecznymi wskaźnikami wyników leczenia. Oznacza to jednak również, że kiedy informacje o sposobie leczenia (podawanym leku) są jawne, tak, jak w normalnej praktyce, zawsze ryzykuje się, że tendencyjne nastawienie pacjentów i klinicystów

będzie determinowało wyniki leczenia. Ponieważ w badaniach otwartych z jawną informacją o rodzaju leczenia zawsze istnieje ryzyko tendencyjności obserwatora, wadę tę trzeba zrównoważyć wskazując na fakt, że możliwość uogólniania wyników jest wtedy większa i że mniej osób wypada z badania. Z metodologicznego punktu widzenia warunkiem *sine qua non* wydaje się randomizacja, jeśli podejmuje się kompromisową decyzję o badaniu z ujawnieniem rodzaju leczenia.

Definicja użytecznych wskaźników wyniku leczenia także była przedmiotem ożywionej dyskusji. Z jednej strony argumentowano, że przerwanie leczenia z dowolnej przyczyny stanowi nieskomplikowany i prymitywny wskaźnik, nawet jeśli wyodrębni się podgrupy, w których przerywano leczenie z powodu braku skuteczności leku, złej tolerancji, lub decyzji pacjenta. Z drugiej strony wysuwa się argument, że niewielka zmiana w globalnym wyniku skali PANSS, czy jeszcze bardziej specyficzne czynniki otrzymane w jakiejś skali szacunkowej mogą mieć tylko marginalną użyteczność kliniczną.

Działa tu swego rodzaju zasada niepewności. Im bardziej kontrolowany i rygorystyczny jest eksperyment, tym większą możemy mieć pewność co do efektu leczenia, ale w tym mniejszym stopniu badanie odpowiada warunkom w rzeczywistości. Im bardziej zbliżone do rzeczywistego życia staje się badanie efektywności, tym mniejszą daje pewność co do skuteczności badanego leku. W zasadzie uważamy, że oba rodzaje badań są pożądane, lecz zawsze razem, a nie jako alternatywa. Co więcej, największą pewność uzyskujemy wówczas, gdy w obydwu typach badań kierunek wyników leczenia jest podobny.

WNIOSKI

Biorąc pod uwagę wszelkie dostępne obecnie dowody naukowe sądzimy, że zarówno eksperymentalne RCTs, jak i bardziej pragmatyczne badania efektywności mają ważne miejsce w psychofarmakologii klinicznej. W idealnym procesie opracowywania leku, po eksploracyjnej fazie I, w której coraz częściej bada się pacjentów, grupy ryzyka lub zdrowych ochotników aby uzyskać potwierdzenie koncepcji teoretycznych, następują dobrze kontrolowane badania metodą ślepej próby [z utajnieniem informacji o rodzaju leku], ze ściśle zdefiniowanym wynikiem leczenia. Takie badania mogą wykazać skuteczność leku, lecz wielkości korzystnych efektów nie można po prostu ekstrapolować na rzeczywiste życie.

Wyniki tych badań fazy II i III należy następnie uzupełnić, być może już w fazie IIIb, przez większe pragmatyczne badania kliniczne, które muszą być tak zaprojektowane, by stawały najważniejsze pragmatyczne pytania kliniczne dotyczące całej populacji pacjentów. W przykładach, jakie tu rozważaliśmy, mogły mieć one różną postać, począwszy od bezpośredniej porównywalności z lekami wcześniejszej generacji, a skończywszy na stosowaniu badanego leku w terapii skojarzonej z innymi farmaceutykami czy interwencjami psychologicznymi. Bardzo skomplikowane plany badawcze ograniczają akceptowalność badania z punktu widzenia pacjentów (i badaczy). Co więcej, wszystkie badania pragmatyczne trzeba podejmować przed rozpoczęciem

intensywnego marketingu nowego leku, zanim opinie o nim zdołają się utrwalić w głowach badaczy. Naszym zdaniem w procesie rejestracji nowych leków wymaga się obecnie zbyt wiele (i coraz więcej) dowodów z początkowych faz badania lekowego, w których możliwość uogólniania wyników jest niewielka. Tymczasowa rejestracja pod warunkiem wdrożenia badań klinicznych na dużą skalę mogłaby zaspokoić niektóre z potrzeb, które dostrzegamy w procesie opracowywania nowych leków.

Badania efektywności należy planować tak, by miały kluczowe właściwości metodologiczne badań klinicznych, a mianowicie randomizację i utajnienie informacji o tym, do jakiej metody leczenia przydzielono pacjenta. Zagwarantuje je planowanie statystyczne, sformułowanie *a priori* jasnych hipotez, oraz koniecznie przestrzeganie standardów dobrej praktyki klinicznej. Doniesienia o przypadkach niepożądanych zdarzeń w takich badaniach lekowych mogłyby wcześniej zasygnalizować występowanie nieoczekiwanych problemów z bezpieczeństwem leku. Stosowanie tego rodzaju schematów badań lekowych we wcześniejszych stadiach procesu opracowywania leku mogłoby do pewnego stopnia zmniejszyć kontrowersje dotyczące praktycznej użyteczności badań prowadzonych w celu uzyskania rejestracji leku, a także umożliwić szybszą ocenę przydatności leku do zaspokajania rzeczywistych potrzeb klinicznych.

PIŚMIENNICTWO

- Leucht S, Pitschel-Walz G, Abraham D et al. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophr Res* 1999; 35: 51–68.
- Geddes J, Freemantle N, Harrison P et al. Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. *BMJ* 2000; 321: 1371–6.
- Davis JM, Chen N, Glick ID. A metaanalysis of the efficacy of second generation antipsychotics. *Arch Gen Psychiatry* 2003; 60: 553–64.
- Tandon R, Fleischhacker WW. Comparative efficacy of antipsychotics in the treatment of schizophrenia: a critical assessment. *Schizophr Res* 2005; 79: 145–55.
- Walsh BT, Seidman SN, Sysko R et al. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; 287:1840–7.
- Robinson D, Woerner MG, Pollack S et al. Subject selection biases in clinical trials: data from a multicenter schizophrenia treatment study. *J Clin Psychopharmacol* 1996; 16: 170–6.
- Hofer A, Hummer M, Huber R et al. Selection bias in clinical trials with antipsychotics. *J Clin Psychopharmacol* 2000; 20: 699–702.
- Hummer M, Fleischhacker WW. Do phase III trials have clinical value? *J Clin Psychopharmacol* 1999; 19:391–2.
- Kemmler G, Hummer M, Widschwendter C et al. Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. *Evid Based Ment Health* 2006; 9: 70.
- Leucht S, Davis JM, Engel RR et al. Defining “response” in antipsychotic drug trials: recommendations for the use of scale-derived cutoffs. *Neuropsychopharmacology* 2007; 32: 1903–10.
- Fleischhacker WW, Kemmler G. The clinical relevance of percentage improvements on the PANSS score. *Neuropsychopharmacology* 2007; 32: 2435–6.
- Stroup TS, McEvoy JP, Swartz MS et al. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophr Bull* 2003; 29:15–31.
- Lieberman JA, Stroup TS, McEvoy JP et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005; 353:1209–23.
- McEvoy JP, Lieberman JA, Stroup TS et al. Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior atypical antipsychotic treatment. *Am J Psychiatry* 2006; 163: 600–10.
- Keefe RS, Bilder RM, Davis SM et al. Neurocognitive effects of antipsychotic medications in patients with chronic schizophrenia in the CATIE Trial. *Arch Gen Psychiatry* 2007; 64: 633–47.
- Lewis SW, Barnes TR, Davies L et al. Randomized controlled trial of effect of prescription of clozapine versus other second-generation antipsychotic drugs in resistant schizophrenia. *Evid Based Ment Health* 2007; 10: 57.
- Jones PB, Barnes TR, Davies L et al. Randomized controlled trial of the effect on Quality of Life of second- vs first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Arch Gen Psychiatry* 2006; 63: 1079–87.
- Trivedi MH, Rush AJ, Wisniewski SR et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006; 163: 28–40.
- McEvoy JP, Lieberman JA, Perkins DO et al. Efficacy and tolerability of olanzapine, quetiapine, and risperidone in the treatment of early psychosis: a randomized, double-blind 52-week comparison. *Am J Psychiatry* 2007; 164: 1050–60.
- Fleischhacker WW, Keet IP, Kahn RS et al. The European First Episode Schizophrenia Trial (EUFEST): rationale and design of the trial. *Schizophr Res* 2005; 78: 147–56.
- Kahn RS, Fleischhacker WW, Boter H et al. Effectiveness of antipsychotic drugs in firstepisode schizophrenia and schizophreniform disorder: an open randomised clinical trial. *Lancet* 2008; 371: 1085–97.
- Rush AJ, Trivedi MH, Wisniewski SR et al. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med* 2006; 354: 1231–42.
- Trivedi MH, Fava M, Wisniewski SR et al. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med* 2006; 354: 1243–52.
- Thase ME, Friedman ES, Biggs MM et al. Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report. *Am J Psychiatry* 2007; 164: 739–52.
- Sachs GS, Nierenberg AA, Calabrese JR et al. Effectiveness of adjunctive antidepressant treatment for bipolar depression. *N Engl J Med* 2007; 356:1711–22.
- Goodwin GM, Bowden CL, Calabrese JR et al. A pooled analysis of 2 placebo-controlled 18-month trials of lamotrigine and lithium maintenance in bipolar I disorder. *J Clin Psychiatry* 2004; 65: 432–41.
- Bowden CL, Calabrese JR, McElroy SL et al. A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. *Arch Gen Psychiatry* 2000; 57:481–9.