



Schemat badawczy w badaniach klinicznych: właściwy schemat na właściwym miejscu

Clinical trial design: horses for courses

JOHN R. GEDDES

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

Fleischacker i Goodwin włączają się do toczącej się obecnie dyskusji na temat względnej wartości badań tak zwanej „skuteczności” i „efektywności” leków. W porównaniach badań lekowych trzeba uwzględnić fakt, że różne badania zaprojektowano tak, by uzyskać odpowiedź na różne pytania kliniczne i że decyzje metodologiczne zawsze wiążą się z kompromisem [1, 2]. Proponowano różne sposoby opisu odmiennych priorytetów i planów badawczych w badaniach klinicznych: np. badanie wyjaśniające w odróżnieniu od pragmatycznego [3], praktyczne w odróżnieniu od dużego prostego, w odróżnieniu od badania skuteczności [2], oraz duże proste w odróżnieniu od małego złożonego [4].

Schematy badania wyjaśniającego [„skuteczności”] charakteryzują się większym stopniem kontroli nad trafnością wewnętrzną [teoretyczną] oraz wyższą proporcją sygnału do szumu, lecz dzieje się to zwykle kosztem mniejszej trafności zewnętrznej [tzn. przystawalności do pacjentów w świecie rzeczywistym] [3, 5]. Celem badań wyjaśniających jest ustalenie, czy określona interwencja eksperymentalna *może działać* w kontrolowanych, zoptymalizowanych warunkach [6]. Natomiast celem bardziej pragmatycznych badań [„efektywności”] jest ustalenie, czy dana interwencja *rzeczywiście działa* w realnym świecie praktyki klinicznej, czyli w środowisku niemal z definicji mniej kontrolowanym i z większym poziomem szumu. Wprawdzie w badaniach pragmatycznych uczestniczą mniej wyselekcjonowani i bardziej reprezentatywni pacjenci i środowiska kliniczne, oraz stosuje się mniej standaryzowane, a bardziej rutynowe miary wyników klinicznych, lecz nieuniknione w badaniach efektywności kompromisowe schematy badawcze, z reguły zwiększają zmienność (i wobec tego szum statystyczny), a często także tendencyjność.

Fleischacker i Goodwin mają zatem rację zwracając uwagę na problem, który powstaje wówczas, gdy wydaje się, że wyniki badań wyjaśniających i bardziej pragmatycznych są rozbieżne. Czy badania wyjaśniające, czy też badania pragmatyczne są z natury rzeczy bardziej rzetelne i jest bardziej prawdopodobne, że dadzą ściślejszą ocenę „rzeczywistego” efektu badanego leku? Jest to złe postawione pytanie. Tak naprawdę zdecydowana większość badań znajduje się na kontinuum między wyidealizowanymi schematami badań wyjaśniających a pragmatycznymi planami badawczymi. Zalety każdego badania trzeba krytycznie ocenić pod kątem prawdopodobnych źródeł tendencyjności i szumu.

Fleischacker i Goodwin uważają, że randomizacja wraz z adekwatnym utajnieniem przydziału do grup jest warunkiem *sine qua non* uczciwego porównania dwóch (lub więcej) sposobów leczenia. Badania empiryczne wskazują jednak, że inne właściwości schematu badania, takie, jak zaślepienie próby mogą również istotnie wpływać na szanse, że wyniki badania nie będą tendencyjne [7]. Rzeczywiście, w badaniach, w których wyniki leczenia oceniane są subiektywnie, brak zaślepienia wydaje się równie ważną przyczyną tendencyjności, jak nieadekwatne utajnienie przydziału do grup [8]. Brak zaślepienia próby może prowadzić zarówno do *tendencji do postępowania* (wiedza o tym, do jakiej metody leczenia pacjent został przydzielony prowadzi do systematycznej zmiany zachowania się lekarza i pacjenta), jak i do *tendencji do oceny* (wiedza o przydzieleniu do określonej metody leczenia prowadzi do systematycznie odmiennej oceny wyników leczenia w grupie leczonej i kontrolnej). Można przewidywać, że zaślepienie będzie szczególnie ważne przy wysokim poziomie zarówno możliwości, jak prawdopodobieństwa tego rodzaju tendencyjności. Będzie tak wówczas, gdy zachowanie i wyniki leczenia łatwo ulegają modyfikacji oraz gdy nie ma naprawdę równego traktowania, zaś badacz i/lub uczestnicy badania wyraźnie preferują którąś z porównywanych metod leczenia.

Potrzebę uwzględniania schematu badawczego poszczególnych badań można dobrze zilustrować na przykładzie programu, w którym porównywano leki przeciwpsychotyczne pierwszej i drugiej generacji [PG i DG]. Kontekst był taki, że wiązano duże nadzieje z lekami DG, spodziewając się, że przyniosą znaczny postęp w leczeniu schizofrenii. W konsekwencji, początkowo patrzono przez palce na ograniczenia metodologiczne prowadzonych przez firmy farmaceutyczne badań lekowych (mieszczących się na kontinuum schematów badawczych bliżej bieguna wyjaśniającego) [9], przeceniano zalety leków DG [10], oraz szybko przechodzono na leki DG w praktyce klinicznej i stosowano je zamiast leków PG [11]. Wyniki systematycznych prac przeglądowych i meta-analiz badań lekowych sponzorowanych przez przemysł farmaceutyczny w zasadzie były podobne [12–14], chociaż autorzy jednego z przeglądów doszli do znacznie bardziej pozytywnych wniosków dotyczących leków DG niż inni [14, 15].

Opublikowano obecnie doniesienia z szeregu randomizowanych, kontrolowanych badań niesponsorowanych przez firmy farmaceutyczne, porównujących leki PG i DG [16–19].

Tak wiele niezależnych badań lekowych jest w psychiatrii czymś niezwykłym: świadczy zarówno o niezwykle istotnym rozwoju, jak i rzadko spotykanej ciągłej niepewności i doniosłości zagadnienia. Owe niezależne badania rozpatrywane jako grupa wydają się wskazywać, że chociaż mogą występować niewielkie różnice w skuteczności poszczególnych leków, nie da się wykazać, że takie korzyści są opłacalne, a ponadto równoważy je większa częstość pewnych niepożądanych skutków. Ten ogólny wniosek nie powinien jednak przesłaniać faktu, iż w omawianych badaniach stosowano bardzo różne schematy badawcze i miały one na celu uzyskanie odpowiedzi na różne, chociaż dopełniające się pytania. Brak miejsca nie pozwala na pełną krytyczną ocenę każdego z tych badań, omówię zatem tylko wybrane zagadnienia.

Program CATIE, określony przez swoich autorów jako badanie „pragmatyczne”, miał pewne cechy pragmatyczne (reprezentatywna grupa pacjentów, zmienne dawkowanie, dość długi okres katamnetyczny), ale stosowano w nim metodę zaślepienia i zadbano o wysoką jakość oceny wyników leczenia [2]. Rezultaty CATIE nie były zaskakujące, chociaż bardzo cenne w tym sensie, że potwierdziły obraz wyłaniający się z rozmaitych źródeł, m.in. zarówno z meta-analiz, jak i z danych obserwacyjnych dotyczących bezpieczeństwa leków [20].

Program CUtLASS mieścił się na kontinuum bliżej krańca pragmatycznego, był bowiem bez zaślepienia i umożliwiał wybór leków zarówno PG, jak i DG [18]. Rozpatrując to badanie bez kontekstu nie da się rozstrzygnąć, w jakim stopniu brak obserwowanych różnic w CUtLASS oznacza, że naprawdę nie było żadnych różnic, czy że w badaniu tym było zbyt dużo „szumów”, aby można było wykryć różnice. Wyniki CUtLASS dotyczące opłacalności są jednak wysoce zbieżne z wynikami innych niezależnych badań.

EUFEEST było ambitnym programem badania pacjentów z pierwszym epizodem, lecz ze względu na zastosowany schemat badania otwartego okazał się niezwykle podatny na tendencyjność postępowania (a w konsekwencji tendencyjność ocen), co spowodowało, że autorzy programu nie byli w stanie sformułować żadnych klarownych wniosków [19], chociaż wydaje się, że Fleischacker i Goodwin są obecnie bardziej na to przygotowani.

Brak miejsca uniemożliwia omówienie innych badań efektywności w innych zaburzeniach omawianych przez Fleischackera i Goodwina, lecz podobne uwagi odnoszą się do wszystkich. Niewątpliwie żadne z badań – ani „skuteczności”, ani „efektywności” – nie daje większego prawdopodobieństwa oszacowania „prawdy”. Wszystkie badania mają ograniczenia, a schemat badawczy jest sztuką kompromisu. Wszystkie badania należy zatem ocenić krytycznie. Mimo różnych niedociągnięć metodologicznych badań lekowych nowej generacji, podjęcie takich badań – oraz chęć ich finansowania przez budżet państwowy i organizacje charytatywne – nastąpiło z dużym opóźnieniem. Nadal brak odpowiedzi na wiele ważnych pytań klinicznych, którymi nie zajmowano się w badaniach zaprojektowanych wyłącznie na wąsko rozumiane potrzeby przemysłu i nadzoru farmaceutycznego [1, 21]. Ci, którzy planują badania o profilu bardziej pragmatycznym, na pewno muszą zidentyfikować istotne źródła tendencyjności w każdym poszczególnym badaniu. Ogromne znaczenie ma schemat badania

– powinien być tak solidny, aby wyniki badania były zarówno wiarygodne, jak i użyteczne. W przeciwnym razie owe z trudem uzyskane wyniki będą narażone na krytykę ze strony osób, którym się one nie spodobają!

PIŚMIENNICTWO

- Geddes JR. Large simple trials in psychiatry: providing reliable answers to important clinical questions. *Epidemiol Psychiatr Soc* 2005; 14: 122–6.
- Stroup TS, Geddes JR. Randomized controlled trials for schizophrenia: study designs targeted to distinct goals. *Schizophr Bull* (in press).
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis* 1967; 20: 637–48.
- Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995; 48: 23–40.
- Sackett DL. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *CMAJ* 2001; 165: 1226–37.
- Haynes B. Can it work? Does it work? Is it worth it? *BMJ* 1999; 319: 652–3.
- Schulz KF, Chalmers I, Hayes RJ et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408–12.
- Wood L, Egger M, Gluud LL et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336: 601–5.
- Bagnall AM, Jones L, Ginnelly L et al. A systematic review of atypical antipsychotic drugs in schizophrenia. *Health Technol Assess* 2003; 7: 1–193.
- Cipriani A, Geddes JR. Comparison of systematic and narrative reviews: the example of the atypical antipsychotics. *Epidemiol Psychiatr Soc* 2003; 12: 146–54.
- Geddes J. Generating evidence to inform policy and practice: the example of the second generation “atypical” antipsychotics. *Schizophr Bull* 2003; 29: 105–14.
- 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.
- Leucht S, Pitschel WG, 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.
- Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. *Arch Gen Psychiatry* 2003; 60: 553–64.
- Lieberman JA. Comparative effectiveness of antipsychotic drugs: a commentary on Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1) and Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE). *Arch Gen Psychiatry* 2006; 63: 1069–72.
- 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.
- Rosenheck R, Perlick D, Bingham S et al. Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: a randomized controlled trial. *JAMA* 2003; 290: 2693–702.
- Jones PB, Barnes TRE, Davies L et al. Randomized controlled trial of the effect on quality of life of second- vs. first-generation