

Nauka a etyka w przeprowadzaniu badań psychologicznych oraz analizowaniu i przedstawianiu ich wyników *

Robert Rosenthal **

Department of Psychology, Harvard University, Cambridge

SCIENCE AND ETHICS IN CONDUCTING, ANALYZING, AND REPORTING PSYCHOLOGICAL RESEARCH ***

The relationship between scientific quality and ethical quality is considered for three aspects of the research process: conduct of the research, data analysis, and reporting of results. In the area of conducting research, issues discussed involve design, recruitment, causism, scientific quality, and costs and utilities. The discussion of data analysis considers data dropping, data exploitation, and meta-analysis. Issues regarding reporting of results include misrepresentation of findings, misrepresentation of credit, and failure to report results as a result of self-censoring and external censoring.

Celem artykułu jest omówienie szeregu problemów o charakterze naukowym i etycznym związanych z przeprowadzaniem badań psychologicznych oraz analizą i przedstawianiem ich wyników. Główna teza tego artykułu głosi, iż istnieje ścisły związek pomiędzy etyczną i naukową jakością badań. Jeżeli w tych samych warunkach przeprowadza się dwa różne badania, to najprawdopodobniej badanie lepsze pod względem naukowym będzie równocześnie miało większą wartość etyczną. Trudno bowiem na gruncie etycznym uzasadnić marnowanie czasu osób badanych, pieniędzy instytucji sponsorujących¹ i miejsca na szpaltach czasopism naukowych związane z przeprowadzaniem badań o niskiej jakości naukowej. Badanie wartościowe naukowo jest z kolei dobrą społeczną inwestycją czasu, pieniędzy i papieru na rzecz rozwoju nauki oraz osób się nią zajmujących.

* Artykuł został opracowany na podstawie referatu zaakceptowanego przez Board of Scientific Affairs of the American Psychological Association (APA) i wygłoszonego na dorocznym zjeździe APA (Washington, DC; 15.08.1992). Za jego treść odpowiada wyłącznie autor. Przygotowanie artykułu możliwe było dzięki pomocy ze strony Spencer Foundation. Autor wyraża swoje podziękowania dla Elizabeth Baldwin, Petera Blancka i Ralpa Rosnowa za ich słowa zachęty i wsparcia.

** Korespondencję dotyczącą artykułu można kierować pod adresem: Robert Rosenthal, Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138, USA.

PRZEPROWADZANIE BADAŃ PSYCHOLOGICZNYCH

Pierwszą poruszoną w artykule kwestią jest problem wyboru planu badawczego i procedur badawczych oraz problem sposobu doboru osób badanych. Problemy etyczne związane z tą ostatnią kwestią można podzielić na dwie kategorie: pierwsza to sprawa zapewnienia uczestnikom badania bezpieczeństwa, druga dotyczy bardziej subtelnych aspektów przeprowadzania badania. Nikogo nie dziwi fakt, że badania narażające uczestników na niebezpieczeństwo budzą sprzeciw moralny. Okazuje się jednak, że wątpliwości natury etycznej mogą wiązać się również z przeprowadzaniem badań całkowicie bezpiecznych. Ich wątpliwa wartość etyczna wynika bowiem z braków i niedociągnięć wykorzystanych w nich planów badawczych.

PROBLEM PLANU BADAWCZEGO

Można wyobrazić sobie sytuację, w której przed członkami komisji decydującej o przyznawaniu funduszy na badania znalazł się projekt, który ma na

*** Artykuł w pierwotnej wersji ukazał się w *Psychological Science*, 1994, 5, 3, 127-134 i został za zgodą Cambridge University Press przetłumaczony przez Karolinę Krysińską.

¹ W Polsce m.in. KBN i Fundusz na Rzecz Nauki Polskiej (przyp. tłum.).

celu weryfikację hipotezy dotyczącej wpływu typu szkolnictwa na funkcjonowanie intelektualne dzieci. Hipoteza głosi, że szkoły prywatne w większym stopniu niż szkoły państwowe stymulują rozwój funkcji intelektualnych uczniów. Projekt obejmuje przeprowadzenie dokładnych badań dzieci z losowo wybranych szkół prywatnych i państwowych, a testem prawdziwości hipotezy ma być porównanie wyników uzyskanych przez obydwie grupy. Jakie problemy o charakterze etycznym wiążą się z przeprowadzeniem tego badania?

Choć uczestnicy badania nie są narażeni na żadne niebezpieczeństwo, to jednak dyskusyjna jest jego wartość etyczna. Wynika ona z nieadekwatności wykorzystanego planu badawczego. Celem badania ma być uzyskanie nowych informacji na temat związku pomiędzy typem szkolnictwa (prywatne vs. państwowe) a poziomem intelektualnego funkcjonowania dzieci, ale wykorzystywany w nim plan badawczy nie pozwoli na wyciągnięcie sensownych wniosków o charakterze przyczynowo-skutkowym. W planie brak randomizacji czy jakiegokolwiek próby przetestowania alternatywnej hipotezy (Cook, Campbell, 1979).

Dlaczego nieadekwatność zastosowanego w tych badaniach planu wzbudza obiekcje natury moralnej? Należy sobie uświadomić, że jego przeprowadzenie wiąże się z wykorzystaniem czasu uczniów, nauczycieli i pracowników administracyjnych szkoły, czasu który mógłby być spędzony o wiele lepiej. Należy sobie również uświadomić, że na podstawie niepoprawnie dobranego planu badawczego wyciągnię się nieuzasadnione, niewłaściwe i szkodliwe dla społeczeństwa, które w sposób bezpośredni lub pośredni sfinansowało te badania, wnioski. Nie można również zapominać, że poświęcone na to badanie pieniądze i czas, płynące z ograniczonej przecież puli, mogły być przeznaczone na przeprowadzenie o wiele cenniejszych naukowo badań.

Powyższej dyskusji o wartości etycznej tego badania można by uniknąć gdyby postawiono w nich hipotezę odpowiednią do wybranego planu badawczego – gdyby jego celem było jedynie sprawdzenie czy występują różnice w funkcjonowaniu intelektualnym uczniów ze szkół prywatnych i państwowych, to wybrany model nie budziłby żadnych wątpliwości.

PROBLEM DOBORU OSÓB DO BADAŃ

Kwestie etyczne związane z problem doboru osób badanych były już szczegółowo analizowane przez American Psychological Association Committee for

the Protection of Human Participants in Research, Committee on Standards in Research i jednego z pionierów badań w tej dziedzinie – Herberta Kelmana (APA, 1982; Blanck et al., 1992; Grisso et al., 1991; Kelman, 1968). Tutaj omówione będą tylko niektóre aspekty tego zagadnienia.

Z Ralphem Rosnowem, moim długoletnim przyjacielem i współpracownikiem, opracowaliśmy na podstawie analizy literatury psychologicznej szereg procedur mających na celu zmniejszanie błędów spowodowanych niereprezentatywnością próby złożonej z ochotników (*volunteer bias*)² i tym samym zwiększenie możliwości generalizacji wyników badań (Rosenthal, Rosnow, 1979, 1991; Rosnow, Rosenthal, 1993). Wykorzystując te procedury, w pewnym momencie zdaliśmy sobie sprawę z faktu, że współpracujące z nami osoby badane również są naszymi „sponsorami”. Ludzie ci przecież muszą podjąć decyzję, czy warto ofiarować badaczom swój czas, swoją uwagę i współpracę. Jeżeli traktuje się ich jako „sponsorów”, to nie można zapominać o udzieleniu im informacji na temat długofalowych korzyści płynących z przeprowadzanych badań. Należy być bardzo ostrożnym by nie obiecywać im zbyt wiele.

Obiecywanie zbyt wielkich korzyści, które mają płynąć z badań

Obiecywanie zbyt wielu korzyści z badań wiąże się z mówieniem potencjalnym uczestnikom badań, współpracownikom, administracji i samemu sobie, że planowane badania przyczynią się do osiągnięcia celów, których nie są w stanie osiągnąć. Można przypuszczać, że instytucje sponsorujące, koledzy psychologowie i administratorzy potrafią trafnie ocenić prawdziwość tych obietnic. Potencjalne osoby badane nie dysponują jednak odpowiednią wiedzą fachową, która by im umożliwiła ocenę prawdziwości twierdzeń badacza i dlatego też jest on zobowiązany do uczciwego przedstawienia potencjalnych korzyści płynących z badań. Przykładem obiecywania zbyt wielkich korzyści może być twierdzenie, że dzięki badaniom na pewno odkryje się lek na schizofrenię, depresję, nerwicę lękową czy raka.

Spostrzeganie nieistniejącej przyczynowości

Bliskie wyżej opisanemu udzielaniu nadmiernych obietnic jest zjawisko postrzegania nieistniejącej

² Zjawisko „volunteer bias” omawiają J. Brzeziński i S. Kowalik [w:] Sęk, H. (Red.), *Spoleczna psychologia kliniczna*, PWN, Warszawa, 1991, str. 273-274 (przyp. tłum.).

przyczynowości. Zjawisko to związane jest z tendencją badacza do mówienia o relacji przyczynowo-skutkowej w sytuacji gdy na jej istnienie nie wskazują żadne dane.

PRZEJAWY I EFEKTY POSTRZEGANIA NIEISTNIEJĄCEJ PRZYCZYNOWOŚCI

Ze zjawiskiem postrzegania nieistniejącej przyczynowości mamy do czynienia wtedy gdy pomimo braku obiektywnych podstaw do stwierdzenia zależności przyczynowo-skutkowej badacz posługuje się zwrotami implikującymi jej występowanie (np. zamiast „X związane jest z Y”, „na podstawie X można przewidywać Y”, czy „na podstawie X można wnioskować Y”, stwierdza, że „Y jest efektem X”, „X wpływa na Y” i „Y jest konsekwencją X”). Ważny jest w tym zjawisku też fakt odnoszenia przez badacza pewnych korzyści płynących z takiego przedstawiania danych. Mówienie o istnieniu zależności przyczynowo-skutkowych zwiększa przecież znaczenie uzyskanych wyników i czyni je bardziej godnymi uwagi.

Jeżeli badacz postrzega występowanie nieistniejących związków przyczynowych i nie jest świadomy popełnianego przez siebie błędu to jest to świadectwem jego braku doświadczenia i wiedzy naukowo-metodologicznej. Jeżeli jednak w sposób celowy i świadomy zniekształca wyniki badań to jest to już przykład nieetycznego i bezczelnego oszustwa.

Innym badaczom-psychologom, posiadającym odpowiednią wiedzę, nie sprawia problemu zidentyfikowanie przypadku błędnego postrzegania relacji przyczynowej. W innej sytuacji są jednak osoby badane: one z łatwością mogą paść ofiarą nieuczciwego opisu planowanych badań. I znowu, tego typu opis może być albo niepoprawnym narzędziem stworzonym przez nieświadomego popełnianego błędu naukowca-ignoranta, albo celowo zastosowaną przez nieuczciwego badacza przynętą, która ma zwiększyć gotowość ludzi do wzięcia udziału w badaniach. Będąc członkiem komisji decydującej o przyznawaniu środków na badania, muszę z zalem stwierdzić, że analizując przedstawiane nam propozycje badań, już nieraz spotkałem się z podobnym zjawiskiem.

EFEKTEM ZŁEJ NAUKI JEST ZŁA ETYKA

Opisane wyżej zjawiska to zaledwie kilka przykładów złej nauki. Etyczne uzasadnienie przeprowadzania jakichkolwiek badań jest tym mniejsze im gorsza jest jakość wykorzystanych w nich planów

badawczych, im gorsza jest jakość przeprowadzanej analizy wyników i im gorszy jest sposób przedstawienia rezultatów badań. O aspekcie etycznym badań można mówić nie tylko wtedy gdy wiążą się one z okłamywaniem, niewygodą czy zakłopotaniem uczestników badań, ale także wtedy gdy osoby badane nie są narażone na żadne niebezpieczeństwo. Jeżeli badanie jest przeprowadzane niepoprawnie i tym samym nie może w żaden sposób przyczynić się do rozwoju nauki, to jak można usprawiedliwić fakt wykorzystania na jego przeprowadzenie czasu, wysiłku i uwagi osób badanych oraz pieniędzy, pomieszczeń, wyposażenia i innych środków? W tym kontekście konieczna wydaje się pewna zmiana w zasadach funkcjonowania komisji decydujących o rozdziale środków na badania: komisje te muszą uwzględniać również techniczno-metodologiczne kompetencje badaczy, którzy zgłaszają swe wnioski. Niewątpliwie zwiększy to ilość pracy członków komisji i zmieni jej skład, tak by znalazło się w niej miejsce także dla osoby o pewnym poziomie wiedzy metodologicznej. Na pewno nie zawsze łatwe będzie ocenienie kompetencji naukowych badacza i metodologicznej jakości jego projektu, ale nie zawsze jest przecież łatwo osądzić inne, bardziej oczywiste aspekty etyczne proponowanych badań.

Zła jakość przeprowadzanych badań wiąże się również ze złą jakością kształcenia studentów psychologii. W sytuacji, gdy udział w przeprowadzanych badaniach ma dla studentów charakter quasi-obowiązkowy, argumentem na rzecz słuszności tego typu organizacji studiów jest twierdzenie, że wiąże się to z korzyściami o charakterze edukacyjnym. Ale przecież rezultatem nakłaniania studentów do brania udziału w złych jakościowo badaniach jest przyswojenie sobie przez nich błędnych przekonań na temat charakteru psychologii i nauki w ogóle. Z sytuacją taką mamy do czynienia wtedy, gdy celem badania ma być na przykład „ustalenie wpływu cech osobowości na funkcjonowanie poznawcze”, a badanie to polega na korelowaniu wyników uzyskanych przez studentów w kwestionariuszu osobowości z ocenami uzyskanymi przez nich na sprawdzianach i egzaminach. Jest to zła edukacja i oszustwo naukowe.

KOSZTY I ZYSKI

Koszty przeprowadzenia badania

Kiedy przed naukowcami lub komisjami przyznającymi fundusze stoi zadanie ocenienia projektu badań, wtedy najczęściej dokonują oni analizy kosztów i zysków z nim związanych. Analiza ta polega na

porównaniu kosztów przeprowadzenia badania (jego ewentualny negatywny wpływ na osoby badane, czas, pieniądze, narzędzia, wysiłek, inne środki) z zyskami z niego płynącymi (zyski dla jego uczestników, przyszłych pokoleń, nauki, świata, badacza). Przewidywane korzyści płynące z badań lepszej jakości i badań dotyczących istotnych problemów są większe niż przewidywane korzyści z badań o niższej jakości i dotyczących mniej ważnych problemów. Tego typu analizę kosztów i zysków można przedstawić w formie dwuwymiarowego diagramu, w którym jednym wymiarem są koszty, a drugim zyski (Rosenthal, Rosnow, 1984, 1991; Rosnow, 1990). Nie budzi wątpliwości zasadność przeprowadzania badań charakteryzujących się wysokimi zyskami i niskimi kosztami, i nie przeprowadzania badań o niskiej użyteczności i wysokich kosztach. Najtrudniejsze jest podjęcie decyzji w przypadku badań, w których koszty równoważą zyski.

Koszty nieprzeprowadzenia badania

Niekiedy okazuje się, że opisany model analizy kosztów i zysków jest niewystarczający. Nie uwzględnia on bowiem analizy kosztów i zysków związanych z nieprzeprowadzeniem badania (Rosenthal, Rosnow, 1984, 1991; Rosnow, 1990; Rosnow, Rosenthal, 1993).

Fakt nieprzeprowadzenia badania należy oceniać na gruncie etycznym tak samo jak fakt jego przeprowadzenia. Przykładem może tu być zespół onkologów stający przed szansą odkrycia leku zapobiegającego rakowi, którzy jednak z szansy tej nie korzystają ponieważ uważają swoją pracę za nudną i nieciekawą. Ich decyzję można oceniać w kategoriach moralnych, podobnie jak i decyzję naukowca, który podejmuje ryzykowne badania nad nowotworami. Innym przykładem może być odmowa przeprowadzenia badania wymagającego okłamania uczestników, a które mogłoby przyczynić się do zmniejszenia zjawiska przemocy, rasizmu, czy seksizmu. Badacz, który nie zgadza się na przeprowadzenie takiego badania nie rozwiązał dylematu etycznego, a jedynie wymienił jeden dylemat etyczny na drugi. Może się wydawać, że problemy moralne są takie same w przypadku badań związanych z praktyką, jak i w przypadku badań podstawowych. Okazuje się jednak, że trudniejsze może być ocenienie prawdopodobieństwa wynalezienia leku na raka, czy sposobu zmniejszenia rasizmu w przypadku badań podstawowych niż w przypadku drugiego rodzaju badań.

John Caplan ze Stanford University Law School (1988) w dyskusji z działaczami ruchu obrony praw zwierząt na temat wykorzystywania zwierząt w badaniach naukowych zręcznie posługuje się pojęciem „zmarowanych szans”. Pisze: „na ile możemy sobie pozwolić na badania przy użyciu zwierząt ... Nie możemy znać kosztów nie przeprowadzonych eksperymentów i nie podjętych badań. Kto przemawia w imieniu chorych i cierpiących? Kto przemawia w imieniu przyszłości?” (str. 839).

W przytoczonych przed chwilą przykładach analizowano koszty nie przeprowadzenia badania, które zapłaci obecne lub przyszłe pokolenie, lecz nie wspomniano o kosztach nie przeprowadzenia badania, dotyczących jego potencjalnych bezpośrednich uczestników. Nie można jednak zapominać, że niekiedy wzięcie udziału w badaniu wiąże się dla badanych z ważnymi ubocznymi korzyściami. Korzyści takie należy również uwzględniać, tak jak należało to zrobić w opisywanym niżej przypadku.

Kilka lat temu poproszono mnie jako eksperta w dziedzinie etyki badań bym wziął udział w obradach pewnej komisji odpowiedzialnej za przyznawanie funduszy na badania. Moim zadaniem była ocena projektu badawczego w dziedzinie genetyki, w ramach którego miano badać dzieci po to, by wykryć u nich obecność chromosomu XYY (który jak przypuszczano ma związek z zachowaniem o charakterze przestępczym). Projekt obejmował badania podłużne tych dzieci aż do momentu, w którym osiągną dorosłość, a potem określenie stopnia korelacji między typem chromosomalnym i występowaniem zachowań przestępczych. Poproszono mnie bym zreferował wyniki swoich badań nad efektem oczekiwań interpersonalnych, obawiano się bowiem, że efektem niezastosowania w tym badaniu procedury „podwójnie ślepej próby”, mogło być wystąpienie zjawiska samospełniającego się proroctwa (Rosenthal, 1966; Rosenthal, Jacobson, 1968, 1992). Niewątpliwie wykorzystanie „podwójnie ślepej próby” mogło rozwiązać ten problem, ale tak czy inaczej komisja nie wyraziła zgody się na przeprowadzenie badań.

Analizując ów projekt badawczy nie uwzględniono wielkich kosztów jego nieprzeprowadzenia, kosztów które poniosły same potencjalne osoby badane. Jakiego typu były to koszty? Była nimi utrata możliwości korzystania przez 20 lat z darmowej dobrej opieki pediatrycznej, na którą nigdy nie byłoby stać rodziców badanych dzieci. Czy wystarczającym moralnym uzasadnieniem tej decyzji była obawa przed koniecznością zastosowania procedury „podwójnie ślepej próby”, która praktycznie nie wiązała się z

zadnym ryzykiem? Komisja mogła przynajmniej poważnie przedyskutować koszty nieprzeprowadzenia badania. Nie zrobiła tego jednak.

ANALIZA DANYCH A ETYKA

POMIJANIE DANYCH

Kwestie etyczne związane z analizą danych obejmują zarówno problemy całkiem oczywiste, jak i bardziej subtelne. Najpoważniejszym i bezdyskusyjnym przykładem pierwszej z tych kategorii są przypadki analizy danych nigdy nie istniejących (zmyślonych). Częstszym zjawiskiem jest jednak pomijanie danych niezgodnych z przewidywaniami i hipotezami badacza oraz uznawaną przez niego teorią.

Odrzucanie danych o wartościach skrajnych

W dziedzinie analizy danych panuje szacowna, licząca już ponad 200 lat, tradycja mówiąca o tym, co należy robić z wynikami o skrajnych wartościach (Barnett, Lews, 1978). Związane są z nią pewne kwestie etyczne i techniczne. Kwestie techniczne dotyczą możliwości stworzenia jak najlepszych sposobów radzenia sobie z tego typu wynikami, etyczne związane są z implikacjami zastosowania tych metod dla stopnia rzetelności weryfikacji hipotez i teorii badacza. Na podstawie obserwacji można stwierdzić, że wyniki skrajne są inaczej traktowane wtedy, gdy potwierdzają hipotezę badacza, a inaczej wtedy, gdy są z nią niezgodne (Rosenthal, 1978; Rosenthal, Rubin, 1971). Przy pomijaniu tego typu danych powinno się w przypisach albo podawać wyniki, które by uzyskano gdyby wyników skrajnych nie pominięto, albo przynajmniej wspomnieć o fakcie ich odrzucenia.

Selekcja osób badanych

Innym rodzajem pomijania danych jest dokonywanie selekcji grup badawczych, która wiąże się z odrzucaniem z analizy pewnej części wyników. Również z tym zjawiskiem związane są pewne problemy natury technicznej i etycznej. Niekiedy z technicznego punktu widzenia słuszne jest odrzucenie pewnej części danych, np. wtedy gdy grupa badawcza jest nieliczna lub wtedy gdy jej odrzucenie umożliwia porównywanie wyników uzyskanych w różnych badaniach. Może się to jednak wiązać z kwestiami etycznymi; niekiedy odrzuca się przecież te wyniki, które nie potwierdzają założeń badacza. O każdym pominięciu pewnej części danych należałoby więc

poinformować czytelników, podobnie jak należałoby podać wyniki, które uzyskano by przy uwzględnieniu tych danych. Uwagi te dotyczą także nieuwzględniania wyników związanych z jedną lub większą ilością zmiennych.

EKSPLOATOWANIE JEST PIĘKNE

Fakt, że pomijanie danych ma pewien wymiar etyczny nie budzi żadnych wątpliwości. Mniej oczywisty jest wymiar etyczny eksploatacji. Naganne jest oczywiście eksploatawanie osób badanych, studentów, pracowników i kolegów-psychologów. Istnieje jednak pewien rodzaj eksploatacji, który godzien jest pochwały. Jest nim eksploatawanie danych.

Naukowców uczy się, że błędem technicznym, a nawet moralnym, jest analizowanie na różne sposoby uzyskanych przez nich danych (czyli „grzebanie w danych”). Uczy się ich testowania hipotezy za pomocą jednego testu i sprawdzania istotności statystycznej na poziomie $\alpha = .05$. I to tyle. Jeżeli wynik nie jest istotny statystycznie na poziomie $\alpha = .05$, to powinni dzielnie zagryźć wargi i zrezygnować z dalszej analizy uzyskanych przez siebie danych. Taka reanaliza mogłaby bowiem wykazać istotność statystyczną uzyskanych przez nich danych na poziomie $\alpha = .05$, a nie mają oni do niej prawa. Problem ten stanowi wspaniały temat na moralitet i przypomina wiersz o utraconej już na zawsze nie wybranej drodze napisany przez Roberta Frosta. Niemniej jednak jest to przykład złej nauki i złej etyki.

Jest to przykład złej nauki, bowiem chociaż „grzebanie w danych” wpływa na wielkość wartości p , zwiększając prawdopodobieństwo odkrycia czegoś nowego, ciekawego i ważnego (Turkey, 1977). Jest to przykład złej etyki ponieważ dogmat zakazujący grzebania w danych wiąże się z marnowaniem czasu, wysiłku, pieniędzy i innych środków, przy wykorzystaniu których przeprowadzono badanie. Jeżeli uznano, że warto przeprowadzić badanie, to powinno się też uznać, że warto wszechstronnie przeanalizować uzyskane w nim dane. W ten sposób istnieje bowiem możliwość spłacenia długu zaciągniętego u osób badanych, instytucji sponsorujących, nauki i społeczeństwa.

Warto jeszcze raz zaznaczyć, że „grzebanie w danych” rzeczywiście wpływa na wielkość wartości p . Istnieją jednak sposoby statystycznego dopasowania [np. Bonferroniego (Estes, 1991; Howell, 1992; Rosenthal, Rubin, 1984)], warte zastosowania. Najważniejsza jest jednak konieczność powtórzenia badań, bez względu na to czy w danych grzebano, czy też nie.

METAANALIZA JAKO IMPERATYW ETYCZNY³

Metaanaliza to zbiór pojęć i procedur służących ilościowej analizie wyników określonej dziedziny badań (Glass et al., 1981; Rosenthal, 1991). Praktyka wykazuje, że procedury metaanalizy są dokładniejsze, głębsze, systematyczniejsze i cenniejsze pod względem statystycznym niż tradycyjne metody podsumowywania wyników badań (Cooper, Rosenthal, 1980; Hedges, Olkin, 1985; Mosteller, Bush, 1954). Procedury te pozwalają na lepsze wykorzystanie danych i uzyskanie z nich większej ilości informacji i tym samym umożliwiają dokonywanie trafniejszej oceny ogólnej wielkości efektu czy zależności będącej przedmiotem badań, trafniejszej oceny ogólnego poziomu istotności statystycznej w danej dziedzinie badań oraz uzyskanie cennych informacji na temat zmiennych wpływających na wielkość badanego efektu czy zależności.

RETROAKTYWNY WZROST KORZYŚCI PŁYNĄCYCH Z BADAŃ

Metaanaliza w znacznym stopniu zwiększa retroaktywnie korzyści płynące z analizowanych przy jej zastosowaniu badań, umożliwia bowiem ekstrakowanie z nich dodatkowych informacji. Zwiększa zyski płynące z poszczególnych badań, zwiększa zasadność decyzji o zainwestowaniu w nie czasu, uwagi i wysiłku uczestników badania, pieniędzy, pomocy badawczych, pomieszczeń, czasu i wysiłku badacza oraz innych środków. Dzięki tej wyszukanej metodzie analizy statystycznej, polegającej na porównaniu ze sobą wyników wielu badań, możliwe jest zwiększenie korzyści z nich płynących.

Rezygnowanie z przeprowadzania metaanalizy tam, gdzie jest ona możliwa, staje się więc problemem natury etycznej – jest to zaprzeczanie szansy retrospektywnego zwiększenia korzyści z badań. Dokonywanie przeglądu literatury na dany temat jest finansowane z funduszy społecznych, należy się więc zastanowić czy pieniądze te wykorzystywane są rozsądnie i zgodnie z wymogami etyki. Po opracowaniu metaanalizy procedur analizy danych nie powinno się już przeprowadzać pre-metaanalizy analiz danych ilościowych – byłoby to marnowanie pieniędzy.

Metaanaliza wyników badań jest nie tylko zwykłą ogólną oceną wielkości efektu podstawowego ale

jest również próbą wyjaśnienia nieuniknionej zmienności wielkości efektów uzyskanych w różnych badaniach.

Jeszcze jedno – nie powinno się finansować badań mających na celu rozwiązanie jakiejś kontrowersji (np. czy dana metoda leczenia jest skuteczna), o ile badacz nie przeprowadził metaanalizy wyników wcześniejszych badań. Po metaanalizie przeprowadzonej przez Glassa (1976) i jego współpracowników (Smith et al., 1980) nie ma już dziś potrzeby podejmowania kolejnych badań poświęconych skuteczności psychoterapii – problem ten nie stanowi już dłużej kontrowersji.

Pseudokontrowersje

Metaanaliza jest bardzo pomocna przy rozwiązywaniu istniejących kontrowersji, ponieważ pozwala na wyeliminowanie dwóch częstych problemów związanych z oceną replikacji badań. Pierwszym problemem jest panujące powszechnie przekonanie, że w sytuacji gdy w jednym badaniu uzyskano istotny statystycznie efekt, a w próbie jego powtórzenia takiego efektu nie otrzymano, to replikacja nie powiodła się. Nie zawsze musi być to prawdą, o powodzeniu replikacji świadczy przecież uzyskana wielkość różnicy pomiędzy wielkościami efektów w obydwu badaniach. Drugim problemem jest przekonanie, że jeżeli w danym zjawisku rzeczywiście istnieje prawdziwy efekt, to każde badanie tego zjawiska powinno w sposób istotny statystycznie potwierdzać jego istnienie. Tymczasem jest tak, że jeżeli dany efekt jest całkiem duży, np. $r = .24$ i w każdym badaniu wykorzystuje się grupę na przykład 64 osobową, to wielkość efektu wyniesie .50 (Cohen, 1962, 1988; Rosenthal, 1994; Seldmeier, Gigerenzer, 1989). Jest to sytuacja, z którą bardzo często spotykamy się w psychologii, a jest w niej tylko jedna szansa na cztery, że wyniki uzyskane w obu badaniach będą istotnie statystycznie na poziomie $\alpha = .05$. W przypadku trzech badań istnieje tylko jedna szansa na osiem, że w nich wszystkich uzyska się efekt istotny statystycznie, nawet wtedy gdy efekt rzeczywiście istnieje i jego wielkość jest istotna.

Testowanie istotności

Metaanaliza zwiększa użyteczność badań również poprzez wpływ na wybór sposobu testowania na poziomie krytycznym. Metaanaliza uwzględnia rzeczywisty poziom uzyskanej istotności i nie jest ważny fakt czy rezultaty badania osiągnęły określony po-

³ Dostępna w j. polskim prezentacją metaanalizy jest artykuł F. L. Schmidta „Co naprawdę oznaczają dane? Wyniki badawcze, metaanaliza i wiedza kumulatywna w psychologii”, *Czasopismo Psychologiczne*, 1995, 1, 1-2, 19-33 (przyp. tłum.).

ziom (np. $p = .05$, test dwustronny). Do tego celu, zamiast wartości p wykorzystuje się odpowiadające jej normalne odchylenie standardowe. Wynik istotny statystycznie na poziomie $\alpha = .05$ (test jednostronny, w przewidywanym kierunku) zapisuje się więc w postaci $Z = +1.645$, a w sytuacji istotności statystycznej na tym samym poziomie α (test jednostronny, w nieprzewidywanym kierunku) jako $Z = -1.645$ (minus oznacza wynik w nieoczekiwanym kierunku). Oznaczone odchylenia standardowe służą jako cechy informujące o rezultatach prezentowanych badań raczej w sposób ciągły niż dychotomiczny. Dzięki ich wykorzystaniu zwiększa się ilość informacji płynących z przeprowadzonego badania, zwiększa się jego użyteczność i zmienia się stosunek kosztów do zysków. Dzięki temu wzrasta również wartość etyczna badania.

Małe efekty nie są małe

Jest jeszcze jeden sposób w jaki metaanaliza zwiększa użyteczność badań i ich wartość etyczną. Dzięki zastosowaniu metaanalizy możliwe jest bowiem oszacowanie wielkości, które mogą być bardzo ważne nawet wtedy, gdy ich wartość wynosi $r^2 = .00$. Szczególnie korzystne jest ocenianie praktycznego znaczenia dobrze ocenionych wielkości efektów. Nieprzydatna okazuje się tu wykorzystująca wielkość r^2 metoda $p.R^2 = .00$ może na przykład występować w kontekście badań nad nową metodą leczenia, która zmniejsza wskaźnik śmiertelności o .07 (Rosenthal, Rubin, 1982). Uświadomienie sobie, że efekty wielkości $r^2 = .05, .10$ i $.20$ (r^2 s wynosi kolejno .00, .01 i .04) wiążą się z uratowaniem życia pięciu, dziesięciu lub dwudziestu osobom na sto umożliwia dokonanie trafniejszej oceny kosztów i zysków związanych z badaniem.

PRZEDSTAWIANIE WYNIKÓW BADAŃ PSYCHOLOGICZNYCH

ZNIEKSZTAŁCANIE WYNIKÓW BADAŃ

Poznanie sekretów matki natury jest na tyle trudne, że stwarzanie dodatkowych utrudnień jest niepotrzebne. Utрудnienia takie mogą przyjmować formę podawania nieprawdziwych danych i publikowania nieuzasadnionych wniosków. Nauce szkoda wszystkie przypadki zniekształcania danych, niektóre z nich budzą jednak większy sprzeciw moralny niż inne.

Zamierzone zniekształcanie wyników badań

Najbezzwzględniejszą formą zamierzonego zniekształcania wyników jest przedstawianie danych, których nigdy nie uzyskano (Broad, Wade, 1982). Wykrycie tego procederu kończy (lub powinno kończyć) karierę naukowca-oszusta. Bardziej subtelną formą zamierzonego zniekształcania wyników jest przydzielanie do grup kontrolnych i eksperymentalnych takich osób, których zachowanie i reakcje można z góry przewidzieć, a które potwierdzą postawioną hipotezę. Z zamierzonym zniekształcaniem danych ma się do czynienia także wtedy gdy prowadzący badania osobiście rejestrują odpowiedzi osób badanych wiedząc, do której grupy badawczej one należą oraz wtedy gdy odpowiedzi te rejestrują pomocnicy badacza znający hipotezę badawczą i skład grup. Jeżeli w badaniu świadomie nie stosuje się procedury „podwójnie ślepej próby” i nie ukrywa się tego faktu, to nie można mówić o zniekształcaniu danych, niemniej jednak sytuacji takich należy unikać.

Niezamierzone zniekształcanie wyników badań

Zniekształcanie wyników badań może mieć również charakter niezamierzony i wynikać z błędów popełnionych w procesie zbierania danych. Nieprawdziwe wyniki mogą być rezultatem popełnienia błędów w zapisie, liczeniu i analizie danych (Broad, Wade, 1982; Rosenthal 1966). Zwykle nie rozpatruje się ich w kategoriach etycznych, nie można jednak zapominać, że ich popełnianie zmniejsza użyteczność badań, niekorzystnie zmienia relację koszty-zyski i zmniejsza uzasadnienie ich przeprowadzania.

Spotkać można jeszcze subtelniejsze formy (zwykle) niezamierzonego zniekształcania danych. Ich przykładem jest omawiane wyżej nieuzasadnione posługiwanie się terminami implikującymi występowanie zależności przyczynowo-skutkowej i zjawisko wątpliwej generalizowalności.

Wątpliwa generalizowalność

Celem hipotetycznych badań może być porównanie zdolności do tworzenia więzi terapeutycznej charakteryzującej terapeutów-kobiety i terapeutów-mężczyzn, ocenianej przez ich pacjentów. Badacz dysponuje grupą sześciu terapeutów (trzy kobiety i trzech mężczyzn), do których losowo przydziela 60 pacjentów. W takim przypadku analiza wariancji wspomina o trzech źródłach zmienności: płeć terapeutów

($df = 1$), terapeuci wgnieźdzeni w płec ($df = 4$) i pacjenci wgnieźdzeni w terapeutów ($df = 54$). Popularnym sposobem analizy takich danych jest podzielenie $MS_{płec}$ przez $MS_{pacjent}$, uzyskanie wartości testu F i potraktowanie terapeutów jako efektów stałych. Przedstawienie uzyskanych wyników jako $F(1,54) = 7,13; p = .01$, umożliwi jednak ich generalizację tylko na innych pacjentów leczonych przez tych terapeutów, a nie na żadnych innych terapeutów (Estes 1991; Snedecor, Cochran, 1989).

ZNIEKSZTAŁCANIE ZASŁUG BADAWCZYCH

W poprzednich akapitach omówiono problem zniekształcania wyników badań, czyli „co tak naprawdę odkryto?”. Teraz poruszone będą problemy autorstwa, czyli „kto tak naprawdę to odkrył?”.

Problem zasług autorskich

W nauce, m.in. w psychologii, nieuniknione jest pojawienie się skomplikowanych problemów związanych z przyznawaniem autorstwa. Wiele artykułów pisanych jest przez więcej niż jednego autora i pojawiają się pytania o to, kto jest autorem, kto współautorem, a o kim należy wspomnieć w odnośniku. Który ze współautorów powinien być wpisany jako pierwszy, kto jako kolejny a kto jako ostatni. Kwestii przyznawania zasług autorskich poświęcono wiele dyskusji i analiz, których rezultatem było opracowanie ogólnych wskazówek (APA, 1981, 1987; także Costa, Gatz, 1992). Niemniej jednak wskazane byłoby przeprowadzenie dalszych badań empirycznych, w ramach których autorzy, redaktorzy, wydawcy, profesorowie, studenci i praktycy proszeni byłiby o ocenienie zasług autorskich poszczególnych osób biorących udział w procesie badawczym.

Problemy pierwszeństwa

Problem zasług autorskich występuje w ramach jednej grupy badawczej, problem pierwszeństwa pomiędzy różnymi grupami. Jednym z najnowszych przykładów świadczących o występowaniu tego problemu jest sprawa „intelektualnej kradzieży” odkrycia wirusa wywołującego AIDS i wykorzystywanego do przeprowadzania testów na HIV odkrytego przez francuskich badaczy, rzekomo dokonanej przez Roberta C. Gallo i jego współpracowników (Palca, 1992). Z takimi problemami można spotkać się także w psychologii. Tu jednak pojawia się kwestia przyznawania pierwszeństwa w stwarzaniu teorii i koncepcji.

NIE PUBLIKOWANIE LUB NIE PRZEDSTAWIANIE DANYCH

Kwestie etyczne wiążą się nie tylko z wyżej omówionymi problemami zniekształcania wyników badań, czy kwestią zasług i pierwszeństwa w dokonaniu odkrycia, ale także z faktem nie przedstawiania wyników przeprowadzonych badań i przyczyn ich nie publikowania. Przyczyną wystąpienia takiej sytuacji może być autocenzura lub cenzura zewnętrzna.

Autocenzura

Niekiedy autocenzura godna jest pochwały. Nie opublikowanie sprawozdania ze źle przeprowadzonego badania może być prawdziwą przysługą wyświadczoną nauce i społeczeństwu. Niekiedy motywy autocenzury są szczerne, lecz jej rezultatem jest marnowanie informacji. Niektórzy badacze uważają na przykład, że nie powinni powoływać się na wyniki swoich (lub cudzych) badań, o ile wyniki te nie były wcześniej opublikowane i poddane ocenie. Wydaje się jednak, że gdy dane te uzyskano w poprawny sposób, to powinny być one wykorzystywane i brane pod uwagę przy dokonywaniu metaanalizy.

Niekiedy przyczyny autocenzury są akceptowalne i są przykładami złego uprawiania nauki i złej etyki. Niektórzy badacze nie publikują danych, które są sprzeczne z wcześniej uzyskanymi przez nich wynikami badań, z ich teorią czy uznawanymi wartościami. Przyczyn usprawiedliwiających niepublikowanie niewygodnych danych można znaleźć lub wymyślić bardzo dużo. Uczestnicy badania mogli na przykład dopiero rozpoczynać studia, mogli właśnie zdawać egzaminy, być tuż po egzaminie lub akurat kończyć studia. Dla dobra nauki w ogóle i dla dobra zachowania jej etycznej czystości, należy pamiętać o następującej zasadzie – należy przedstawiać wszystkie wyniki, które przyczyniają się do poznania danej sfery badawczej i które mogą być wykorzystane przez innych badaczy.

Niewątpliwie istnieją wyniki bardziej spektakularne niż inne. Jeżeli dane uzyskane w badaniach nad nowo opracowaną procedurą terapeutyczną wykazują jej skuteczność w zapobieganiu lub leczeniu chorób psychicznych lub somatycznych, to rzeczywiście wydaje się, że warto poświęcić jej więcej miejsca w prestiżowych czasopismach naukowych, niż wynikom wykazującym jej nieskuteczność. Nie można jednak zapominać o konieczności przedstawiania i mniej efektownych wyników, umożliwiających badaczom poznanie nowych faktów.

Cenzura zewnętrzna

Cenzurze zewnętrznej zawdzięczać można zarówno postęp jak i zastój w nauce. Gdyby nie istniała cenzura dokonywana przez recenzentów, komitety programowe i wydawców, to w nauce prawdopodobnie zapanowałby jeszcze większy chaos niż obecnie. Recenzenci pełnią rolę strażników nie dopuszczających do zalewu czasopism naukowych przez naukę o kiepskiej jakości.

Cenzura zewnętrzna ma dwie podstawy. Pierwszą z nich jest ocena badań pod względem metodologicznym. Jestem zdecydowanym zwolennikiem cenzury tego typu – nie powinno się publikować doniesień z badań bardzo złych jakościowo.

Drugą podstawą cenzury zewnętrznej jest ocena wyników uzyskanych w badaniach. W ciągu 35 lat pracy w psychologii niejednokrotnie spotykałem komentarze „takie wyniki nie są możliwe”, czy „te wyniki nie mają sensu”. Często uwagi te są całkowicie trafne, lecz kryterium „prawdopodobieństwa” wyników nie stanowi dobrego kryterium cenzury. Ukrywanie nieprawdopodobnych lub nieodpowiednich wyników jest przykładem złej nauki i złej etyki (Rosenthal, 1975, 1994).

WNIOSKI

W artykule omówiono niektóre problemy naukowe i etyczne związane z przeprowadzaniem badań psychologicznych oraz analizą i przedstawianiem ich wyników. Główną tezą artykułu było stwierdzenie, że jakość etyczna badań psychologicznych jest ściśle związana z ich jakością naukową. Powyższej prezentacji przyświecały dwa cele. Pierwszym z nich było uspokojenie zaniepokojonych, czyli przedstawienie możliwych sposobów poprawy jakości psychologii jako nauki i równocześnie możliwych sposobów poprawy jakości jej aspektu etycznego. Drugim celem było zaniepokojenie uspokojonych, poprzez przypomnienie im, że w obu tych sferach pozostało jeszcze bardzo dużo do zrobienia.

BIBLIOGRAFIA

- American Psychological Association. (1981). Ethical principles of psychologists. *American Psychologist*, 36, 633-638.
- American Psychological Association. (1982). *Ethical principles in the conduct of research with human participants*. Washington, DC (Rosenthal, R.).
- American Psychological Association. (1987). *Casebook on ethical principles of psychologists*. Washington, DC (Rosenthal, R.).
- Barnett, V., Lewis, T. (1978). *Outliers in statistical data*. New York: Wiley.
- Blanck, P.D., Bellack, A.S., Rosnow, R.L., Rotheram-Borus, M.J., Schooler, N.R. (1992). Scientific rewards and conflicts of ethical choices in human subjects research. *American Psychologist*, 47, 959-965.
- Broad, W., Wade, N. (1982). *Betrayers of the truth*. New York: Simon and Schuster.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.
- Cook, T.D., Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cooper, H.M., Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87, 442-449.
- Costa, M.M., Gatz, M. (1992). Determination of authorship credit in published dissertations. *Psychological Science*, 3, 354-357.
- Estes, W.K. (1991). *Statistical models in behavioral research*. Hillsdale, NJ: Erlbaum.
- Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Glass, G.V., McGaw B., Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Grisso, T., Baldwin, E., Blanck, P.D., Rotheram-Borus, M.J., Schooler, N.R., Thompson, T. (1991). Standards in research: APA's mechanism for monitoring the challenges. *American Psychologist*, 46, 758-766.
- Hedges, L.V., Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Howell, D. C. (1992). *Statistical methods for psychology (3rd ed.)*. Boston: PWS-Kent.
- Kaplan, J. (1988). The use of animals in research. *Science*, 242, 839-840.
- Kelman, H.C. (1968). *A time to speak: On human values and social research*. San Francisco: Jossey-Bass.
- Mosteller, F., Bush, R.R. (1954). Selected quantitative techniques. [w:] Lindzey, G. (Ed.), *Handbook of social psychology: Vol. 1. Theory and method*, (str. 289-334). Cambridge, MA: Addison-Wesley.
- Palca, J. (1992). „Verdicts” are in on the Gallo probe. *Science*, 256, 735-738.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Rosenthal, R. (1975). On balanced presentation of controversy. *American Psychologist*, 30, 937-938.
- Rosenthal, R. (1978). How often are our numbers wrong?. *American Psychologist*, 33, 1005-1008.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research (rev.ed.)*. Newbury Park, CA: Sage
- Rosenthal, R. (1994). On being one's own case study: Experimenter effects in behavioral research - 30 years later. [w:] Shadish, W.R., Fuller, S. (Eds.), *The social psychology of science*, (str. 214-229). New York: Guilford Press.
- Rosenthal, R., Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winson.
- Rosenthal, R., Jacobson, L. (1992). *Pygmalion in the classroom (expanded ed.)*. New York: Irvington.
- Rosenthal, R., Rosnow, R.L. (1975). *The volunteer subject*. New York: Wiley.

- Rosenthal, R., Rosnow, R.L. (1984). Applying Hamlet's question to the ethical conduct of research: A conceptual addendum. *American Psychologist*, 39, 561-563.
- Rosenthal, R., Rosnow, R.L. (1991). *Essentials of behavioral research: Methods and data analysis (2nd ed.)*. New York: McGraw-Hill.
- Rosenthal, R., Rubin, D.B. (1971). Pygmalion reaffirmed. [w:] Elashoff, J.D., Snow, R.E. (Eds.), *Pygmalion reconsidered* (pp.139-155). Worthington, OH: CA. Jones.
- Rosenthal, R., Rubin, D.B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.
- Rosenthal, R., Rubin, D.B. (1984). Multiple contrasts and ordered Bonferroni procedures. *Journal of Educational Psychology*, 76, 1028-1034.
- Rosnow, R.L. (1990). Teaching research ethics through role-play and discussion. *Teaching of Psychology*, 17, 179-181.
- Rosnow, R.L., Rosenthal, R. (1993). *Beginning behavioral research: A conceptual primer*. New York: Macmillan
- Seldmeier, P., Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309-316.
- Smith, M.L., Glass, G.V., Miller, T.I. (1980). *The benefits of psychotherapy*. Baltimore: John Hopkins University Press.
- Snedecor, G. W., Cochran, W.G. (1989). *Statistical methods* (8th ed.). Ames: Iowa State University Press.
- Tukey, J.W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Тлумачыла: *Karolina Krysińska*