

Witold Simon

**META-ANALIZA W BADANIACH NAD SKUTECZNOŚCIĄ PSYCHOTERAPII.
CZĘŚĆ II: RODZAJE WIELKOŚCI EFEKTU, BINOMINALNA WIELKOŚĆ EFEKTU,
TESTOWANIE HOMOGENICZNOŚCI, ZMIENNE MEDIUJĄCE I MODERUJĄCE**

**META-ANALYSIS IN PSYCHOTHERAPY OUTCOME RESEARCH.
PART II: TYPES OF EFFECT SIZE, BINOMIAL EFFECT SIZE DISPLAY, HOMOGENEITY
TESTING, MEDIATORS AND MODERATORS**

Klinika Nerwic Instytutu Psychiatrii i Neurologii, Warszawa.

Kierownik: prof. dr hab. med. Maria M. Siwiak-Kobayashi.

Streszczenie

Jest to drugi artykuł z cyklu poświęconego meta-analizie w badaniach nad skutecznością psychoterapii. W pracy przedstawiono różne rodzaje wielkości efektu oraz omówiono procedury wzajemnego przeliczania ich wartości. Zaprezentowano również sposób obliczania binominalnej wielkości efektu. Ponadto, zestawiono ze sobą wielkość efektu oraz istotność statystyczną. Zaprezentowano też procedurę testowania homogeniczności meta-analizy. Omówiono również rolę zmiennych mediujących i moderujących w meta-analizie.

Abstract

This is the second paper related to meta-analysis in psychotherapy outcome research. The different types of effect sizes are discussed here, which includes explanations of the procedures used to convert between those types. Also, binomial effect size display is presented. The comparison between effect sizes statistics and statistical significance is provided. Also, homogeneity testing procedure is outlined. Furthermore, the role of mediators and moderators in meta-analytic research is explained.

Słowa kluczowe: meta-analiza, psychoterapia, wielkość efektu, binominalna wielkość efektu, testowanie homogeniczności, zmienne mediujące i moderujące

Key words: meta-analysis, psychotherapy, effect size, binomial effect size display, homogeneity testing, mediators and moderators

Wstęp

Niniejszy artykuł stanowi drugą część cyklu poświęconego meta-analizie w badaniach nad skutecznością psychoterapii. Jest on w całości poświęcony procedurom związanym z analizą danych, a w szczególności: 1) rodzajom wielkości efektu, 2) procedurze testowania homogeniczności, 3) roli zmiennych mediujących i moderujących.

Analizowanie danych

Rodzaje wielkości efektu

Podstawową procedurą statystyczną stosowaną w meta-analizie jest miara wielkości efektu (*effect size*). Do najczęściej stosowanych wielkości efektu należą: wynikający z korelacji r Pearsona [1] oraz oparte o średnie d Cohena [1], g Hedgesa [2] i Δ Glassa [3]. Jako że dwa pierwsze są najczęściej stosowane,

omówiono je bardziej szczegółowo niż pozostałe.

Wielkość efektu d można obliczać na różne sposoby [4,5] w zależności od procedur statystycznych użytych w analizowanych badaniach, np. w oparciu o:

1) test t-Studenta:

$$d = \frac{\overline{M}_t - \overline{M}_c}{SD_{\text{łączone}}}, \text{ przy czym } SD_{\text{łączone}} = \sqrt{\frac{(n_t - 1)SD_t^2 + (n_c - 1)SD_c^2}{n_t + n_c}},$$

gdzie: d – wielkość efektu, M – średnie dla poszczególnych grup, $SD_{\text{łączone}}$ – odchylenie standardowe łączone, SD_t oraz SD_c – odchylenie standardowe grupy badanej i kontrolnej, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna;

2) test t-Studenta, gdy nie są dostępne ani odchylenia standardowe ani błędy standardowe:

$$d = tS \sqrt{\left(\frac{n_t + n_c}{n_t n_c}\right) \left(\frac{n_t + n_c}{n_t + n_c - 2}\right)},$$

gdzie: d – wielkość efektu, tS – test t-Studenta, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna.

3) test t-Studenta, gdy są dostępne błędy standardowe, ale nie odchylenia standardowe:

$$SD = SE \sqrt{n},$$

gdzie: SD – odchylenie standardowe, SE – błąd standardowy, n – wielkość próbeki.

4) test F- Snedecora:

$$d = \frac{\overline{M}_t - \overline{M}_c}{\sqrt{MSE \left(\frac{n_t + n_c - 2}{n_t + n_c}\right)}}$$

gdzie: d – wielkość efektu, MSE – średnia błędów kwadratowych, M – średnie dla poszczególnych grup, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna.

Warto przy tym pamiętać, że MSE nie umożliwia precyzyjnego obliczenia d , w przypadku gdy porównywane są więcej niż dwie grupy. W takim przypadku należy stosować standardowy wzór d ,

obliczany w oparciu o test t-Studenta.

5) test F- Snedecora, gdy nie jest dostępny średni błąd kwadratowy:

$$d = \sqrt{F \left(\frac{n_t + n_c}{n_t n_c} \right) \left(\frac{n_t + n_c}{n_t + n_c - 2} \right)},$$

gdzie: d – wartość wielkości efektu, F – wartość testu jednoczynnikowej analizy wariancji, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna.

Druga najczęściej wykorzystywana wielkość efektu r , znajduje zastosowanie dla danych ciągłych oraz binarnych. Również i ten typ wielkości efektu można obliczać w zależności od procedur statystycznych użytych w poszczególnych meta-analizowanych badaniach [5,6], np. w oparciu o:

1) test t-Studenta:

$$r = \frac{tS}{\sqrt{tS^2 + n_t + n_c - 2}},$$

gdzie: r – wielkość efektu, tS - test t-Studenta, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna.

2) test F- Snedecora:

$$|r| = \sqrt{\frac{F}{(F + n_t + n_c - 2)}},$$

gdzie: $|r|$ – wartość bezwzględna wielkości efektu, F – F-test jednoczynnikowej analizy wariancji, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna.

3) test chi-kwadrat:

$$|r| = \sqrt{\frac{\chi^2}{N}},$$

gdzie: $|r|$ – wartość bezwzględna wielkości efektu, χ^2 - test chi-kwadrat dla $df = 1$, N - całkowita wielkość próby. Uwaga: wzór ten należy stosować tylko dla tablic liczebności 2 x 2.

4) ciągle miary (dwuszeregowo punktowe r) [4]:

$$r = \frac{\overline{M}_t - \overline{M}_c / SD_{\text{łączone}}}{\sqrt{\left(\frac{\overline{M}_t - \overline{M}_c}{SD_{\text{łączone}}}\right)^2 + \frac{1}{p(1-p)}},$$

gdzie: r – wielkość efektu, SD - łączone odchylenie standardowe dla zmiennej zależnej, M - średnie dla poszczególnych grup, p - proporcja całościowej próby względem jednej z dwóch prób, t – grupa badana, c – grupa kontrolna.

Wielkość efektu typu r posiada następujące zalety w porównaniu z wielkością efektu typu d [7]: 1) r przedstawia związek pomiędzy dwoma poziomami zmiennych niezależnych a wynikami zmiennych zależnych, podczas gdy konwertowanie zmiennej ciągłej r na zmienną dychotomiczną jaką jest d powoduje utratę tej informacji; 2) r pozwala na analizowanie trendów więcej niż dwóch grup, podczas gdy d jest zawężone tylko do dwóch grup; 3) r jest łatwiej interpretowalne w porównaniu z d .

Cohen [1] zaproponował następującą kategoryzację bezwzględnych wartości r : mała wielkość efektu 0,10, średnia 0,30, duża wielkość efektu 0,50. Wartości r wahają się do -1 do 1, gdzie -1 oznacza idealnie negatywny liniowy związek, 1 oznacza idealny pozytywny liniowy związek, a 0 oznacza brak linearnej relacji między dwoma zmiennymi.

Główna zaleta d w porównaniu z r wynika z faktu, że wyniki obliczone w oparciu o d są publikowane o wiele częściej. Niektórzy [4,5,8] uważają d wręcz za standard w raportowaniu wyników meta-analizy. Cohen [1] zaproponował następujące punkty odcięcia dla wielkości efektu d : mała 0,20, średnia 0,50, duża wielkość efektu 0,80.

Inne stosunkowo często stosowane wielkości efektu to:

1) g Hedgesa [2] , obliczane według wzoru:

$$g = \frac{M_t - M_c}{SD_{\text{łączone}}}, \text{ przy czym } SD_{\text{łączone}} = \sqrt{\frac{(n_t - 1)SD_t^2 + (n_c - 1)SD_c^2}{n_t + n_c - 2}},$$

gdzie: g – wielkość efektu, M_t oraz M_c - średnie grup badanej i kontrolnej, SD - łączone odchylenie standardowe, SD_t oraz SD_c – odchylenie standardowe grupy badanej i kontrolnej, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna.

W przypadku test t-Studenta, g Hedgesa oblicza się według wzoru:

$$g = \frac{tS \sqrt{n_t + n_c}}{\sqrt{n_t n_c}},$$

gdzie: g – wielkość efektu, tS – test t-Studenta, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna;

2) Δ Glassa [3], obliczane według wzoru:

$$\Delta = \frac{M_t - M_c}{SD_c}, \text{ przy czym } SD_c = \sqrt{\frac{\sum (M_c - \bar{M}_c)^2}{n_c - 1}}$$

gdzie: Δ – wielkość efektu, M_t oraz M_c – średnie grup badanej i kontrolnej, M_c – średnia grupy kontrolnej zważona względem wielkości grupy kontrolnej, SD_c – odchylenie standardowe grupy kontrolnej, n_c – liczebność grupy kontrolnej, t – grupa badana, c – grupa kontrolna.

Stosowanie Δ Glassa zaleca się w przypadku porównywania skuteczności kilku różnych terapii względem tej samej grupy kontrolnej [9]. Korzystanie wyłącznie z odchylenia standardowego grupy kontrolnej zapobiega zaś potencjalnemu nadmiernemu zróżnicowaniu wielkości efektu, do którego dochodzi jeżeli odchylenia standardowe grup badanych i grupy kontrolnej są znacznie rozbieżne.

Przeliczanie między poszczególnymi typami wielkości efektu

Wzory służące przeliczanie między poszczególnymi typami wielkości efektu [6] są następujące:

$$\begin{aligned} r \leftarrow d & \quad r = \sqrt{\frac{d^2}{d^2 + 4}}, & r \leftarrow g & \quad r = \sqrt{\frac{g^2 n_t n_c}{g^2 n_t n_c + (n_t + n_c) df}}, \\ d \leftarrow r & \quad d = \frac{2r}{\sqrt{1 - r^2}}, & d \leftarrow g & \quad d = g \sqrt{\frac{n_t + n_c}{df}}, \\ g \leftarrow r & \quad g = \frac{r}{\sqrt{1 - r^2}} \sqrt{\frac{df (n_t + n_c)}{n_t n_c}}, & g \leftarrow d & \quad g = \frac{d}{\sqrt{\frac{n_t + n_c}{df}}}, \end{aligned}$$

gdzie: d , r , g – wielkości efektu, df – stopnie swobody, n_t oraz n_c – liczebności grupy badanej i kontrolnej, t – grupa badana, c – grupa kontrolna.

Ujmowanie wielkości efektu w postaci centyli i procentów

W dotychczasowych rozważaniach wielkość efektu była ujmowana jako numeryczne oszacowanie stopnia skuteczności psychoterapii. Tego rodzaju ujęcie jest pomocne, kiedy poszukuje się odpowiedzi na pytanie jak bardzo średnie grupy badanej oraz grupy kontrolnej różnią się od siebie.

Jeżeli natomiast pytanie dotyczy tego jaki procent statystycznych leczonych pacjentów jest w lepszym stanie niż nieleczeni respondenci, wówczas wielkość efektu przedstawia się np. w postaci centyli wyników osiągniętych przez przeciętnego pacjenta z grupy eksperymentalnej w zestawieniu z wynikami respondenta z grupy kontrolnej (Tabela 1.) [1]. Przykładowo, wielkość efektu 0,00 oznacza, że średnia dla leczonej grupy jest na poziomie 50-tego centyla grupy nieleczonej. Natomiast, wartość wielkości efektu 0,80 oznacza, że średnia grupy leczonej jest na poziomie 79-tego centyla grupy nieleczonej, itd.. Możliwe jest również przeliczenie wielkości efektu na procent pacjentów, których stan się poprawił w porównaniu z osobami nieleczonymi (Tabela 1.). I tak, np. wielkość efektu 0,00 oznacza, że dystrybucja wyników obu grup pokrywa się całkowicie, co implikuje, że zerowy procent przeciętnych leczonych pacjentów będzie w lepszym stanie niż nieleczeni respondenci. Natomiast wielkość efektu 1,00 oznacza, że statystyczny pacjent będzie miał się lepiej niż 55,4 % osób z nieleczonej grupy [1], itd.

Tabela 1. Przeliczenie wielkości efektu na procenty i centyle

Wielkość efektu	centyl	Procent rozdzielnej dystrybucji wyników
2.0	97.7	81.1%
1.9	97.1	79.4%
1.8	96.4	77.4%
1.7	95.5	75.4%
1.6	94.5	73.1%
1.5	93.3	70.7%
1.4	91.9	68.1%
1.3	90	65.3%
1.2	88	62.2%
1.1	86	58.9%
1.0	84	55.4%
0.9	82	51.6%
0.8	79	47.4%
0.7	76	43.0%
0.6	73	38.2%
0.5	69	33.0%
0.4	66	27.4%
0.3	62	21.3%
0.2	58	14.7%
0.1	54	7.7%
0.0	50	0%

Źródło: zaadaptowano z [1]

Binominalna wielkość efektu

Binominalna wielkość efektu, BESD (*binomial effect size display*) [10], jest kolejną procedurą umożliwiającą przełożenie numerycznych wielkości efektu na równoważny współczynnik sukcesu. Kontrastuje on wyniki grupy badanej i grupy porównawczej poprzez nałożenie dystrybucji wyników z obu grup. Procedura BESD opiera się na wielkości efektu typu r , dlatego należy przeliczyć inne wielkości efektu właśnie na ten rodzaj. Sukces terapeutyczny w BESD oblicza się według wzoru [6]:

$$0,50 + r/2,$$

a sukces warunków kontrolnych następująco:

$$0,50 - r/2.$$

Przykładowo, dla $r = 0,40$ wielkość sukcesu dla grupy eksperymentalnej wynosi 0,70, a dla warunków kontrolnych 0,30. Innymi słowy, współczynnik sukcesu danej formy terapii wynosi 70% dla grupy eksperymentalnej do 30% dla grupy kontrolnej.

Wielkość efektu a istotność statystyczna

Wielu badaczy przedmiotu twierdzi, że wielkość efektu umożliwia dokładniejszą, w porównaniu z istotnością statystyczną, ocenę skuteczności psychoterapii [1,3,4,6,7]. Statystyczny test istotności informuje bowiem o prawdopodobieństwie z jakim wyniki osiągnięte w grupie eksperymentalnej różnią się od przypadkowych rezultatów. Natomiast miara wielkości efektu wskazuje na relatywną wielkość uzyskanego wyniku. Obliczanie istotności statystycznej dla wielkości efektu jest oczywiście kluczowe dla oceny wartości uzyskanych wyników. Obie procedury, istotność statystyczna oraz wielkość efektu, odnoszą się do siebie w następujący sposób [6]:

$$\text{test istotności} = \text{wielkość efektu} \times \text{liczebność próby}.$$

W oparciu o powyższy schemat, przedstawiono poniżej zależności między istotnością statystyczną a poszczególnymi rodzajami wielkości efektu:

1) dla wielkości efektu r

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{df}$$

gdzie: t – test istotności, r – wielkość efektu, df – stopnie swobody;

2) dla wielkości efektu d

$$t = d \frac{\sqrt{df}}{2}$$

gdzie: t – test istotności, d – wielkość efektu, df – stopnie swobody;

3) oraz dla wielkości efektu g

$$t = \frac{M_t - M_c}{SD} \sqrt{\frac{n_t n_c}{n_t + n_c}}$$

gdzie: t – test istotności, M_t oraz M_c – średnie dla grup badanej i kontrolnej, SD – odchylenie standardowe, n_t oraz n_c – liczebność próby dla grup badanej i kontrolnej.

Analizując powyższe wzory, można dostrzec, że dwa badania, uwzględnione w jednej meta-analizie, charakteryzujące się dokładnie taką samą wielkością efektu, lecz różną liczebnością prób, mogą się znacznie różnić między sobą na poziomie istotności statystycznej. Innymi słowy, wyniki pochodzące z różnych badań, stanowiących podstawę danej meta-analizy, choć mogą być nieistotne statystycznie, to mogą dostarczać o wiele solidniejszych dowodów, niż pojedynczy statystycznie istotny wynik. Przykładowo, dwa wyniki na poziomie $p = 0,06$ mają o wiele silniejsze podstawy ($p = 0,014$) niż hipoteza zerowa weryfikowana przez jeden wynik na poziomie $p = 0,05$ [7].

Testowanie homogeniczności

Uważa się, że o stopniu rzetelności meta-analizy decyduje to, na ile spełnia ona kryteria homogeniczności (jednorodności) zmiennych [11]. Warto przy tym pamiętać, że heterogeniczność do pewnego stopnia jest naturalnym zjawiskiem odzwierciedlającym realia kliniczne. Przyczyny heterogeniczności mogą być różnorakie, począwszy od zmienności między badaniami (*between-study variability*) spowodowanej różnorodnością zmiennych klinicznych w poszczególnych badaniach (np. typem leczenia, długością terapii, operacjonalizacją oceny skuteczności leczenia, doświadczeniem terapeutów itp.), po zmienność wewnątrz badania (*within-study variability*) wynikającą np. z błędu losowego.

Homogeniczność można oszacować klinicznie (w oparciu o subiektywne porównywanie populacji, interwencji, wyników, itp.) lub statystycznie, analogicznie do sposobu testowania homogeniczności testów statystycznych. W tym drugim przypadku korzysta się przeważnie z:

1) testu Q [9,12]:

$$Q = \sum (\omega_i (ES_i - \overline{ES})^2)$$

gdzie: Q – test Q , \overline{ES} – zważona średnia d pochodząca z meta-analizowanych badań, ES_i – pojedyncza wielkość efektu dla każdego badania, ω_i – odwrotność wariancji dla każdej wielkości efektu, i – liczba badań uwzględnionych w meta-analizie.

Test Q opiera się na teście chi-kwadrat z $k-1$ stopniami swobody, gdzie k oznacza liczbę badań. Stosując test Q trzeba pamiętać, że informuje on nie tyle o stopniu heterogeniczności, ale raczej o statystycznej istotności. Jeżeli wartość testu Q przekroczy graniczną wartość dla testu chi-kwadrat, wtedy zerowa hipoteza homogeniczności jest odrzucona. Innymi słowy, statystycznie istotne wartości testu Q wskazują na heterogeniczną dystrybucję wyników. Należy przy tym pamiętać, że test Q ma

stosunkowo małą siłę [13], więc próg istotności jest przeważnie ustawiony wyżej, np. na poziomie 0,0001.

2) testu τ [12]:

$$\tau^2 = \frac{Q - (k-1)}{\sum \omega_i - \frac{\sum \omega_i^2}{\sum \omega_i}} \quad \text{dla } Q > (k-1),$$

$$\text{zaś } \tau^2 = 0 \text{ dla } Q \leq (k-1),$$

gdzie: τ - test τ , ω_i - zważony czynnik względem liczby badań, Q - test Q , k - liczba badań.

Test τ określa jak bardzo poszczególne wielkości efektu pochodzące z danego badania różnią się między sobą. Ponieważ jednak zależy on od statystyki używanej przy ocenie wielkości efektu, nie jest możliwe dokonywanie porównań pomiędzy badaniami, które stosowały różne sposoby obliczania wielkości efektu [9].

3) testu I^2 [14]:

$$I^2 = \frac{Q - (k-1)}{Q} \times 100\%, \quad \text{dla } Q > (k-1),$$

$$\text{zaś } I^2 = 0 \text{ dla } Q \leq (k-1),$$

gdzie: I^2 - test I^2 , Q - test Q , $k-1$ oznacza liczbę stopni swobody [9].

Test I^2 pozwala do pewnego stopnia ominąć ograniczenia testu Q oraz testu τ , ponieważ umożliwia on obliczenie procentu całkowitej różnorodności wielkości efektów, spowodowanej prawdziwą heterogenicznością zmiennych pochodzących z różnych badań. Innymi słowy, I^2 jest równoważne procentowi całkowitej wariancji wszystkich badań spowodowanej raczej heterogenicznością niż przypadkiem. Wartości I^2 lokują się między 0% a 100%. Wartość 0% oznacza potwierdzenie hipotezy homogeniczności, czyli brak heterogeniczności, co pozwala interpretować ew. obserwowalną różnorodność między wielkościami efektu jako spowodowaną tylko przez błąd losowy. Wartości 25%, 50%, 75% stanowią próg dla niskiej, średniej i wysokiej heterogeniczności [14].

Homogeniczność można również badać oceniając na ile badania statystycznie różnią się między sobą. Można to uczynić, np. w oparciu o:

1) odchylenia standardowe Z [8]:

$$Z = \frac{Z_1 - Z_2}{\sqrt{2}},$$

gdzie: Z – standardowe normalne odchylenie.

Podejście to stosuje się, gdy porównuje się tylko dwa poziomy istotności pochodzące z dwóch badań. Generalnie, odchylenie standardowe jest dobrą miarą zmienności wielkości efektu, gdyż nie zależy od wielkości próbki. Szczegółowe tabele z dystrybuantą Z , które są często niezbędne by uzyskać jednostronną istotność, są dostępne w literaturze [15].

2) test dyfuzji [8,16]:

$$\chi^2 = \sum (Z - \bar{Z})^2 \text{ dla } df = k - 1, \quad \left| \right.$$

gdzie: Z – standardowe normalne odchylenie, \bar{Z} – standardowe odchylenie zważone względem wielkości próby, df – stopnie swobody, k – ilość badań.

Test dyfuzji znajduje zastosowanie w przypadku testowania homogeniczności więcej niż dwóch niezależnych badań. Jeżeli badania wykazują znaczącą heterogeniczność, należy wtedy zbadać wartości odstające od dystrybucji oraz efekt zmiennych mediujących, które mogą się przyczyniać do heterogeniczności. Znaczenie zmiennych mediujących w meta-analizie przedstawiono poniżej.

3) Z -Fishera [15]:

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}, \quad \left| \right.$$

gdzie: Z – normalne odchylenie standardowe, z – Z -Fishera dla każdej grupy, n_1 oraz n_2 – liczebność poszczególnych grup.

Ta metoda umożliwia ocenę, na ile r pochodzące z poszczególnych badań różnią się statystycznie między sobą. W tym celu oblicza się wartości Z -Fishera odnoszące się do logarytmicznych transformacji r [16]. Przykładowo, jeżeli porówna się dwie wielkości efektu (np. 0,45 i 0,40), z dwóch różnych badań o takiej samej liczbie respondentów (np. po 120), to w oparciu o tabelę transformacji r w Z -Fishera [17], Z -Fishera dla każdej wielkości efektu wyniesie odpowiednio 0,485 oraz 0,424. W następnym kroku w oparciu o powyższy wzór, oblicza się normalne odchylenie standardowe, uzyskując w tym przypadku wartość: 0,47. Następnie określa się – w oparciu o stosowne tabele [18] – istotność jednostronną dla danego normalnego odchylenia standardowego, co w danym przypadku

wynosi 0,3192. Istotność dwustronną osiąga wtedy wartość 0,6384. Uzyskany w tym przypadku wynik wskazuje na to, że oba badania nie różnią się istotnie statystycznie między sobą.

Łączenie dwóch wielkości efektu z powyższego przykładu (0,45 i 0,40) oblicza się uzyskując średnią dla obu *Z-Fishera*, czyli $(0,485 + 0,424) / 2 = 0,45$. Po czym w oparciu o tabelę transformacji *Z-Fishera* w r [17], uzyskuje się estymację wielkości efektu dwóch łączonych badań, w tym przypadku wynoszącą 0,422.

Zmienne mediujące i moderujące

Potwierdzona statystycznie heterogeniczność implikuje konieczność zidentyfikowania możliwych zmiennych mediujących i moderujących [19]. Ich znajomość może również pomóc w spełnieniu postulatu zakładającego określenie „jaka forma psychoterapii prowadzona przez jakiego terapeuty i w jakich warunkach będzie najskuteczniejsza dla danej osoby ze specyficznymi problemami” [20].

Zmienne moderujące (*moderators*) są jakościowymi (np. rok publikacji, rasa respondentów, płeć badacza) lub ilościowymi (np. poziom nagrody dla uczestnika badań) zmiennymi wpływającymi na kierunek i/lub siłę relacji pomiędzy niezależnymi lub predykcyjnymi zmiennymi a zmienną zależną lub zmienną kryterium. Przykładem zmiennej moderującej jest doświadczenie zawodowe terapeuty, które nie wpływa istotnie na wyniki leczenia psychoterapią behawioralną, natomiast w zasadniczy sposób rozróżnia skuteczność terapii niebehawioralnych [9,21].

Zmienne mediujące (*mediators*) odnoszą się natomiast do mechanizmów przyczynowo-skutkowych i jako takie są zmiennymi przez które zmienne niezależne wpływają na zmienne zależne. Np. mediatorem wyniku psychoterapii często bywa proces psychoterapii lub stopień zgodności terapii z podręcznikiem [9,21].

Niektórzy autorzy [19] uważają, że zmienne mediujące i moderujące podważają istnienie tzw. Efektu Ptaka Dodo (*Dodo Bird Effect*) [22], zakładającego zbliżoną skuteczność wszystkich szkół psychoterapeutycznych. Tymczasem, po uwzględnieniu w modelu zmiennych mediujących i moderujących, Efekt Ptaka Dodo wydaje się pozostawać jedynie artefaktem, niesłusznie zakładającym, że orientacja terapeutyczna ma bezpośredni wpływ na wynik leczenia. W rzeczywistości, w modelu mediującym orientacja terapeutyczna wpływa na mediujące zmienne i dopiero poprzez nie na wynik terapii. Natomiast w modelu moderującym, orientacja wpływa na wynik leczenia, w stopniu zależnym od szeregu różnorodnych zmiennych [19]. Szczegółowe opisy procedur testujących zmienne mediujące i moderujące znajdują się w literaturze [19].

Adres do korespondencji:

Witold Simon,
Klinika Nerwic IPiN, ul. Sobieskiego 9,
02-957 Warszawa,
wsimon@ipin.edu.pl

Piśmiennictwo:

1. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum; 1988.
2. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*. 1981; 6 (2): 107–128.
3. Glass GV, McGaw B, Smith ML. Meta-analysis in social research. Newbury Park, CA: Sage; 1981.
4. Lipsey MW, Wilson DB. Practical meta-analysis. Thousand Oaks, CA: Sage Publication; 2001.
5. Thalheimer W, Cook S. How to calculate effect sizes from published research articles: A simplified methodology. Uzyskano 15 marca 2009 z http://work-learning.com/effect_sizes.htm
6. Rosenthal R. Parametric measures of effect size. W: Cooper H, Hedges LV. (red.). *The Handbook of research synthesis*. New York: Russel Sage Foundation; 1994, 231-244.
7. Rosenthal R, DiMatteo MR. Meta-analysis: Recent developments in quantitative methods for literature review. *Annual*

Review of Psychology. 2001; 52: 59-82.

8. Wolf FM. Meta-analysis: Quantitative methods for research synthesis. Newbury Park, CA: Sage Publication; 1986.
9. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistics or I-square index? Psychological Methods. 2006; 11 (2): 193-206.
10. Rosenthal R, Rubin DB. A simple, general purpose display of magnitude of experimental effect. Journal of Educational Psychology. 1983; 74 (2): 166-169.
11. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. British Medical Journal. 2004; 327: 557-560.
12. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954; 10: 101-129.
13. Thompson SG, Higgins JPT. Can meta-analysis help target interventions at individuals most likely to benefit? The Lancet. 2005; 365: 341-346.
14. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine. 2002; 21: 1539-1558.
15. Rosenthal R. Meta-analytic procedures for social research. Newbury, CA: Sage; 1991.
16. Rosenthal R. Assessing the statistical and social importance of the effects of psychotherapy. Journal of Consulting and Clinical Psychology. 1983; 51: 4-13.
17. Snedecor GW, Cochran WG. Statistical methods. Ames: Iowa State University; 1989.
18. Siegel S. Nonparametric statistics. New York: McGraw-Hill; 1956.
19. Shadish WR, Sweeney R. Mediators and moderators in meta-analysis: There is a reason we don't let dodo bird tell us which psychotherapies should have prizes. Journal of Consulting and Clinical Psychology. 1991; 6: 883-893.
20. Paul G. Strategy of outcome research in psychotherapy. Journal of Consulting Psychology. 1967; 31: 109-118.
21. Baron RM, Kenny D. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986; 51: 1173-1182.
22. Luborsky L, Singer B, Luborsky L. Comparative studies of psychotherapies: Is it true that "Everyone has won and all must have prizes?" Archives of General Psychiatry. 1975; 32: 995-1008.