PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 96
Badania marketingowe – nowe wyzwania 2010

**Małgorzata Rószkiewicz**
Szkoła Główna Handlowa w Warszawie

# MOŻLIWOŚCI I OGRANICZENIA METAANALIZY W OBSZARZE ILOŚCIOWYCH BADAŃ MARKETINGOWYCH

**Streszczenie:** Rozwój problematyki badań społecznych prowadzi do poszerzania się obszarów realizowanych badań empirycznych. Duża podaż wyników badań często odnoszących do tych samych zagadnień stwarza pytanie, jak prowadzić analizę sumaryczną, ustalającą, co w zasadzie wiadomo z przebadanych obszarów [Grave, Griffith 1971]. Realizowanie poszczególnych badań na próbach losowych sprawia, iż uzyskiwane wyniki nie są jednakowe. Ich zakres zmienności opisuje wariancja losowa estymatora. W artykule rozważono problemy związane z analizą fluktuacji wyników badań, których rozstrzygnięcie wyznacza paradygmat statystycznej metaanalizy.

**Słowa kluczowe:** analiza sumaryczna, spójność wyników, czynniki zakłócające, model statystyczny.

## 1. Istota i przesłanki rozwoju metaanalizy

Rozwój problematyki badań społecznych prowadzi do poszerzania się obszarów realizowanych badań empirycznych. Duża podaż wyników, odnoszących do tych samych zagadnień rodzi pytanie, jak prowadzić analizę sumaryczną, ustalającą, co w zasadzie wiadomo z przebadanych obszarów [Grave, Griffith 1971]. Zagadnienie to nie dotyczy jedynie nauk społecznych. Ma to miejsce w fizyce, chemii, biologii eksperymentalnej, medycynie, zdrowiu publicznym i w wielu innych dziedzinach. W każdym z tych obszarów badawczych akumulacja wyników badań prowadzi do rozwoju metod ilościowych odnoszących się do sposobów dokonywania tego typu syntezy [Cooper, Hedges 1994]. Do określenia tych metod używa się nazwy metaanaliza [Glass 1976].

Przegląd literatury sugeruje, że początki tego typu analiz sięgają lat siedemdziesiątych XX w. [Glass 1976]. Od tego czasu opublikowano bardzo wiele rezultatów metaanalizy, które odnosiły się do bardzo szerokiego spektrum zagadnień, w tym badań marketingowych, co wskazuje, że potrzebę prowadzenia tego typu analiz zgłaszają badacze niezależnie od domeny badań. Jedną z najczęściej przytaczanych przyczyn uzasadniających takie podejście badawcze jest wzrost mocy statystycznej

[Becker, Cohen 2003]. Jednakże obok opinii bardzo pozytywnych wyrażano również poglądy sceptyczne, głoszące, że metaanaliza jest swoistą alchemią, która nieistotne wyniki badań przekształca w wynik istotny [Feinstein 1995]. Warto zatem rozważyć, czym jest metaanaliza, a co nią nie jest, z jakimi problemami wiąże się to podeście badawcze w obszarze badań ilościowych.

## 2. Procedura metaanalizy

Przegląd literatury obejmującej pozycje publikowane od lat siedemdziesiątych prowadzi do dość zaskakującego spostrzeżenia. W literaturze tematu brakuje jednej powszechnie uznanej definicji metaanalizy, na którą powoływaliby się wszyscy autorzy. Większość z nich prezentuje swoją własną definicję, a zgłaszane propozycji obejmują zarówno bardzo ogólne stwierdzenia głoszące, że metaanaliza jest analizą rezultatów wielu badań w celu wyprowadzenia generalnych wniosków i konkluzji [Konstantopoulos 2006], jak i lakoniczne określające ją jako syntezę lub agregację wyników z domeny badań [Olkin 1995] lub po prostu analizę danych zastanych [Durlak, Lipsey 1991]. Taki stan rzeczy sprawia, że bardzo często metaanaliza mylona jest z tradycyjnym, narracyjnym przeglądem literatury.

Niewątpliwie w ramach metaanalizy taki przegląd literatury jest dokonywany, lecz zbieżność metaanalizy i tradycyjnego przeglądu literatury odnosi się jedynie do takich kwestii, jak:

– dokonanie podsumowania wyników z domeny badań,
– rozpoznanie, jak bardzo wyniki badań mogą się różnić ze względu na kluczowe cechy tych badań,
– sformułowanie rekomendacji dla przyszłych badań i wniosków dla praktyki.

Natomiast metaanalizę od tradycyjnego, narracyjnego przeglądu literatury odróżnia to, iż metaanaliza w zakresie badań ilościowych jest statystyczną analizą wyników badań zrealizowanych w ramach wyznaczonego obszaru. Celem metaanalizy w obszarze badań ilościowych jest bowiem:

– wyznaczenie ilościowej reprezentacji kluczowych wyników badań,
– dokonanie statystycznej analizy zróżnicowania wyników między badaniami,
– powiązanie wyników badań z cechami tych badań.

Przegląd literatury należy jedynie do wstępnego etapu procedury, która powinna być zrealizowana w ramach metaanalizy. Procedura ta obejmuje bowiem następującą sekwencję działań: sformułowanie pytań badawczych, przegląd literatury, przygotowanie danych do analizy, ocenę wyników udostępnionych do metaanalizy, estymację wyniku zagregowanego oraz konkluzje i interpretacje. W każdym z tych etapów badacz musi dokonać zasadniczych rozstrzygnięć.

W pierwszym etapie w ramach formułowania pytań badawczych należy przede wszystkim rozstrzygnąć, czym jest szczególne pytanie badawcze, formalna hipoteza lub zmienne określone wprost jako kluczowe dla obszaru badań. Ważne jest, by ustalić, jaka była podstawa ich formułowania lub identyfikacji, czy wynikały one

z rozważań teoretycznych i były w pełni zdefiniowane w literaturze tematu, czy też wywodzą się z dotychczasowych doświadczeń badawczych. Dokonując identyfikacji problemu badawczego, konieczne jest podanie kryteriów jego wyodrębniania.

W ramach przeglądu literatury należy rozstrzygnąć, czy na podstawie literatury można zidentyfikować wszystkie badania zrealizowane dotychczas w interesującym obszarze, uwzględniając badania, których wyniki opublikowano, oraz te, których wyników nie opublikowano. Konieczne jest ustalenie, czy badania te przeprowadzono na reprezentatywnych, nieobciążonych próbkach. Jeśli zaś wyniki badań nieopublikowanych są niedostępne, lecz w opinii ekspertów są znaczące w domenie badań, konieczne jest oszacowanie skutków ich pominięcia dla wyniku końcowego metaanalizy, analogicznie jak w badaniach reprezentacyjnych ocenia się efekt braków odpowiedzi. Przeglądu literatury należy dokonać wszelkimi dostępnymi metodami. Dostępne badania powinny zostać wyszczególnione wraz z ich charakterystyką.

Podczas prac związanych z przygotowaniem danych do analizy dokonywana jest procedura kodowania pozwalająca oznakować charakterystyczne cechy badań umożliwiające identyfikowanie w zagregowanych zbiorach dane pochodzące z poszczególnych badań lub ich wyniki cząstkowe.

W etapie obejmującym ocenę wyników udostępnionych do metaanalizy należy sprawdzić, czy miernik statystyczny szacowany podczas poszczególnych badań, a który jest podstawą uogólnień w metaanalizie, przedstawia rezultaty badań w sposób łatwy do zinterpretowania i jest porównywalny między tymi badaniami. Chodzi o to, by miernik ten przekładał wyniki badań na wspólną skalę, tak by były one łatwo interpretowane, porównywane i łączone. Postulat ten zostanie spełniony wówczas, gdy zmienne podlegające obserwacji i będące podstawą konstrukcji tego miernika będą we wszystkich badaniach jednakowo zoperacjonalizowane. Ważne jest też, by znany był rozkład jego estymatora, a także by znajomość jego wariancji umożliwiała wnioskowanie statystyczne. Postulat ten również jest na ogół spełniony, gdyż w większości przypadków przedmiotem metaanalizy jest analiza przyczynowości, w której wykorzystuje się najczęściej takie mierniki, które jako estymatory parametrów w populacji mają rozkład normalny lub asymptotycznie normalny. W zależności bowiem od tego, jak były skalowane zmienne dające podstawę do wyznaczenia miernika zależności jest to wskaźnik ilorazu szans, standaryzowana różnica, zwana również statystyką *d* Cohena, lub współczynnik korelacji po transformacji Fishera [Rószkiewicz 2002]. W takich przypadkach w mocy jest założenie, że w poszczególnych badaniach ($i = 1, 2, ..., k$) obserwowana jest statystyka $T_i$, co do której przyjmuje się, że ma rozkład normalny o wartości oczekiwanej $\Theta_i$ i znanej wariancji $\upsilon_i$, dla $i = 1, 2, \ldots, k$.

W ramach estymacji wyniku zagregowanego, czyli wyniku końcowego metaanalizy, należy rozstrzygnąć kwestie odnoszące się do zakresu zmienności wykorzystywanych wyników badań. Fakt, iż wykorzystywane w metaanalizie wyniki pochodzą z badań przeprowadzanych na próbach statystycznych różniących się nie tylko rozmiarem, ale i techniką losowania, sprawia, iż uzyskiwane wyniki nie są

jednakowe. Ich zakres zmienności opisuje błąd losowy estymatora. Uwidacznia się on np. w różnej rozpiętości przedziałów ufności otrzymywanych z kolejnych prób. Analiza tego zakresu dotyka kilku kwestii, których rozstrzygnięcie wyznacza paradygmat statystycznej metaanalizy.

Po pierwsze, jeśli wszystkie zestawiane w metaanalizie badania mają szacować wartość tego samego parametru w populacji, to wydaje się rozsądne, by kombinacja oszacowań z poszczególnych badań (np. biorąc ich wartość średnią) redukowała całkowitą nieprzewidywalność losową prób przez wyrównanie fluktuacji poszczególnych badań. Po drugie, różna dokładność poszczególnych badań sugeruje ponadto, by obliczając efekt średni z kilku badań, brać pod uwagę wielkość wariancji błędu i nadawać większą wagę tym wynikom, które są osiągane z większą precyzją. Po trzecie, podczas oceny przedziałów ufności uzyskanych z poszczególnych badań często można odnotować ich pokrywanie się, lecz w niektórych przypadkach otrzymane oszacowania mogą się znajdować poza przedziałami wyznaczonymi na podstawie wyników innych badań. Nasuwa się zatem pytanie, czy wyniki otrzymane w tego typu badaniach mogą się różnić bardziej niż wynikałoby to z wartości wariancji błędu. Innymi słowy, czy zasadne jest założenie, że wszystkie te badania szacują tę samą wartość parametru i różnią się w tych oszacowaniach tylko wariancją błędu próby? A jeśli wyniki z niektórych badań różnią się znacznie od pozostałych, to czy należy traktować je jako faktyczną odmiennością, czy też można je pominąć, a jeśli tak się stanie, to czy regularności pozostałych badań będą bardziej zwarte i logiczne? Po czwarte wreszcie, obserwowane odmienności wyników poszczególnych badań, np. jeśli są rozłożone w czasie, mogą być wynikiem zmian systematycznych, np. trendu. Powstaje zatem pytanie, w jaki sposób dokonać rozróżnienia, czy trend ten jest statystycznie istotny, czy też jest tylko artefaktem wywołany wariancją losową?

Pierwsze ze sformułowanych pytań odnosi się do sposobu oszacowania wartości parametru na podstawie wyników badań. Jednym z podstawowych celów metaanalizy jest bowiem oszacowanie wartości parametru populacji, a ponadto wzrost precyzji tej estymacji. By cele te osiągnąć, metaanalizę można prowadzić na dwa sposoby. Można oprzeć analizę na założeniu, że wszystkie wykorzystywane w analizie próby pochodzą z tej samej populacji o parametrze $\Theta$. Parametr ten jest szacowany na podstawie estymatora $T$, który w kolejnych badaniach przyjmuje wartość $T_i$, czyli:

$$T_i = \theta + \varepsilon_i, \tag{1}$$

gdzie $\varepsilon_i$ ma rozkład normalny z wartością oczekiwaną 0 i wariancją $v_i$.

Takie sformułowanie problemu jest tożsame z uznaniem, że badane zagadnienie opisuje model z efektami ustalonymi. W modelu tym najlepszym estymatorem parametru $\Theta$, wyznaczonym MNW i MNK jest średnia ważona wyników estymacji z poszczególnych prób. Ważony średni wynik wyznacza formuła (za: [Hedges, Olkin 1985]):

$$\overline{T} = \left( \frac{1}{\sum w_i} \right) \sum w_i T_i, \tag{2}$$

gdzie $w_i = \frac{1}{\upsilon_i}$, czyli wyniki o mniejszej wariancji błędu, znaczą (ważą) więcej w estymacji parametru niż wyniki o wariancji większej.

Z kolei wariancja błędu ważonego średniego wyniku jest szacowana jako:

$$V_{.} = \frac{1}{\sum w_i}, \tag{3}$$

co pozwala na estymację przedziałową wartości parametru populacji.

Jeśli liczba prób wzrasta, to wariancja błędu oceny średniej maleje i w konsekwencji maleje jej błąd standardowy, co zwiększa precyzję estymacji. Uzyskanie tego typu rezultatu uznaje się za główne osiągnięcie metaanalizy. Badanie istotności wyniku opiera się na statystyce, która przy ogólnym standardowym założeniu ma rozkład normalny $N$ (0, 1) przy prawdziwości hipotezy zerowej, że wartość parametru w populacji wynosi 0. Statystyka testująca przyjmuje postać:

$$z = \frac{\overline{T}}{\sqrt{V_{.}}}. \tag{4}$$

Jeśli hipoteza zerowa jest fałszywa i parametr populacji przyjmuje wartość różną od zera, np. $\Theta_1$, to wartość średnia statystyki testującej wynosi wówczas (za: [Hedges, Pigott 2001]):

$$\lambda = \frac{1}{\sqrt{V_{.}}}(\theta_1 - \theta). \tag{5}$$

Hedges i Pigott wykazali, że moc dwustronnego testu sprawdzającego wartość $\Theta_1$ wynosi:

$$1 - \beta = 1 - \Phi(z_{\alpha/2} - \lambda) + \Phi(-z_{\alpha/2} - \lambda), \tag{6}$$

gdzie: $\Phi$ jest dystrybuantą standaryzowanego rozkładu normalnego, $z_{\alpha/2}$ jest wartością krytyczną dwustronnego obszaru o poziomie istotności $\alpha$.

Biorąc pod uwagę tę formułę, należy stwierdzić, że moc testu wzrasta wraz ze wzrostem $\lambda$, co może być spowodowane zarówno spadkiem zmienności losowej (czyli $V_{.}$), jak i spadkiem różnicy między $\Theta_1$ i $\Theta$.

Metaanalizę można również oprzeć na założeniu, że wyniki osiągane z prób losowych są realizacjami zmiennej losowej o określonym rozkładzie. W takim przypadku przyjmuje się, że $\Theta_i$ jest nieznanym parametrem w $i$-tym badaniu, $T_i$ jest zaś

jego oszacowaniem na podstawie *i*-tego badania, a $\upsilon_i$ jest jego wariancją losową. Zatem dane ze zbioru *k* badań generują oszacowania $T_1$, $T_2$, …$T_k$ i ich wariancje $\upsilon_1$, $\upsilon_2$, …, $\upsilon_k$. Takie sformułowanie problemu jest tożsame z modelem o efektach losowych. Naturalnym sposobem prezentacji tego typu zagadnienia jest dwupoziomowy hierarchiczny model, w którym pierwszy poziom wyznaczają wyniki określonego badania, a drugi poziomi odnosi się do zmienności wyniku między badaniami. Na pierwszym poziomie wartość parametru jest szacowana jako $T_i$, przyjmując, że jest to wartość parametru skorygowana błędem próby, czyli:

$$T_i = \theta_i + \varepsilon_i, \tag{7}$$

gdzie $\varepsilon_i$ ma rozkład normalny z wartością oczekiwaną 0 i wariancją $\upsilon_i$ .

Parametr Θ jest średnim poziomem wartości parametrów szacowanych na podstawie wszystkich badań. W przypadku określonego zbioru *k* badań chodzi o zbiór parametrów $\Theta_1$, $\Theta_2$, …, $\Theta_k$, które są szacowane w poszczególnych badaniach. Przy tym nie chodzi o zwykłe uśrednienie tych próbkowanych parametrów, lecz o przestrzeń wszystkich prób. Na drugim poziomie, poziomie między badaniami, wartości parametrów są kształtowane przez średni poziom Θ plus specyficzny dla badania efekt losowy $\eta_i$, czyli:

$$\theta_i = \theta + \eta_i, \tag{8}$$

gdzie $\eta_i$ ma rozkład normalny z wartością oczekiwaną 0 i wariancją $\tau^2$ .

W tym modelu $\eta_i$ opisuje różnice między wartościami parametrów szacowanych w poszczególnych badaniach. Parametr $\tau^2$ jest często określany jako wariancja międzygrupowa, jeśli się przyjmie, że grupy wyznaczają kolejne badania. W modelu z efektami losowymi przedmiotem estymacji jest zarówno wartość średnia rozkładu estymatora, jak i jego wariancja. W modelach z efektami losowymi rozważa się zatem dwa źródła zmienności obserwowanych wartości z prób, tj. zmienność błędu losowego próby i zmienność wyników estymacji otrzymywanych z poszczególnych prób.

Model o efektach losowych jest identyczny z hierarchicznym modelem linowym często stosowanym w analizach danych społecznych, jednakże ma dwie cechy, które go od tego podstawowego modelu odróżniają. Po pierwsze, w tradycyjnym hierarchicznym modelu liniowym zakłada się, że wariancja losowa na pierwszym poziomie jest identyczna dla wszystkich jednostek pierwszego poziomu. W modelu metaanalizy wariancje te są różne dla poszczególnych badań, co uwidacznia się w różnych wartościach wariancji losowych ($\upsilon_i$). Po drugie, w tradycyjnym modelu wariancja ta jest nieznana i musi być szacowana na podstawie danych z próby. W modelu metaanalizy przyjmuje się, że jest ona znana. Model dwupoziomowy metaanalizy można zatem zapisać jako równanie:

$$T_i = \theta + \eta_i + \varepsilon_i = \theta + \xi_i, \tag{9}$$

gdzie $\xi_i$ jest sumą efektów losowych ($\xi_i = \eta_i + \varepsilon_i$).

Wynik z każdego badania jest oszacowaniem Θ z wariancją, która zależy zarówno od $\upsilon_i$, jak i od $\tau^2$. W podejściu tego typu konieczne jest rozróżnienie między wariancją błędu, która odnosi się do założenia o ustalonej wartości parametru $\Theta_i$ (zwaną wariancją warunkową błędu i oznaczaną jako $\upsilon_i$) i wariancją związaną ze zmiennością $\Theta_i$ (zwaną wariancją bezwarunkową błędu i oznaczaną jako $\upsilon_i^* = \upsilon_i + \tau^2$, gdyż co do błędu próby $\varepsilon_i$ i efektu losowego $\eta_i$ przyjmuje się, że są niezależne). Estymatory MNW i MNK parametru Θ definiują wzór (za: [Hedges, Vevea 1998]):

$$\overline{T}_{\cdot} = \frac{\sum_{i=1}^{k} w_i^* T_i}{\sum_{i=1}^{k} w_i^*}, \tag{10}$$

gdzie: $w_i^* = \dfrac{1}{(\upsilon_i + \hat{\tau}^2)} = \dfrac{1}{\upsilon_i^*}$.

Warto zauważyć, że w modelu o efektach losowych estymator jest również średnią ważoną wartości $T_i$; wagami są wartości odwrotne do wariancja błędu, co oznacza, że większą wagę nadaje się tym wynikom, które są bardziej precyzyjne. Błąd tego estymatora wyznacza formuła:

$$\sqrt{\upsilon_{\bullet}^*} = \sqrt{\left(\sum_{i=1}^{k} w_i^*\right)^{-1}}. \tag{11}$$

W związku z tym, że w tak zdefiniowanym modelu estymator parametru Θ ma rozkład normalny, można dokonać estymacji przedziałowej wartości parametru oraz poddawać ocenie istotność uzyskanego wyniku. Sprawdzianem testu jest statystyka, która przy prawdziwości hipotezy zerowej, że wartość parametru wynosi 0, ma rozkład *t*-Studenta o ($k$ – 1) stopniach swobody.

Wariancja międzygrupowa $\tau^2$ jest szacowana na podstawie danych z prób. Estymacja wariancji międzygrupowej może przebiegać przy założeniu określonego rozkładu efektów losowych. Wówczas rekomendowana jest procedura oparta na metodzie największej wiarygodności. Jeśli zaś pominie się założenie dotyczące kształtu rozkładu tych efektów, to rekomendowana jest procedura nieparametryczna. Trudności z weryfikacją założeń dotyczących rozkładu efektów losowych sprawiają, że najczęściej stosowane jest podejście nieparametryczne wykorzystujące statystykę *Q* Cochrana według formuły:

$$\hat{\tau}^2 = \begin{cases} \dfrac{Q-(k-1)}{\sum_{i=1}^{k} w_i - \dfrac{\sum_{i=1}^{k} w^2{}_i}{\sum_{i=1}^{k} w_i}} & \text{dla } Q \geq (k-1) \\ 0 & \text{dla } Q < (k-1) \end{cases} \tag{12}$$

$$Q = \sum_{i=1}^{k} \frac{\left(T_i - \overline{T}_.\right)^2}{\upsilon_i}, \tag{13}$$

gdzie: $w_i = 1/\upsilon_i$,

$\overline{T}_.$ jest wartością parametru Θ, przy założeniu zerowania się wariancji międzygrupowej, czyli gdy: $\tau^2 = 0$.

Jeśli zatem wartość statystyki $Q$ przekracza liczbę stopni swobody, to uznaje się wariancję międzygrupową za różną od zera.

Jak już wspomniano, w modelu o efektach losowych występują dwa źródła zmienności powodujące różnice między obserwowanymi wynikami z prób, tj. błędy próby oraz zmienność pomiędzy wartościami, które są szacowane. Jeśli wariancja międzygrupowa oszacowana na podstawie $k + 1$ badań wzrośnie w stosunku do wariancji międzygrupowej oszacowanej na podstawie $k$ badań, to błąd standardowy szacowania wartości średniej może być wyższy, powodując spadek mocy statystycznej. Dlatego też w modelu z efektami losowymi wzrost liczby badań nie musi powodować spadku standardowego błędu i wzrostu mocy statystycznej wnioskowania.

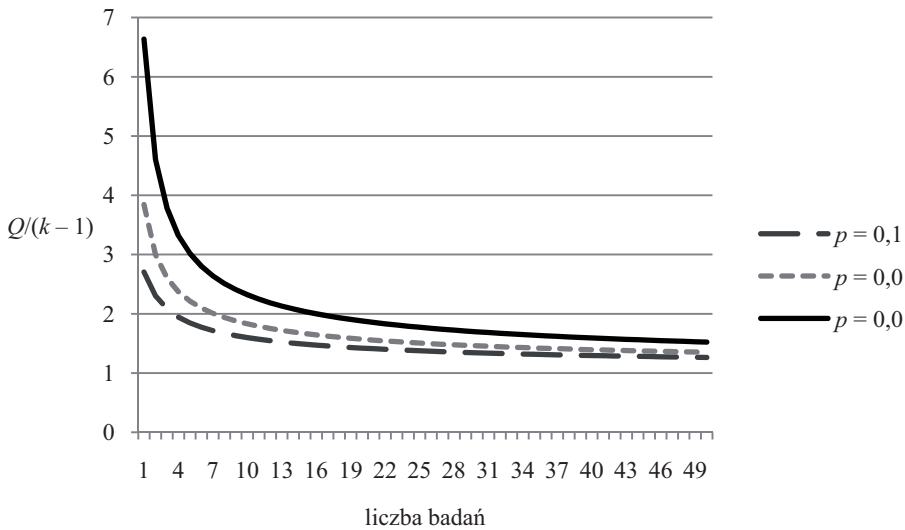
Powiększanie wariancji losowej próby ($\upsilon_i$) o wariancję efektu losowego (oszacowaną wartość $\tau^2$) prowadzi do dwóch widocznych konsekwencji. Po pierwsze, powoduje wzrost wartości błędu szacunku wyniku końcowego, czyli wzrost $\sqrt{V_.}$ do $\sqrt{V_.^*}$, po drugie zaś, poszerza przedział ufności i tym samym redukuje prawdopodobieństwo wystąpienia błędu pierwszego rodzaju. Redukcja prawdopodobieństwa wystąpienia błędu pierwszego rodzaju jest główną przyczyną postulowania podejścia wykorzystującego efekty losowe zamiast podejścia z efektami ustalonymi. Jak pokazują Hunter i Schmidt, dla heterogenicznych wyników z poszczególnych badań podejście oparte na ustalonych efektach może powodować ryzyko błędu pierwszego rodzaju 10-krotnie większe niż zalecane 0,05 (por. [Hunter, Schmidt 2000, s. 279]). Ponadto szerszy przedział ufności, który jest rezultatem podejścia opartego na efektach losowych, ogranicza precyzję estymacji w stosunku do precyzji oczekiwanej. Może to skutkować tym, że w analizie zakładającej efekty losowe uzyska się uogólnienie wykraczające poza zakres wyników badań wykorzystanych do tego uogólnienia. Analiza z efektami ustalonymi jest bardziej restrykcyjna i dopuszcza wnioskowanie jedynie w zakresie uzyskiwanych wyników. Na tej podstawie Hun-

ter i Schmidt sugerują, że modele z efektami ustalonymi rzadko, jeśli w ogóle, są właściwe dla rzeczywistych danych wykorzystywanych do metaanalizy i dlatego też, zdaniem tych autorów, powinny być preferowane modele oparte na efektach losowych.

Warto jednak w tym miejscu odnotować ograniczenia podejścia opartego na efektach losowych. Po pierwsze, badania wykorzystywane w metaanalizie prawdopodobnie nie zawsze są badaniami na próbach losowych z obszaru będącego domeną badań, co narusza podstawowe założenie modelu o efektach losowych. Ponadto badania te nie są losową reprezentacją badań prowadzonych w obszarze domeny, co ponownie narusza założenie modelu o efektach losowych. Z tych powodów podejście oparte na efektach losowych dostarcza błędnego wnioskowania. Ważniejsze jest jednak, że podejście oparte na efektach ustalonych i podejście oparte na efektach losowych są adresowane do fundamentalnie innych pytań badawczych, których istota często nie jest rozróżniana. W podejściu opartym na efektach ustalonych pytanie badawcze dotyczy najlepszego oszacowania parametru populacji odnoszącego się do pewnego zagadnienia (np. siły zależności) oraz rozstrzygnięcia, czy z punktu widzenia praktycznego lub teoretycznego jest ono ważne. Hipoteza zerowa zakłada wówczas, że parametr $\Theta = 0$. Oszacowanie średniej z wyników poszczególnych badań oraz zbudowanie przedziału ufności rozstrzyga tę kwestię. Natomiast w podejściu opartym na efektach losowych celem nie jest oszacowanie pojedynczego wyniku dla populacji, gdyż model ten zakłada rozkład wyników parametru populacji, z którego badana jest próba losowa. Dlatego też wyznaczenie wartości średniej tego rozkładu dostarcza innych informacji niż obliczenie wartości średniej z wyników prób losowych. Pytanie badawcze, na które odpowiada model z efektami losowymi, dotyczy zakresu rozkładu wyników dla populacji oraz cech tego rozkładu.

Kluczową częścią metaanalizy jest ocena spójności wyników poszczególnych badań, od której zależy wybór modelu statystycznego metaanalizy. Jeśli nie wiemy, na ile spójne są wyniki badań, nie możemy dokonywać ich uogólnienia. Statystyką testującą istotność wariancji międzygrupowej przy założeniu homogeniczności wyników badań jest zdefiniowana wzorem (13) statystyka $Q$ Cochrana. Statystyka ta, przy prawdziwości hipotezy zerowej, ma rozkład chi-kwadrat o $(k - 1)$ stopniach swobody. Warto jednak pamiętać, że im mniejszy zbiór badań (małe $k$), tym mniejsza moc tego testu [Sterne, Egger 2001]. Jak ukazuje rys. 1, test ten zwiększa swą moc, wykazując z kolei nadmierną skłonność do odrzucania hipotezy zerowej i wskazywania na heterogeniczność wyników badań, jedynie gdy liczba badań jest znaczna, a szczególnie gdy badania te są realizowane na dużych próbach. Moc ta jest słaba również wówczas, gdy wariancje wewnątrzgrupowe $\upsilon_i$ są znaczne. Może zatem się zdarzyć, że mimo że nie będzie podstaw do odrzucenia hipotezy zerowej, to jednak może wystąpić widoczna (znaczna) wartość $\tau^2 > 0$.

Jak wcześniej zaznaczono, gdy wartość statystyki $Q$ nie przekracza liczby stopni swobody, przyjmuje się homogeniczność międzygrupową wyników badań i tym samym procedura przebiega według modelu o efektach ustalonych, co może być




**Rys. 1.** Relacja krytycznej wartości statystyki $Q$ zgodnie z rozkładem chi-kwadrat względem liczby stopni swobody ($k$ – 1) dla poziomu istotności 0,01, 0,05 i 0,1

Źródło: opracowanie własne.

jedynie efektem statystycznym niezgodnym z faktycznym stanem rzeczy. Chcąc przełamać te ograniczenia Higgins, Thompson, Deeks, Altman [Higgins i in. 2003; Higgins, Thompson 2002] zaproponowali podejście umożliwiające pomiar efektu heterogeniczności wyników badań, proponując miarę stopnia niespójności. Wskaźnik oznaczony jako $I^2$ opisuje procent całkowitej zmienności między badaniami (zmienność międzygrupową), która jest spowodowana heterogenicznością w większym stopniu niż przypadkową. Wskaźnik ten wyznacza formuła:

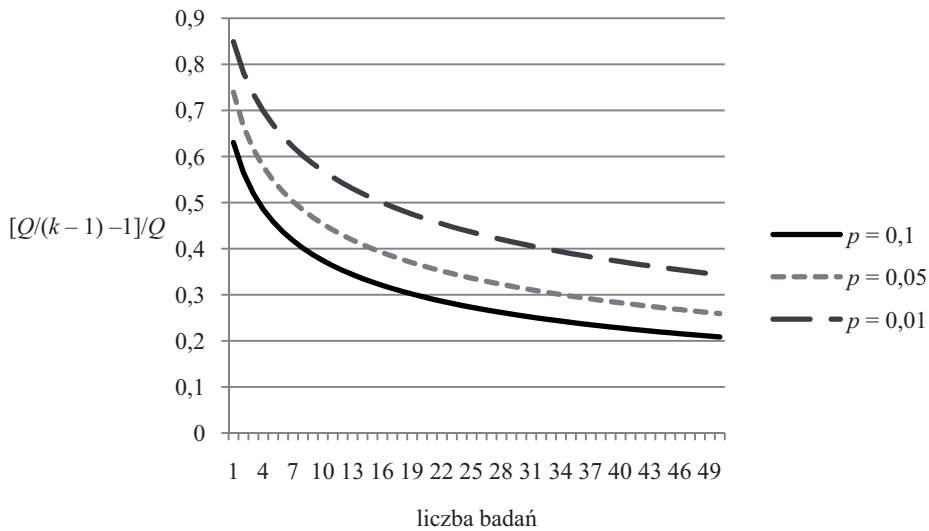
$$I^2 = 100\% \times \frac{(Q - df)}{Q}, \tag{14}$$

gdzie: $Q$ jest statystyką Cochrana obliczoną według wzoru (13),
$df = (k - 1)$ to liczba stopni swobody.

Zaproponowana miara uwalnia ocenę heterogeniczności wyników badań i tym samym wybór modelu statystycznego metaanalizy od sposobu skalowania analizowanego wyniku oraz od rezultatu testu statystycznego, który jak podkreśla Freeman, może być dyskusyjny [Freeman 1989].

Warto w tym miejscu zaznaczyć, że konstrukcja tego miernika jest zbieżna z konstrukcją współczynnika korelacji międzyklasowej opisującego zakres relacji międzygrupowych. Wartość tego współczynnika powyżej 0,1 skłania do uznania heterogeniczności międzygrupowej i tym samym uznawania zróżnicowania nie tylko

na poziomie jednostek, ale i na poziomie grup jednostek. Jeśli zatem przyjmie się tę graniczną wartość za kryterium oceny stopnia heterogeniczności wyników badań, to, jak pokazuje to rys. 2, przy małej liczbie badań wartości proponowanego wskaźnika $I^2$ przekroczą tę wartość progową, skłaniając do wyboru modelu z efektami losowymi jako właściwego do opisu wyników metaanalizy, mimo że wyniki testu $Q$ będą sugerować nieistotność tego typu prawidłowości.




**Rys. 2.** Wartości wskaźnika heterogeniczności wyników badań dla krytycznych wartości statystyki $Q$ względem liczby stopni swobody (liczby badań) dla trzech poziomów istotności: 0,01, 0,05 i 0,1

Źródło: opracowanie własne.

Wystąpienie zmienności międzygrupowej skłania do poszukiwania przyczyn takiego stanu rzeczy. Kwestia ta odnosi się do zależności wyników badań względem kluczowych cech tych badań, którym przypisuje się rolę sprawczą tej zmienności. Chodzi tu o rozstrzygnięcie, na ile specyficzne cechy poszczególnych badań wpływają na ich wyniki i czy odmienności obserwowane między poszczególnymi badaniami powodują efekty systematyczne, czy też artefakty wynikające ze zmienności losowej poszczególnych badań. Wybór procedury oceniającej tego typu oddziaływanie zależy od sposobu skalowania kluczowych cech badań, które podejrzewa się, iż mogą mieć wpływ na ich wyniki. Jeśli cechy te są nominalne, to po dokonaniu klasyfikacji wyników według tych cech można dokonać dekompozycji wartości statystyki $Q$ na składnik odnoszący się do zmienności utworzonych wewnątrz grup oraz składnik odnoszący się do zmienności między tak utworzonymi grupami, analogicznie do dekompozycji sumy kwadratów różnic, która ma miejsce w analizie wariancji, gdyż licznik statystyki $Q$ jest tego typu wielkością. Ocena istotności składnika

tej dekompozycji odnoszącego się do różnic międzygrupowych rozstrzyga kwestię wpływu cech nominalnych badań na różnicowanie się ich wyników, pod warunkiem że składnik odnoszący się do różnic wewnątrzgrupowych okaże się nieistotny. Jeśli tak się nie stanie, to konieczne jest poszukiwanie dodatkowych cech badań, które pozwoliłyby osiągnąć nieistotność zróżnicowania międzygrupowego przy istotności zróżnicowania międzygrupowego. Jeśli zaś kluczowe cechy badań skalowane są na skalach mocnych (są mierzalne), to najdogodniejszą procedurą statystyczną rozstrzygającą tę kwestię są modele mieszane.

Modele mieszane wykorzystywane w metaanalizie należą do grupy ogólnych liniowych modeli mieszanych. Jeśli się założy, że efekty ustalone są opisane za pomocą $p$ predyktorów $X_1, X_2, ..., X_p$, które są powiązane liniowo z wynikiem badania (cechy badania), to model drugiego poziomu dla wyniku pochodzącego z $i$-tego badania przyjmie postać:

$$T_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \eta_i, \tag{15}$$

gdzie: $x_{ij}$ to poziom $j$-tego predyktora w $i$-tym badaniu,
$\eta_i$ to efekt losowy dla $i$-tego badania, o rozkładzie $N(0, \tau^2)$.

Wówczas model dwupoziomowy przy założeniu efektów losowych przyjmuje postać:

$$T_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \eta_i + \varepsilon_i \tag{16}$$

lub

$$T_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \xi_i. \tag{17}$$

Jeśli $\tau^2$ byłoby znane, to współczynniki regresji mogłyby być szacowane ważoną metodą najmniejszych kwadratów, a ich istotność rozstrzygałaby o efektach systematycznych związanych ze specyfiką poszczególnych badań. Na ogół jednak $\tau^2$ nie jest znane i wówczas możliwe są cztery sposoby postępowania:

1. Oszacowanie $\tau^2$ na podstawie dostępnych wyników badań i wykorzystanie oceny tej wariancji do estymacji współczynników regresji ważoną metodą najmniejszych kwadratów, co umożliwi weryfikację istotności efektów systematycznych poszczególnych badań.

2. Łączne oszacowanie wektora współczynników i wariancji $\tau^2$ za pomocą REML nierestrykcyjnej metody największej wiarygodności.

3. Przyjęcie dla tej wariancji wartości zero (jak to ma miejsce w analizie efektów ustalonych), co jak już wspomniano, może powodować obciążenie wyników, ale jest ono tym mniejsze, im liczniejszy jest zbiór dostępnych badań.

4. Oszacowanie wektora współczynników regresji dla każdej wartości $\tau^2$ z przedziału prawdopodobnych wartości i użycie ważonej średniej dla tych wyników,

przyjmując za wagi wartości odpowiednie do prawdopodobieństw, tak jak to ma miejsce podejściu bayesowskim.

Ostatnim etapem metaanalizy jest sformułowanie konkluzji i interpretacja wyniku. Treści tych konkluzji powinny się odnosić do zagadnień mocy statystycznej wyników, a także zawierać krytyczną ocenę wykorzystanych badań ze względu na przyjmowane założenia, zrealizowane procedury oraz otrzymane wyniki, by na tej podstawie wypracować modyfikację postępowania badawczego w przyszłości.

## 3. Podsumowanie

Podsumowując, warto podkreślić korzyści analityczne, jakich dostarcza metaanaliza, a które można uznać za szczególne, gdyż są niedostępne w incydentalnych badaniach przekrojowych. Po pierwsze stwarza ona możliwość wykorzystania pełnego zakresu informacji wynikowych zrealizowanych badań niezależnie od tego, czy wyniki te okazały się znamienne statystycznie, czy też nie. Po drugie, umożliwia prowadzenie bardziej szczegółowych podziałów i klasyfikacji poprzez integrację cząstkowych zbiorów danych w jeden zbiór o znacznym rozmiarze. Ponadto, o czym już wspomniano, stwarza możliwość wychwycenia czynników zakłócających procesy badawcze i wskazania ścieżek modyfikacji postępowania w przyszłych badania na podstawie krytycznej oceny dostępnych badań pod względem przyjmowanych w nich założeń, realizowanych procedur i otrzymywanych wyników.

## Literatura

Becker B.J., Cohen L.D., *How meta-analysis increases statistical Power*, "Psychological Methods" 2003, 3, s. 243-253.

Cooper H.M., Hedges L.V. (red.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York 1994.

Durlak J.A., Lipsey M.W., *A Practitioner's Guide to Meta-analysis*, "American Journal of Community Psychology" 1991, vol. 19, no 3.

Feinstein A.R., *Meta-analysis: Statistical alchemy for the 21st century*, "Journal of Clinical Epidemiology" 1995, no 48, s. 71-79.

Freeman P.R., *The performance of the two-stage analysis of two-treatment, two-period crossover trials*, "Statistics in Medicine" 1989, 8, s. 1421-1432.

Glass G.V., *Primary, secondary, and meta-analysis of research*, "Education Researcher" 1976, no 5, s. 3-8.

Grave W., Griffith B., *Scientific communication: Its role in the conduct of research and creation of knowledge*, "American Psychologist" 1971, no 26, s. 349-361.

Hedges L.V., Olkin I., *Statistical methods for meta-analysis*, Academic Press, New York 1985.

Hedges L.V., Vevea J.L., *Fixed and random effects models in meta-analysis*, "Psychological Methods" 1998, no 3, s. 486-504.

Hedges L.V., Pigott T.D., *The power of statistical test in meta-analysis*, "Psychological Methods" 2001, no. 6, s. 203-217.

Higgins J.P.T., Thompson S.G., *Quantifying heterogeneity in a meta-analysis*, "Statistics in Medicine" 2002, 21, s. 1539-1558.

Higgins J.P.T., Thompson S.G., Deeks J.J., Altman D.G., *Measuring inconsistency in meta-analyses*, "BMJ" 2003, 327, s. 557-560.

Hunter J.E., Schmidt F.L., *Fixed effects vs. random effects meta-analysis models: Implication for cumulative research knowledge*, "International Journal of Selection and Assessment" 2000, no 8, s. 275-292.

Konstantopoulos S., *Fixed and mixed effects models in meta-analysis*, Discussion Paper Series, IZA DP, 2006, no 2198.

Olkin I., *Statistical and theoretical consideration in meta-analysis*, "Journal of Clinical Epidemiology" 1995, vol. 48, no 1, s. 133-146.

Rószkiewicz M., *Metody ilościowe w badaniach marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 2002.

Sterne J.A.C., Egger M., *Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis*, "Journal of Clinical Epidemiology" 2001, 54, s. 1046-1055.

Thompson S.G., Sharp S.J., *Explaining heterogeneity in meta-analysis: A comparison of methods*, "Statistics in Medicine" 1999, 18, s. 2693-2708.

## CAPABILITIES AND CONSTRAINTS OF META-ANALYSIS OF QUANTITATIVE MARKETING RESEARCH AREA

**Summary:** The development of social science research leads to broadening the areas of conducted empirical research. Big supply of research results often relating to the same issues poses a question how to organize total analysis which decides what is known from the researched areas [Grave, Griffith 1971]. The fact that individual research is carried out on random samples causes that obtained results are not equal. Their changeability extent is described by estimator random variance. The paper deals with problems connected with the analysis of research results fluctuations which deciding is set by the paradigm of statistical meta-analysis.