

Małgorzata Rószkiewicz

Szkoła Główna Handlowa w Warszawie

ILOŚCIOWE SYNTEZY BADAŃ POWTARZALNYCH – PODSTAWY METAANALIZY

Streszczenie: Rozwój problematyki badań społecznych prowadzi do poszerzania się obszarów realizowanych badań empirycznych. Duża podaż wyników badań często odnoszących się do tych samych zagadnień rodzi pytanie, jak prowadzić analizę sumaryczną, ustalającą, co w zasadzie wynika z już dokonanych analiz [Grave, Griffith 1971]. To, iż poszczególne badania są realizowane 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.

W literaturze tematu metaanaliza jako odrębna procedura analityczna pojawia się od lat 70. [Glass 1976]. Wyniki analiz opublikowane od tego czasu odnosiły się do bardzo szerokiego spektrum zagadnień, co wskazuje, iż potrzebę prowadzenia tego typu analiz zgłaszają badacze niezależnie od domeny badań. Najczęściej przytaczane przyczyny uzasadniające tego typu podejście analityczne to z jednej strony wzrost podaży wyników badań, co skłania do dokonywania syntez i uogólnień, z drugiej zaś – oczekiwania, iż procedura ta spowoduje wzrost mocy statystycznej [Becker, Cohen 2003]. Jednakże oprócz opinii bardzo pozytywnych wyrażano również poglądy sceptyczne, głoszące, że metaanaliza jest swoistą alchemią, która nieistotne wyniki badań przekształca w wynik istotny [Feinstein 1995].

Podejmowane próby uogólniania wyników badań przyczyniają się do rozwoju metod odnoszących się do sposobów dokonywania tego typu syntezy [Cooper, Hedges 1994]. Ze względu na fakt, że w literaturze tematu brak jest jednej, powszechnie przyjętej definicji metaanalizy, na potrzeby niniejszego opracowania przyjęto, że metaanaliza w ujęciu statystycznym polega na wypracowaniu ogólnej konkluzji z wyników pochodzących z wielu badań. By było to możliwe, miernik statystyczny szacowany w ramach poszczególnych badań i będący podstawą uogólnień powinien przedstawiać rezultat badań w sposób łatwy do zinterpretowania i powinien być porównywalny między tymi badaniami. Ważne jest też, by znany był rozkład jego estymatora, znajomość zaś jego wariancji umożliwia wnioskowanie statystyczne. Najogólniej w metaanalizie zakłada się, że w poszczególnych badaniach ($i = 1, 2, \dots, k$) obserwowana jest statystyka T_i , co do której przyj-

muje się, że ma rozkład normalny o wartości oczekiwanej θ_i i znanej wariancji ν_i , dla $i = 1, 2, \dots, k$.

Fakt, iż oceny estymatora pochodzą z poszczególnych badań realizowanych na próbach losowych, sprawia, iż uzyskiwane wyniki nie są jednakowe. Ich zakres zmienności opisuje wariancja estymatora. Analiza tego zakresu dotyka kilku kwestii, których rozstrzygnięcie wyznacza paradygmat metaanalizy metodami statystycznymi. 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, wyznaczająca wynik końcowy, redukowała całkowitą nieprzewidywalność losową prób. Po drugie, wynik końcowy z kilku badań powinien być w większym stopniu zależny od wyników bardziej precyzyjnych, a w mniejszym stopniu od wyników uzyskiwanych z gorszą precyzją. Po trzecie, oceniając przedziały ufności uzyskane z poszczególnych badań, często można odnotować ich pokrywanie się, lecz w niektórych przypadkach uzyskane wyniki mogą wykraczać poza ten zakres. Nasuwa się zatem pytanie, czy wyniki otrzymane z poszczególnych badań mogą różnić się bardziej, niż wynikałoby to z wartości wariancji błędu. Innymi słowy, czy zasadne jest założenie, że wszystkie te badania szacują tę sama 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 odmiennność, czy też można je pominąć, a jeśli tak się stanie, to czy regularności pozostałych badań będą bardziej zwarte i logiczne. Po piąte wreszcie, obserwowane odmienności wyników poszczególnych badań mogą być rezultatem zmian systematycznych, zależnych od cech tych badań, lub mogą być jedynie artefaktem wywołanym wariancją błędu. Rodzi to pytanie, w jaki sposób dokonać tego rozróżnienia.

W metaanalizie wypracowano dwa podejścia badawcze wykorzystujące dwa odmienne modele statystyczne, tj. podejście zakładające występowanie jedynie efektów ustalonych oraz podejście zakładające występowanie obok efektów ustalonych również efektów losowych. W analizie z efektami ustalonymi zakłada się, że wszystkie wykorzystywane próby pochodzą z tej samej populacji o parametrze θ . Parametr ten jest szacowany na podstawie estymatora T , który w kolejnych badaniach przyjmuje wartość T_i , czyli:

$$T_i = \theta + \varepsilon_i, \quad (1)$$

gdzie: ε_i ma rozkład normalny z wartością oczekiwaną 0 i wariancją ν_i .

Najlepszym estymatorem parametru θ , wyznaczonym MNW i MNK, jest średnia ważona wyników estymacji z poszczególnych prób, przy wagach odwrotnie proporcjonalnych do wariancji losowej poszczególnych wyników. Ważony średni wynik wyznacza formuła (za: [Hedges, Olkin 1985]):

$$\bar{T} = \left(\frac{1}{\sum w_i} \right) \sum w_i T_i, \quad \text{gdzie: } w_i = \frac{1}{v_i}, \quad (2)$$

z wariancją błędu szacowaną jako:

$$V = \frac{1}{\sum w_i}. \quad (3)$$

Tak zdefiniowany estymator parametru Θ spełnia dwa pierwsze postulaty, składające się na ramy paradygmatu metaanalizy. Przyjęty system wag gwarantuje bowiem, że wyniki o gorszej precyzji będą miały mniejsze znaczenie dla wyniku końcowego, natomiast wyniki o lepszej precyzji – znaczenie większe. Z kolei wariancja błędu wyniku końcowego w miarę wzrostu liczby badań będzie maleć i w konsekwencji maleć będzie jego błąd standardowy, co zwiększy precyzję estymacji. Uzyskanie tego typu rezultatów uznaje się za główne osiągnięcie metaanalizy.

W modelu z efektami ustalonymi obowiązuje założenie, że próby będące podstawą metaanalizy są homogeniczne i pozwalają oszacować jeden, nieznaną parametr rozkładu populacji. Odwrotnie jest w przypadku modelu z efektami losowymi, gdzie zakłada się, że wyniki osiągane z prób losowych są realizacjami zmiennej losowej o określonym rozkładzie. W modelach z efektami losowymi rozważa się zatem dwa źródła zmienności obserwowanych wartości z prób, tj. zmienność błędu losowego próby oraz zmienność wyników estymacji otrzymywanych z poszczególnych prób.

W modelu o efektach losowych przyjmuje się, że Θ_i będzie nieznanym parametrem szacowanym w i -tym badaniu za pomocą estymatora T_i o wariancji losowej v_i . Zatem dane ze zbioru k badań generują oszacowania T_1, T_2, \dots, T_k i ich wariancje v_1, v_2, \dots, v_k . Naturalnym sposobem operacjonalizacji tak sformułowanego problemu badawczego jest dwupoziomowy model hierarchiczny, gdzie 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:

$$T_i = \theta_i + \varepsilon_i, \quad (4)$$

gdzie: ε_i ma rozkład normalny z wartością oczekiwaną 0 i wariancją v_i .

Parametr Θ jest średnim poziomem wartości parametrów szacowanych na podstawie wszystkich badań. W przypadku określonego zbioru k badań chodzi o zbiór parametrów $\Theta_1, \Theta_2, \dots, \Theta_k$, które są szacowane w poszczególnych badaniach. Na poziomie wyników badań wartości parametrów są kształtowane przez poziom średni możliwych wyników Θ oraz specyficzny dla badania efekt losowy η_i , czyli:

$$\theta_i = \theta + \eta_i, \quad (5)$$

gdzie: η_i ma rozkład normalny z wartością oczekiwaną 0 i wariancją τ^2 .

Parametr τ^2 jest często określany wariancją międzygrupową, jeśli przyjąć, że grupy wyznaczają kolejne badania. W tego typu podejściu konieczne jest rozróżnienie między wariancją błędu, która odnosi się do założenia o ustalonej wartości parametru θ_i (zwaną wariancją warunkową błędu i oznaczaną jako ν_i), i wariancją związaną ze zmiennością θ_i (zwaną wariancją bezwarunkową błędu i oznaczaną jako $\nu_i^* = \nu_i + \tau^2$, gdyż co do błędu próby ε_i i efektu losowego η_i przyjmuje się, że są niezależne).

Formuła modelu o efektach losowych jest identyczna z hierarchicznym modelem liniowym, często stosowanym w analizach danych społeczno-ekonomicznych. Jednakże ma dwie cechy, które ją 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 tego poziomu. W modelu metaanalizy wariancje wyników poszczególnych badań są różne, co uwidacznia się w różnych wartościach ν_i . Po drugie, w tradycyjnym modelu wariancja ta jest nieznaną i musi być szacowana na podstawie danych z próby. W modelu metaanalizy przyjmuje się, że wariancje losowe wyników poszczególnych badań są znane. Model dwupoziomowy metaanalizy można zatem zapisać jako równanie:

$$T_i = \theta + \eta_i + \varepsilon_i = \theta + \xi_i, \quad (6)$$

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

Estymator MNW i MNK parametru θ definiuje formuła (za: [Hedges, Vevea 1998]):

$$\bar{T} = \frac{\sum_{i=1}^k w_i^* T_i}{\sum_{i=1}^k w_i^*}, \quad \text{gdzie: } w_i^* = \frac{1}{(\nu_i + \hat{\tau}^2)} = \frac{1}{\nu_i^*}. \quad (7)$$

Wariancję tego estymatora wyznacza formuła:

$$V^* = \frac{1}{\sum w_i^*}. \quad (8)$$

Warto zauważyć, że w modelu o efektach losowych estymator jest również średnią ważoną wartości T_i , gdzie wagami są wartości odwrotne do wariancji błędu, co oznacza, że większą wagę nadaje się tym wynikom, które są bardziej precyzyjne. Jednakże w modelu o efektach losowych, jeśli wariancja międzygrupowa

oszacowana na podstawie $(k + 1)$ badań wzrośnie w stosunku do wariancji międzygrupowej oszacowanej na podstawie k badań, to błąd standardowy wyniku końcowego może być wyższy. Dlatego też w modelu z efektami losowymi wzrost liczby badań nie musi powodować spadku standardowego błędu i wzrostu mocy statystycznej wnioskowania. Ponadto powiększanie wariancji losowej próby (ν_i) o wariancję efektu losowego (oszacowana wartość τ^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 w miejsce 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, tabl. 1. s. 279]). Jednakże 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 wynik końcowy wykroczy poza zakres wyników badań wykorzystanych do wyznaczenia jego wartości. Pod tym względem model z efektami ustalonymi jest bardziej restrykcyjny i dopuszcza wnioskowanie jedynie w zakresie uzyskiwanych wyników.

W tym miejscu warto również odnotować inne 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 to założenie. Z tych powodów podejście oparte na efektach losowych dostarcza wnioskowania, które może być błędne.

Formalnie o wyborze modelu rozstrzyga wnioskowanie o wariancji międzygrupowej τ^2 . Statystyką testującą istotność wariancji międzygrupowej wykorzystującą ocenę średniego wyniku z badań, jaki wystąpiłby, gdyby wariancja ta była równa zero, jest statystyka Q (zwana również statystyką χ^2 Cochrańa o postaci [Konstantopoulos, Hedges 2004]):

$$Q = \sum_{i=1}^k \frac{(T_i - \bar{T})^2}{\nu_i}, \quad (9)$$

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

Statystyka ta przy prawdziwości hipotezy zerowej ma rozkład chi-kwadrat o $(k - 1)$ stopniach swobody. Warto jednak pamiętać, że im mniejszy zbiór badań (ma-

le k), tym mniejsza moc tego testu. Moc ta jest słaba również, gdy wariancje wewnątrzgrupowe v_i są znaczne. Może zatem zdarzyć się, że mimo iż nie będzie podstaw do odrzucenia hipotezy zerowej, to jednak może wystąpić widoczna (znaczna) wartość $\hat{\tau}^2 > 0$. Dlatego też niektórzy autorzy rekomendują, by w praktyce nie traktować wiążąco wyniku tego testu i mimo braku podstaw do odrzucenia hipotezy zerowej o zerowym efekcie losowym szacować tę wariancję i uzyskany wynik wykorzystywać w konstrukcji wag do rachunku wyniku końcowego metaanalizy.

Trudności z weryfikacją założeń dotyczących rozkładu efektów losowych sprawiają, że najczęściej w estymacji wariancji międzygrupowej stosowane jest podejście nieparametryczne wykorzystujące statystykę Q według formuły:

$$\hat{\tau}^2 = \begin{cases} \frac{Q - (k - 1)}{\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}} & \text{dla } Q \geq (k - 1) \\ 0 & \text{dla } Q < (k - 1) \end{cases}, \quad (10)$$

gdzie: $w_i = 1/v_i$.

Jeśli wartość statystyki testującej Q przekracza liczbę stopni swobody, uznaje się wariancję międzygrupową za różną od zera. Jeśli zaś liczba stopni swobody przekracza wartość statystyki Q , wówczas wagi w_i uznaje się za stałe i w konsekwencji stwierdza się, że oszacowanie parametru θ nie zależy od założenia co do rozkładu efektów losowych, a to może być jedynie efektem statystycznym, niezgodnym z faktycznym stanem rzeczy. Dlatego też sięga się również do pośrednich procedur weryfikujących istotność efektów losowych [Higgins, Thompson 2002].

Do procedur pośrednich rozstrzygnięcia o istotności zmienności między wynikami badań należy weryfikacja założeń o wpływie cech rozważanych badań na kształtowanie się ich wyników. Chodzi tu o rozstrzygnięcie, na ile cechy specyficzne poszczególnych badań wpływają na ich wyniki i czy odmienności obserwowane między poszczególnymi wynikami badań, rejestrowane przez wariancję międzygrupową $\hat{\tau}^2$, mają charakter systematyczny, czy też są to artefakty wynikające ze zmienności losowej poszczególnych badań. Najdogodniejszą procedurą statystyczną rozstrzygającą tę kwestię są modele mieszane. Jeśli bowiem założyć, że efekty ustalone są opisane za pomocą p predyktorów (cechy badań) X_1, X_2, \dots, X_p , które są powiązane liniowo z wynikiem badań, to model drugiego poziomu dla wyniku pochodzącego z i -tego badania przyjmuje postać:

$$T_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \eta_i, \quad (11)$$

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

Model dwupoziomowy przy założeniu efektów losowych przyjmuje postać:

$$T_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \eta_i + \varepsilon_i. \quad (12)$$

Jeśli τ^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, a tym samym o istotności zmienności między wynikami badań. Na ogół jednak τ^2 nie jest znane i wówczas możliwe są cztery drogi postępowania:

1. Oszacowanie τ^2 na podstawie dostępnych wyników badań i wykorzystanie oceny tej wariancji do estymacji współczynników regresji ważoną metodą najmniejszych kwadratów.

2. Łączne oszacowanie wektora współczynników i wariancji τ^2 za pomocą nierestrykcyjnej metody największej wiarygodności (REML).

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

4. Oszacowanie wektora współczynników regresji dla każdej wartości τ^2 z przedziału prawdopodobnych wartości i użycie ważonej średniej dla tych wyników przez przyjęcie za wagi wartości odpowiednich do prawdopodobieństw, tak jak to ma miejsce podejściu bayesowskim.

Słaba moc testu o istotności efektów losowych skłania wielu autorów do preferowania modelu z efektami losowymi w miejsce modelu z efektami ustalonymi. Jednakże wybór modelu statystycznego w metaanalizie jest nie tylko kwestią statystycznych własności wyników badań, ale i przyjętych założeń o istocie badanego zjawiska. Dokonując rozstrzygnięć o modelu statystycznym wykorzystywanym w metaanalizie, warto mieć na uwadze, 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 i ocena błędu tego szacunku 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.

Literatura

- Becker B.J., Cohen L.D., *How meta-analysis increases statistical power*, „Psychological Methods” 2003 no 3, s. 243-253.
- Cooper H.M., Hedges L.V. (red.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York 1994.
- Feinstein A.R., *Meta-analysis: Statistical alchemy for the 21st century*, „Journal of Clinical Epidemiology” 1995 no 48, s. 71-79.
- 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.
- Higgins J.P.T., Thompson S.G., *Quantifying heterogeneity in a meta-analysis*, „Statistics in Medicine” 2002 no 21, s. 1539-1558.
- 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.
- Konstantopoulos S., Hedges L.V., *Meta-Analysis*, [w:] *Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan (red.), Sage Publications, Inc., New York 2004.

QUANTITATIVE SYNTHESIS OF SURVEYS RESULTS – THE BASE OF META-ANALYSIS

Summary: The growth of the social science research enterprise has led to a large body of related research studies. The sheer volume of research related to the same topics poses a question how to organize and summarize these findings to identify and exploit what is known and focus research on promising areas [Grave, Griffith 1971]. The effect size estimates from the studies are not identical. This is to be expected because the estimates are based on data from samples and random variations due to sampling should introduce fluctuations into estimates. In the paper, the methods for exploring these fluctuations will be introduced as paradigm for explorations that are sensible in meta-analyses generally.