



## ILOŚCIOWA SYNTEZA WYNIKÓW BADAŃ PIERWOTNYCH - METAANALIZA W *STATISTICA ZESTAWIE MEDYCZNYM*

*Michał Kusy, StatSoft Polska Sp. z o.o.*

### Wprowadzenie

Metaanaliza, czyli ilościowa synteza wyników pochodzących z niezależnych badań, to narzędzie badawcze, które w ostatnim czasie zyskało w medycynie, farmakologii i epidemiologii szczególne znaczenie. Metaanalizę badań z grupą kontrolną stawia się najwyżej w hierarchii wiarygodności danych z badań klinicznych. Pozwala ona bardziej precyzyjnie ocenić wpływ działania analizowanego czynnika oraz rozszerzyć wnioski z pojedynczych badań na szerszą populację. W kolejnych rozdziałach przedstawimy przykład przeprowadzenia tego typu analizy za pomocą modułu *Metaanaliza* dostępnego w *STATISTICA Zestawie Medycznym*. Prezentowane narzędzia umożliwią wykonanie obliczeń dla szeregu miar efektu i przedstawienie wyników w postaci szczegółowego raportu oraz wykresu typu *forest plot*. Przyjrzymy się również miarom efektu, niejednorodności badań i wrażliwości efektu łącznego na wyniki poszczególnych badań.

Metaanalizę wykonamy na przykładzie danych pochodzących z analizy skuteczności witaminy D w zapobieganiu upadkom osób starszych (Bischoff-Ferrari [1]). Blisko połowa upadków u osób starszych prowadzi do urazów, które często kończą się hospitalizacją, upośledzeniem sprawności, a nawet powikłaniami prowadzącymi do zgonu. Witamina D wraz z wapniem pełni ważną rolę w utrzymaniu kości w dobrym stanie. Jak się okazuje, może ona również wpływać bezpośrednio na siłę mięśni i zmniejszać przez to ryzyko upadku. Głównym celem analizy będzie zatem ocena łącznego efektu witaminy D w prewencji upadków u osób starszych, na podstawie przeglądu systematycznego badań pierwotnych.

### Miary efektu

W zależności od rodzaju zebranych wyników możemy wyszczególnić kilka rodzajów miar efektu, m.in. miary oparte na średnich, miary ekspozycji na zdarzenie i korelacje. W przypadku, gdy wynik przedstawiony jest na skali ilościowej, wspólnej dla wszystkich badań, metaanaliza może być przeprowadzona bezpośrednio na zwykłej różnicy średnich (*raw mean difference*) w porównywanych grupach:



$$D = \bar{X}_1 - \bar{X}_2.$$

Jeśli skala, na której przedstawiono wynik, nie jest powszechnie znana lub gdy badania używają różnych skal, możemy podzielić różnicę średnich przez wewnętrzgrupowe odchylenie standardowe, tworząc w ten sposób porównywalny wskaźnik. Nazywamy go standaryzowaną różnicą średnich lub *d Cohena*:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{wewn}}.$$

Ponieważ *d* jest współczynnikiem obciążonym, ze skłonnością do przeszacowania estymowanego parametru w przypadku mniej licznych prób, często przemnaża się go dodatkowo przez czynnik korygujący *J*, otrzymując nieobciążony współczynnik nazywany *g Hedgesa*:

$$g = J \times d.$$

W badaniach, w których efekt mierzono na skali ilorazowej, możemy dodatkowo sprawdzić, ile razy silniejszy lub słabszy jest efekt w grupie z interwencją w porównaniu do grupy kontrolnej, korzystając z ilorazu średnich *R (response ratio)*:

$$R = \frac{\bar{X}_1}{\bar{X}_2}.$$

W metaanalizie obliczenia na skonstruowanych w powyższy sposób miarach (tzn. opartych na ilorazie efektów) będziemy przeprowadzać na skali logarytmicznej. Za chwilę przyjrzemy się tej sytuacji bliżej, na przykładzie ilorazu szans.

W przypadku badań, których wynikiem jest liczba pacjentów w porównywanych grupach, u których wystąpiło (lub nie wystąpiło) pewne zdarzenie, najczęściej wyznacza się miary takie jak: współczynnik ryzyka *RR (risk ratio/relative risk)*, iloraz szans *OR (odds ratio)* i różnica ryzyk *RD (risk difference)*:

	<b>Interwencja</b>	<b>Kontrola</b>
<b>Wystąpiło zdarzenie</b>	A	C
<b>Brak zdarzenia</b>	B	D

$$RR = \frac{A/(A+B)}{C/(C+D)}, \quad OR = \frac{A/B}{C/D}, \quad RD = \frac{A}{A+B} - \frac{C}{C+D}.$$

Założmy teraz, że jedno z badań (I) uwzględnionych w metaanalizie wykazało dwukrotnie wyższą szansę wystąpienia zdarzenia w grupie z interwencją, natomiast na podstawie drugiego badania (II) dwukrotnie wyższa jest szansa wystąpienia zdarzenia w grupie kontrolnej. Zakładając jednakowe wagi dla każdego z badań, powinniśmy otrzymać efekt łączny wskazujący na jednakową szansę wystąpienia zdarzenia w grupie z interwencją i grupie kontrolnej (iloraz szans równy 1). Gdybyśmy wyliczyli łączną miarę efektu



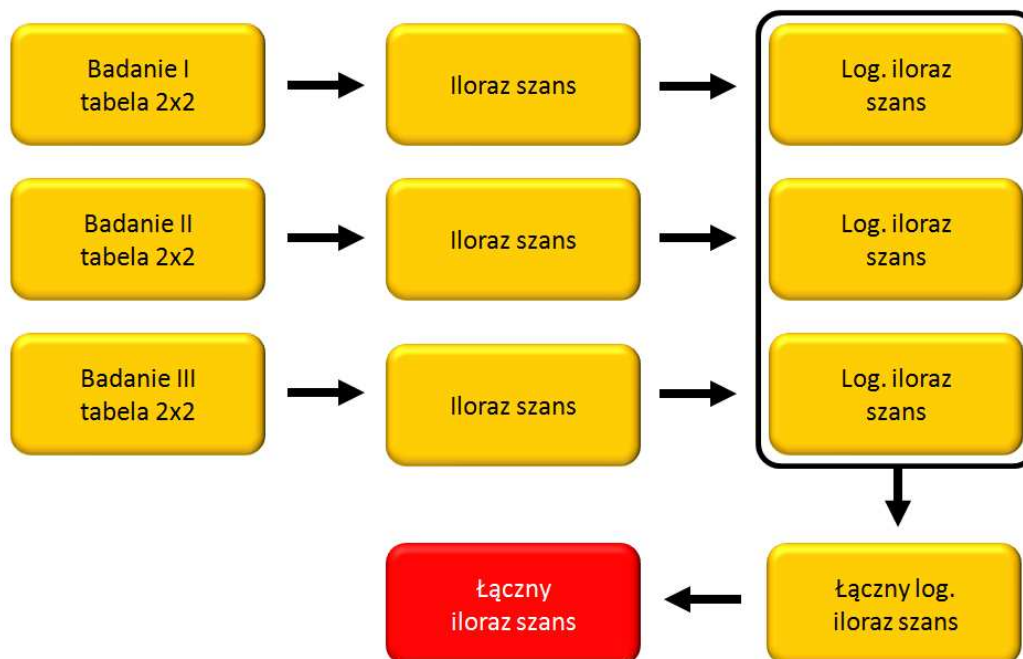
bezpośrednio na podstawie ilorazów szans, otrzymalibyśmy wartość średnią równą 1,25. Dzięki zlogarytmowaniu ilorazów szans dostajemy natomiast dwie przeciwne wartości: 0,693 i -0,693, których średnia wynosi 0. Przedstawienie efektu łącznego na wyjściowej skali daje nam iloraz szans równy 1.

	Szansa (interwencja)	Szansa (kontrola)	Iloraz szans	Log. iloraz szans
<b>Badanie I</b>	4:1	2:1	2	0,693
<b>Badanie II</b>	2:1	4:1	0,5	-0,693

**Łączny iloraz szans:**

<b>1,25</b>	<b>1</b>
-------------	----------

Z tego względu przed przeprowadzeniem metaanalizy dla ilorazów średnich, ryzyk i szans stosujemy przekształcenie logarytmiczne. Dopiero po wyliczeniu efektu łącznego, na podstawie zlogarytmowanych współczynników, przedstawiamy wynik na wyjściowej skali.



Na przekształconych miarach wykonujemy również metaanalizę dla efektów wyrażonych w postaci współczynników korelacji  $r$  między dwiema zmiennymi ciągłymi, choć w tym przypadku transformacja (do  $Z$  Fishera o rozkładzie normalnym) jest nieco bardziej skomplikowana.

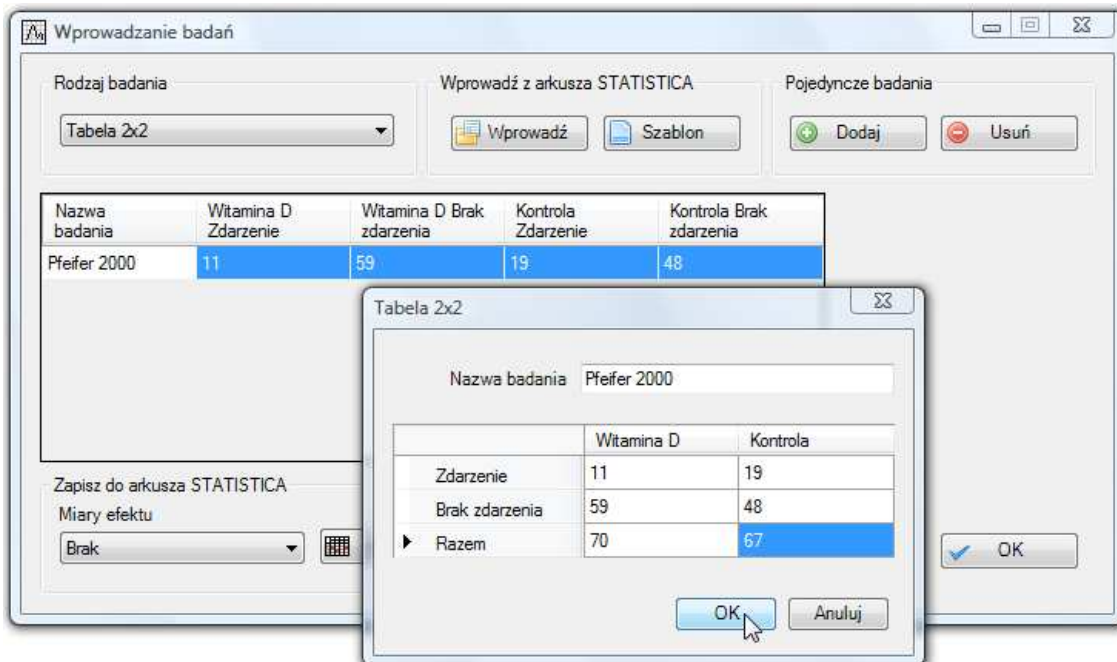
Oprócz miar efektu w metaanalizie będziemy potrzebowali również błędów standardowych. Na podstawie tych wartości wyliczane będą m.in. przedziały ufności i wagi przypisywane poszczególnym badaniom. W przypadku miary efektu opartej na średnich błęd



standardowy wyznaczany będzie na podstawie odchyłeń standardowych i licznosci w porównywanych grupach. Błędy standardowe dla miar ekspozycji na zdarzenie szacuje się natomiast na podstawie licznosci zebranych w tabelach 2x2.

Czasami, mimo że w badaniach korzystano z tej samej skali pomiarowej, wyniki przedstawiane są za pomocą innych miar efektu, np. dla części badań za pomocą różnicy w średnim poziomie ciśnienia skurczowego krwi, a dla innych jako iloraz szans uzyskania ciśnienia niższego od wartości progowej. Okazuje się, że istnieje możliwość konwersji między niektórymi miarami efektu, dzięki czemu można uwzględnić w metaanalizie większą liczbę badań. Oczywiście w takiej sytuacji stajemy zawsze przed wyborem między wykorzystaniem przybliżonych wartości miar efektu i przyjęciem pewnych założeń o włączanych wynikach (które nie zawsze muszą odpowiadać rzeczywistości) a pominięciem części badań, przez co tracimy część informacji i pracujemy na obciążonej próbie. Jeżeli uwzględnimy w metaanalizie dodatkowe badania przedstawiające inne miary efektu, warto przeprowadzić później analizę wrażliwości, w celu sprawdzenia, w jakim stopniu wpłynęły one na efekt łączny.

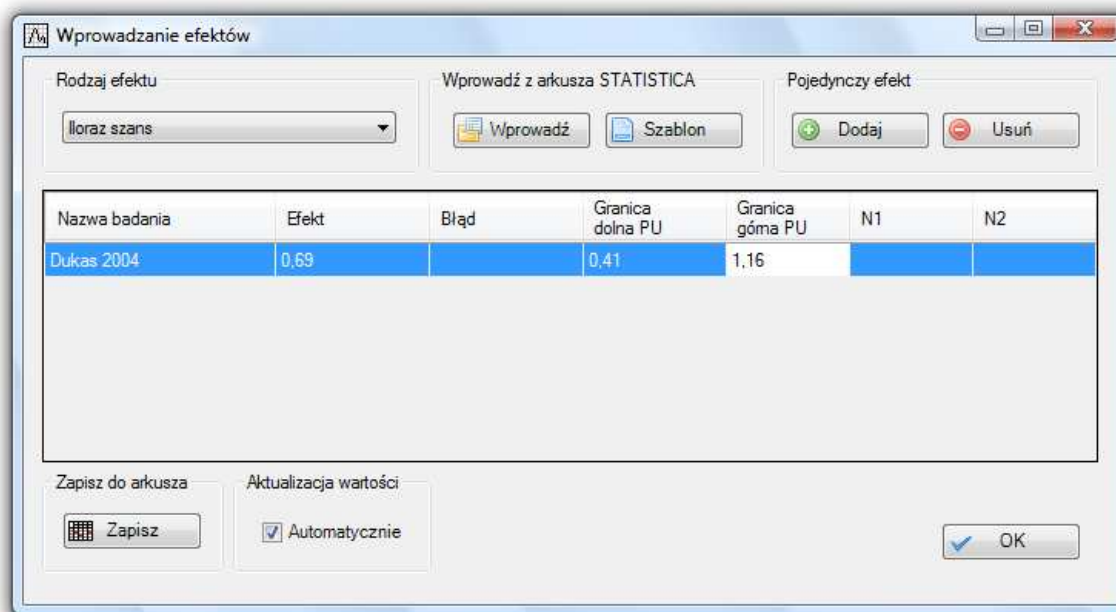
Obliczenia w module *Metaanaliza* rozpoczniemy od wprowadzenia wyników poszczególnych badań. Do analizy włączono jedynie podwójnie zaślepięone badania z randomizacją, opisujące związek między stosowaniem witaminy D a występowaniem upadków w populacji osób starszych (średni wiek nie mniejszy niż 60 lat). Miarą głównego punktu końcowego był iloraz szans wystąpienia przynajmniej jednego upadku w trakcie roku obserwacji wśród osób otrzymujących witaminę D w porównaniu do badanych, którzy witaminy D nie otrzymywali. Ostatecznie spośród 38 zidentyfikowanych badań do właściwej analizy włączono 5 badań spełniających kryteria, a kolejne 5 uwzględniono również w analizie wrażliwości. Bardziej szczegółowe informacje dotyczące rozważanych badań i kryteriów ich włączania do metaanalizy znaleźć można w artykule [1].





W programie *STATISTICA* wywołujemy moduł *Metaanaliza* dostępny w *STATISTICA Zestawie Medycznym*, a następnie przechodzimy do okna, w którym wprowadzimy wyniki badań pierwotnych. Ponieważ interesującą nas miarą jest iloraz szans, wybieramy opcję *Tabela 2x2*, a następnie dodajemy pierwsze badanie (*Pfeifer 2000*). Wystarczy teraz podać liczbę przypadków, w których wystąpiło zdarzenie (przynajmniej jeden upadek w ciągu roku obserwacji), i przypadków bez zdarzenia, w dwóch grupach badanych: otrzymujących witaminę D i nieotrzymujących witaminy D. W badaniu *Pfeifer 2000* upadku doświadczyło 11 spośród 70 osób otrzymujących witaminę D oraz 19 spośród 67 osób, którym nie podawano witaminy D. Iloraz szans i 95% przedział ufności wynoszą w tym badaniu odpowiednio: 0,47 i (0,20; 1,09).

Jeśli dane z badań pierwotnych zostały wcześniej zapisane w arkuszu *STATISTICA*, można je wprowadzić do analizy jednocześnie. Moduł *Metaanaliza* umożliwia również wprowadzenie badań z już wyliczonymi efektami, co okazuje się szczególnie istotne w przypadku, gdy nie mamy dostępu do szczegółowych danych dla badania, dysponujemy jednak wyliczoną na ich podstawie miarą efektu. Z taką sytuacją mamy do czynienia w badaniu *Larsen 2002*, gdzie podano jedynie iloraz szans i przedziały ufności, nie znamy natomiast licznosci z tabeli 2x2. Z opcji wprowadzania gotowych efektów będziemy również korzystać w badaniach, w których wyliczając miarę efektu, zastosowano jakąś korektę, np. uwzględniono dodatkowe czynniki, takie jak: wiek badanych, płeć, BMI itp. (jak w badaniu *Dukas 2004*).

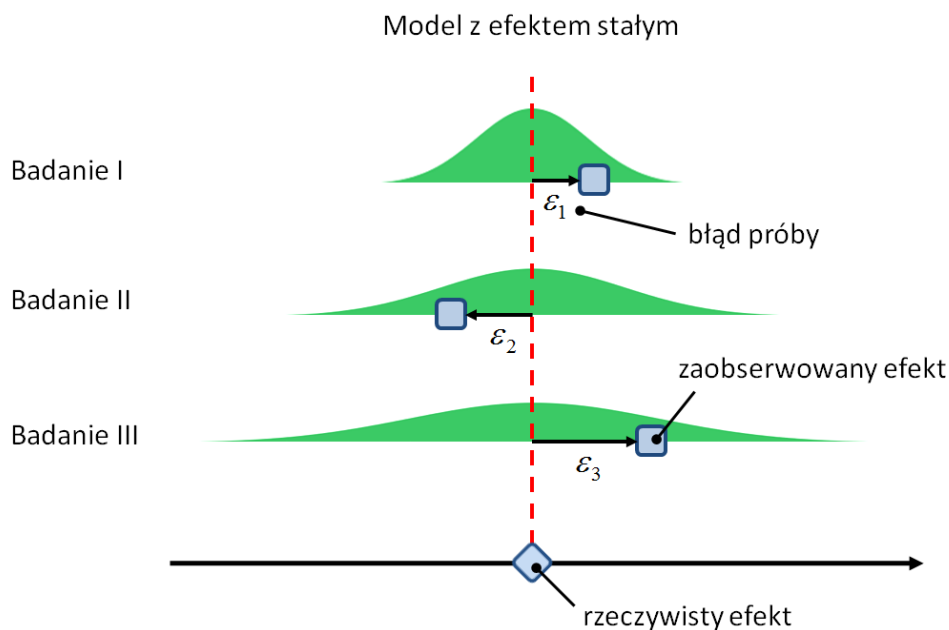


Po wprowadzeniu danych z badań pierwotnych program wylicza automatycznie wszystkie dostępne dla takich danych miary efektu spośród: różnicy średnich, ilorazu średnich, *g* Hedgesa, *d* Cohena, ilorazu szans, ryzyka względnego, różnicy ryzyk i współczynnika korelacji. Dzięki możliwości przeliczania między sobą ilorazów szans *OR*, współczynników *d* Cohena i *g* Hedgesa oraz współczynnika korelacji *r*, moduł *Metaanaliza* daje również możliwość jednoczesnego uwzględnienia w analizie badań z wynikami

przedstawionymi w postaci różnych miar efektu. Badania adekwatne do wskazanej miary efektu będą przez program wybierane automatycznie i włączane do metaanalizy.

## Metaanaliza - model z efektem stałym i zmiennym

W większości przypadków metaanaliza jest oparta na jednym z dwóch modeli statystycznych: modelu z efektem stałym (*fixed-effect model* lub *common-effect model*) lub modelu z efektem zmiennym (*random-effects model*). W modelu z efektem stałym (czy też wspólnym) zakłada się, że wyniki wszystkich badań opisują tę samą, rzeczywistą wartość efektu, a różnice w obserwowanych efektach wynikają z błędu próby. Inaczej mówiąc, wszystkie czynniki, które mogą wpłynąć na wartość efektu są jednakowe w analizowanych badaniach, w związku z czym efekt jest dla wszystkich badań wspólny. W tym przypadku efekt łączny jest wyznaczany na podstawie średniej ważonej efektów, przy czym wagi są wyznaczane jako odwrotności wariancji odpowiednich badań.

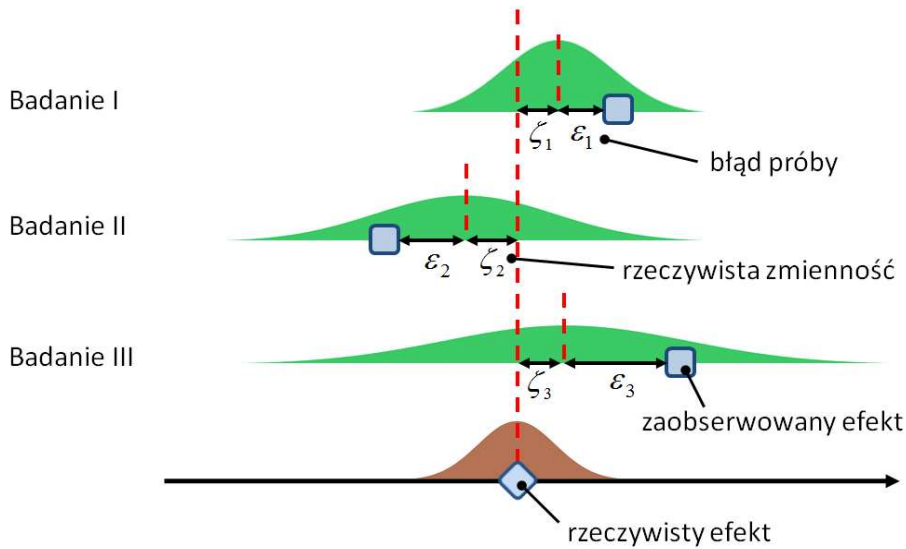


Decydując się włączyć do metaanalizy grupę badań, zakładamy, że mają one na tyle zgodny charakter, że ma sens wykonanie syntezy ich wyników. Najczęściej nie możemy jednak zakładać, że są to jednakowe badania w tym sensie, że rzeczywisty efekt jest dokładnie taki sam w każdym z nich. Efekt może być np. większy wśród pacjentów starszych, mniej schorowanych, może zależeć od bardziej intensywnej rodzaju interwencji. W modelu z efektem zmiennym (nazywanym również modelem efektów losowych) uwzględniamy sytuację, w której nieznane, rzeczywiste efekty różnią się między badaniami. Zwykle zakłada się, że pochodzą one z rozkładu normalnego. Na obserwowaną wartość efektu będzie się zatem składał ogólny średni efekt, odchylenie rzeczywistego efektu w badaniu od średniego efektu i odchylenie wartości obserwowanej od rzeczywistego efektu w badaniu, związane z błędem próby. Również w tym modelu efekt łączny jest wyznaczany na podstawie



średniej ważonej efektów poszczególnych badań z tym, że wagi są wyznaczone jako odwrotności wariancji wewnątrz badań, powiększonej o wariancję pomiędzy badaniami.

Model z efektem zmiennym



Po wprowadzeniu wyników wszystkich pięciu badań, które chcemy uwzględnić, kontynuujemy analizę, przechodząc do okna *Metaanaliza*.

The screenshot shows the 'Metaanaliza' window. The title bar reads 'Metaanaliza'. The main area contains a table with the following data:

Wł.	Badanie	OR	Błąd std.	Granica dolna PU	Granica górna PU	p-value	Udział %
<input checked="" type="checkbox"/>	Pfeifer 2000	0.47	1.53	0.2	1.09	0.077	9.1
<input checked="" type="checkbox"/>	Bischoff 2003	0.68	1.51	0.3	1.53	0.353	9.6
<input checked="" type="checkbox"/>	Gallagher 2001	0.53	1.3	0.32	0.89	0.015	24.4
<input checked="" type="checkbox"/>	Dukas 2004	0.69	1.3	0.41	1.16	0.163	23.3
<input checked="" type="checkbox"/>	Graafmans 1996	0.91	1.25	0.59	1.4	0.658	33.6
	Podsumowanie	0.68	0.09	0.53	0.88	0.003	100



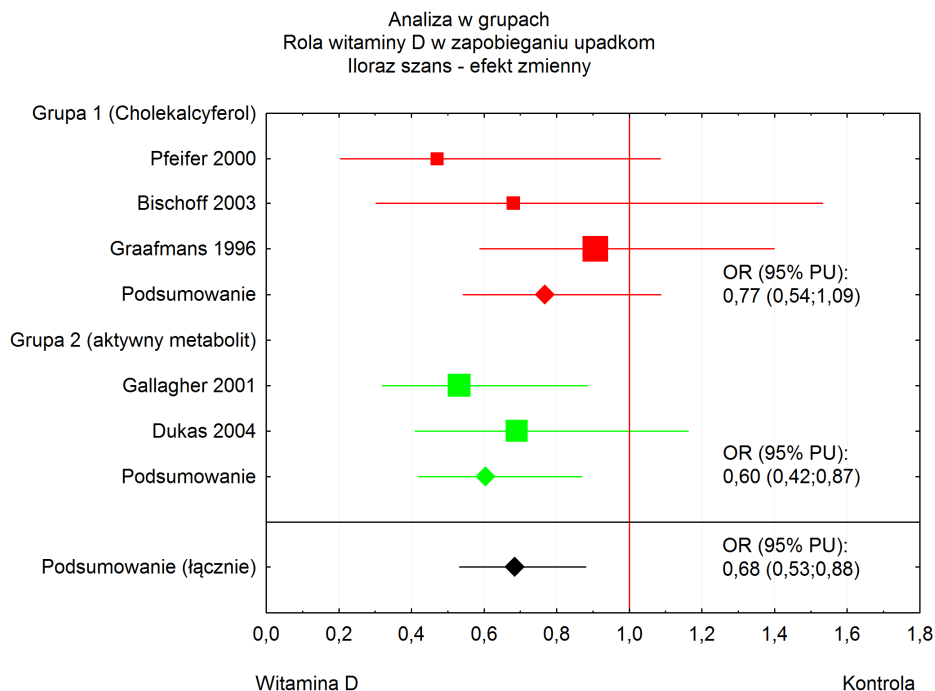
statystycznie efekt łączny. Dodatkowo zmniejszyła się niepewność oceny efektu. Wykres pozwala również zauważyć duży wpływ trzech ostatnich badań na wynik końcowy.

## Niejednorodność wyników

Rozważając niejednorodność wartości efektu w analizowanych badaniach, mamy na uwadze zmienność rzeczywistej wartości efektu. Obserwowany rozrzut wyników uwzględnia jednak zarówno zmienność rzeczywistą, jak i błąd losowy. W związku z tym, aby ocenić niejednorodność zebranych wyników, porównujemy zaobserwowaną zmienność pomiędzy badaniami ze zmiennością, której oczekivalibyśmy, gdyby badania wskazywały efekt wspólny. Różnica między nimi powinna odzwierciedlać rzeczywiste różnice między badaniami i na jej bazie konstruuje się kilka miar niejednorodności.

Przykładową miarą oceniającą niejednorodność wyników jest ważona suma kwadratów  $Q$ . Jest to miara standaryzowana, dzięki czemu omijamy problemy związane ze skalowaniem wyników. Przy założeniu, że każde z  $k$  badań wskazuje na ten sam stały efekt, współczynnik  $Q$  będzie pochodził z rozkładu chi-kwadrat o  $k-1$  stopniach swobody. W ten sposób otrzymujemy test sprawdzający niejednorodność efektów. Na podstawie wartości  $Q$  wyznacza się również  $T^2$  - estymator wariancji rzeczywistych efektów, oraz  $I^2$  - statystykę, która przedstawia udział procentowy zmienności rzeczywistej do łącznej.

Do oceny niejednorodności badań autorzy artykułu wykorzystali między innymi statystyki  $Q$  i  $I^2$ . Wyliczone wartości nie wskazują na niejednorodność badań -  $Q$  jest równe około 3,327, a odpowiadający mu poziom istotności  $p$  wynosi około 0,505. Z kolei szacowana wartość wariancji rzeczywistych efektów  $T^2$  (a co za tym idzie również  $I^2$ ) jest równa 0. Na podstawie tych wyników nie możemy zatem odrzucić hipotezy o jednorodności badań.





Autorzy artykułu sprawdzili również efekty w podgrupach klinicznych wyznaczonych przez formę podawanej witaminy D (cholekalcyferol lub aktywny metabolit), stosowanie suplementacji wapnia, długość leczenia i płeć. Dla przykładu przeprowadzimy teraz analizę w grupach (*subgroup analysis*) wyznaczonych przez formę witaminy D. Cholekalcyferol otrzymywało 613 badanych z trzech badań (*Pfeifer 2000*, *Bischoff 2003* i *Graafmans 1996*), natomiast aktywny metabolit 259 osób z pozostałych dwóch badań (*Gallagher 2001* i *Dukas 2004*). W przypadku pierwszej grupy badanych łączny iloraz szans wyniósł 0,77, wynik nie jest jednak istotny statystycznie (na poziomie istotności  $\alpha=0,05$ ). Istotny jest natomiast wynik w grupie badań, w których podawano pacjentom aktywny metabolit (OR=0,60).

## Analiza wrażliwości

Analiza wrażliwości pozwala ocenić, na ile założenia metaanalizy i uwzględnione w niej badania wpłynęły na uzyskane łączne wyniki. Gdy część wyników znacznie odbiega od pozostałych, można zbadać wpływ poszczególnych badań na efekt łączny, powtarzając wielokrotnie metaanalizę po wyłączeniu każdego z nich z osobna, lub wyłączeniu grupy badań. Podobnie jeśli badanie nie zostało uwzględnione w metaanalizie, np. ze względu na jakość danych, możemy sprawdzić, jak zmieniłaby się łączna wartość efektu, gdyby włączyć je do analizy.

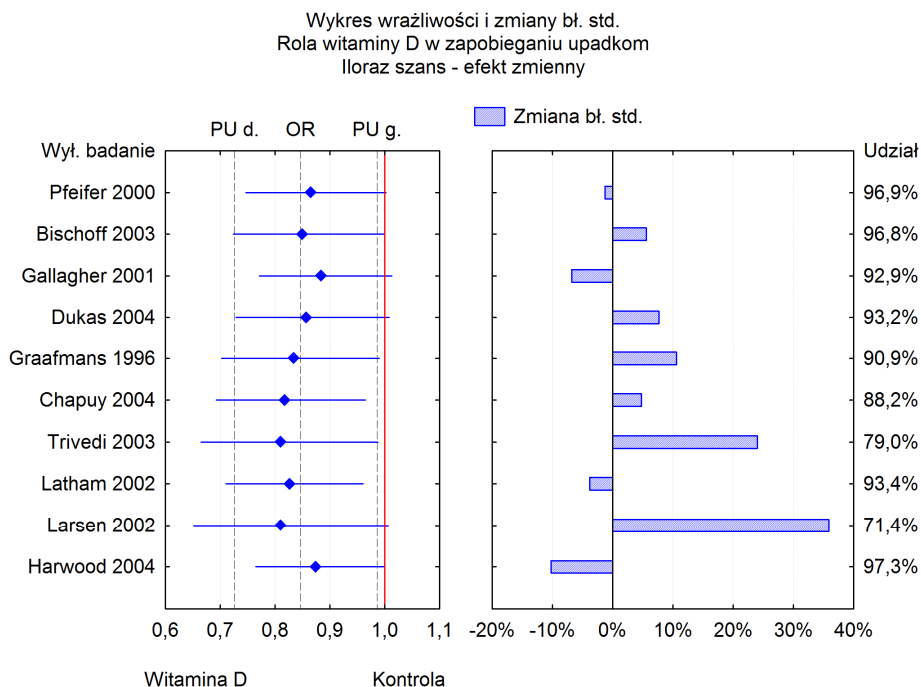
The screenshot shows the 'Metaanaliza' software window. The title bar reads 'Metaanaliza'. The main area is divided into several sections:

- Metaanaliza:** Contains buttons for 'Wczytaj', 'Zapisz', and 'Wprowadź'. The 'Wprowadź' section has buttons for 'Surowe', 'Efekt', 'Podsumowanie', and 'Zamknij'.
- Wprowadź:** Contains input fields for 'Rola witaminy D w zapobieganiu upadkom', 'Czynnik 1' (Witamina D), and 'Czynnik 2' (Kontrola).
- Badania:** A tabbed interface with 'Badania', 'Analiza', and 'Grupy' tabs. The 'Badania' tab is active, showing a table of study results.

Wl.	Badanie	OR	Błąd std.	Granica dolna PU	Granica górna PU	p-value	Udział %
<input checked="" type="checkbox"/>	Pfeifer 2000	0,47	1,53	0,2	1,09	0,077	7,9
<input checked="" type="checkbox"/>	Bischoff 2003	0,68	1,51	0,3	1,53	0,353	8,2
<input checked="" type="checkbox"/>	Gallagher 2001	0,53	1,3	0,32	0,89	0,015	14,2
<input checked="" type="checkbox"/>	Dukas 2004	0,69	1,3	0,41	1,16	0,163	13,9
<input checked="" type="checkbox"/>	Graafmans 1996	0,91	1,25	0,59	1,4	0,658	16,3
<input checked="" type="checkbox"/>	Chapuy 2004	1,08	1,2	0,75	1,54	0,679	18,7
<input type="checkbox"/>	Trivedi 2003						
<input checked="" type="checkbox"/>	Latham 2002	1,31	1,31	0,77	2,23	0,321	13,6
<input checked="" type="checkbox"/>	Larsen 2002						
<input checked="" type="checkbox"/>	Harwood 2004	0,37	1,57	0,15	0,89	0,027	7,2
	Podsumowanie	0,76	0,11	0,58	1,01	0,057	100

W analizie wrażliwości uwzględnimy pięć dodatkowych badań. Badań tych nie rozważano w głównej analizie, ze względu na niespełnienie części kryteriów - między innymi brak definicji „upadku”. Obliczony łączny iloraz szans jest teraz na poziomie 0,85, co oznacza, że zróżnicowanie szansy wystąpienia upadku w porównywanych grupach jest mniejsze niż w podstawowej analizie (OR=0,68). Oprócz dołączenia do analizy nowych badań moduł *Metaanaliza* umożliwi również wskazanie, dla których spośród nich ma zostać obliczony efekt łączny. Dzięki temu można przeprowadzić analizę wrażliwości dla dowolnych kombinacji badań.

W programie możemy dodatkowo utworzyć wykres wrażliwości, który pokazuje, jakie zmiany w efekcie łącznym spowodowałoby wyłączenie z metaanalizy poszczególnych badań. Na takim wykresie pionowymi liniami zaznaczony jest efekt łączny i granice przedziału ufności dla całościowej metaanalizy. Do tych poziomów możemy odnieść wartość efektu łącznego i zakres przedziału ufności dla metaanalizy przeprowadzonej bez danego badania. Widzimy na przykład, że nieuwzględnienie w analizie badania *Bischoff 2003* nie wpływa znacznie na efekt łączny, natomiast metaanaliza przeprowadzona bez badania *Larsen 2002* daje wyraźnie mniejszy iloraz szans przy jednoczesnym zwiększeniu zakresu przedziału ufności. Utworzenie dodatkowego wykresu przedstawiającego procentową zmianę błędu standardowego po wyłączeniu badania z metaanalizy pokazuje, że w tym przypadku błąd standardowy wzrósł o prawie 40%. Okazuje się zatem, że badanie *Larsen 2002* miało silny wpływ na wyniki metaanalizy dziesięciu badań. Było to najliczniejsze spośród analizowanych badań - liczba uczestników wynosiła 5771, a waga, która została mu przyznana w metaanalizie, to około 29%.





## Podsumowanie

W przedstawionych powyżej rozdziałach zobaczyliśmy przykład wykonania metaanalizy wyników badań pierwotnych za pomocą modułu *Metaanaliza* dostępnego w *STATISTICA Zestawie Medycznym*. Analiza umożliwiła ocenę skuteczności witaminy D w prewencji upadków u osób starszych. Wyniki analizy podstawowej przedstawiliśmy w postaci raportu i wykresu typu *forest plot*. Oceniliśmy również niejednorodność wyników poszczególnych badań i przeprowadziliśmy analizę w grupach. Moduł *Metaanaliza* pozwolił przeprowadzić dodatkowo analizę wrażliwości z uwzględnieniem pięciu badań pominiętych w głównej analizie.

## Literatura

1. H.A. Bischoff-Ferrari i wsp., *Effect of Vitamin D on Falls*, JAMA, 28.04.2004, Vol 291, No. 16.
2. M. Borenstein, L.V. Hedges, J.P.T. Higgins, H.R. Rothstein, *Introduction to Meta-Analysis*, John Wiley and Sons Ltd, 2009.
3. P. Gajewski, R. Jaeshke, J. Brożek, *Podstawy EBM czyli medycyny opartej na danych naukowych dla lekarzy i studentów*, Medycyna Praktyczna, Kraków 2008.
4. R. Kay, *Statistical Thinking for Non-Statisticians in Drug Regulation*, John Wiley and Sons Ltd, 2007.