

Metaanaliza in njen pomen za psihološko metodologijo

Andrej Kastrin*
Univerzitetni klinični center Ljubljana, Inštitut za medicinsko genetiko, Ljubljana

Povzetek: Metaanaliza je posebna oblika statistične analize, v kateri na sistematičen način združujemo rezultate posameznih med seboj neodvisnih študij. Glavni namen prispevka je opredeliti metodo metaanalize, opozoriti na nekatere značilnosti in posebnosti pri uporabi ter spodbuditi raziskovalce k njeni uporabi pri raziskovalnem delu. Pravilno izvedena metaanaliza omogoča sistematično spremljanje najnovejših znanstvenih spoznanj, učinkovitejšo izrabo obstoječih podatkov, pomembno prispeva h kakovosti obstoječega znanja o določenem proučevanem fenomenu in lahko služi kot podpora pri gradnji novih raziskovalnih domnev. Zamisel o možnosti združevanja podatkov in rezultatov neodvisnih študij je stara že skoraj pol tisočletja, temelje moderne metaanalize pa je pred dobrimi tremi desetletji postavil Glass, ko je na metodološko rigorozen način ovrgel Eysenckovo domnevo o ničnosti učinka psihoterapije. V članku osvetlimo kronološki potek razvoja metaanalize, predstavimo splošen potek metode ter njene glavne statistične koncepte. Podrobneje predstavimo pojem velikosti učinka, problematiko heterogenosti študij ter dva modela združevanja rezultatov študij v skupno oceno velikosti učinka: model stalnih in model slučajnih učinkov. Na primeru podamo zgled za vizualizacijo rezultatov metaanalize s pomočjo drevesnega in lijakastega diagrama. Z namenom enostavne in hitre podpore k metaanalitični metodologiji smo razvili spletni strežnik RMetaWeb, ki raziskovalcem ponuja možnost interaktivne analize podatkov. Gre za prvo tovrstno spletno orodje, katerega jedro predstavlja okolje R za statistično analizo in grafiko.

Ključne besede: psihološka metodologija, statistika, metaanaliza

Meta-analysis: Its role in psychological methodology

Andrej Kastrin
University Medical Centre Ljubljana, Institute of Medical Genetics

Abstract: Meta-analysis refers to the statistical analysis of a large collection of independent observations for the purpose of integrating results. The main objectives of this article are to define meta-analysis as a method of data integration, to draw attention to some particularities of its use, and to encourage researchers to use meta-analysis in their work. The benefits of meta-analysis include more effective exploitation of existing data from independent sources and contribution to more powerful domain knowledge. It may also serve as a support tool to generate new research hypothesis. The idea of combining results of independent studies addressing the same research question dates back to sixteenth century. Meta-analysis was reinvented in 1976 by Glass, to refute the conclusion of an eminent colleague, Eysenck, that psychotherapy was essentially ineffective. We review some major historical landmarks of meta-analysis and its statistical background. We present the concept of effect size measure, the problem of heterogeneity and two models which are used to combine individual effect sizes (fixed and random

* Naslov/Address: asist. Andrej Kastrin, univ. dipl. psih., Univerzitetni klinični center Ljubljana, Inštitut za medicinsko genetiko, Štajmerjeva ulica 3, 1000 Ljubljana, Slovenija, e-pošta: andrej.kastrin@guest.arnes.si

effect model) in great details. Two visualization techniques, forest and funnel plot graphics are demonstrated. We developed RMetaWeb, simple and fast web server application to conduct meta-analysis online. RMetaWeb is the first web meta-analysis application and is completely based on R software environment for statistical computing and graphics.

Key words: psychological methodology, statistics, meta-analysis

CC = 2200

Kot odgovor na izzive zajemanja, shranjevanja in upravljanja z velikimi količinami podatkov, informacij in znanja se je v zadnjem desetletju uveljavilo raziskovalno področje, ki se imenuje odkrivanje zakonitosti iz podatkov. Gre za odkrivanje implicitnih, doslej neznanih in potencialno uporabnih zakonitosti iz podatkov, z namenom učinkovitejšega odločanja, razvrščanja in napovedovanja. Danes, ko je statistika ena od vodilnih znanstvenih disciplin in eden od paradnih konjev področja odkrivanja zakonitosti iz podatkov, lahko psihologija s ponosom ugotovi, da je mnogo pomembnih konceptov, modelov in teorij v zakladnico statističnega znanja prispevala prav sama. Najbolj pomembni prispevki so Stevensova tipologija merskih lestvic, metoda faktorске analize s konceptom latentnih spremenljivk, multidimenzionalno skaliranje ter široko področje testne teorije. Tudi mnogo pomembnih statistikov je po svoji osnovni izobrazbi psihologov, npr. Donald B. Rubin, Leland Wilkinson, Paul Murrell, če omenimo le nekatere. Če že ne najpomembnejši, pa zagotovo najbolj odmeven metodološki prispevek, vsaj sodeč po številu objav v mednarodnih bibliografskih zbirkah, pa je psihologija dala na področju metaanalize.

Popularna Wikipedia takole povzema razumevanje pojma metaanaliza: "Metaanaliza je statistična metoda, namenjena združevanju rezultatov večjega števila študij, ki se ukvarjajo s proučevanjem podobnega raziskovalnega problema." Metaanalizi nadreden koncept je sistematičen pregled literature (Torgerson, 2003). Gre za metodo pregleda literature, povzemanja in zbiranja kvalitativnih dokazov o nekem raziskovalnem problemu. Metaanalizo opredelimo bolj specifično, kot tehniko pregleda literature, z natančno določeno metodologijo in kvantifikacijo rezultatov podobnih študij s standardno metriko, ki omogoča uporabo statističnih metod kot sredstva analize (Rosenthal, 1991; Wachter, 1988; Wolf, 1986). Pri metaanalizi se bibliografski viri ne uporabljajo za definicijo raziskovalnega problema, ampak njihov pregled predstavlja samostojen problem, ki privede do teoretičnih in empiričnih zaključkov, ki lahko spremenijo ali dopolnijo znanje na nekem področju znanstvenega proučevanja. Njene korenine segajo na področje psihologije in pedagogike, kasneje pa se je močno razširila v praktično vse temeljne in aplikativne znanstvene vede. Metaanaliza, skupaj z nepogrešljivo faktorško analizo in čedalje bolj uporabljeno metodologijo strukturnih enačb, predstavlja močen paket podpore modernemu raziskovalnemu delu v psihologiji.

Zakaj metaanaliza

Povezovanje podatkov in informacij, nenehen razvoj novega znanja in njegovo plemenitenje v praksi morda še nikoli niso bili tako pomembni kot danes. Zlasti v svetu zunaj laboratorijev, inštitutov in univerz pomeni golo kopičenje znanja, brez njegovega pretapljanja v rast in razvoj, izgubo konkurenčnosti.

Kopičenju znanja na nekem znanstvenem področju lahko sledimo z dvema komponentama: zadostnostjo in stabilnostjo (Schmidt, 1992). Komponenta zadostnosti se nanaša na vprašanje količine študij, ki jih potrebujemo, da zadovoljivo opišemo nek fenomen oz. problemsko domeno, komponenta stabilnosti pa na vprašanje skladnosti obstoječega znanja z znanjem, ki ga dobimo na osnovi novih raziskovalnih izsledkov pri ponovljenih merjenjih istega fenomena. Medtem ko na zadostnost vpliva predvsem integracija raziskovalčeve ustvarjalnosti na eni ter zadovoljevanje kriterijev znanstvene uspešnosti na drugi strani, je skladnost bolj objektivna in lažje preverljiva kategorija. V ožjem pomenu besede jo lahko skrcimo na zanesljivost merjenja določenega predmeta znanstvenega proučevanja.

Z vprašanjem zanesljivosti se srečujemo v vseh znanstvenih disciplinah, ki poskušajo svoje raziskovalne domneve preveriti z empiričnimi izsledki. Zanesljivi rezultati so ključnega pomena za doseganje osnovnega cilja znanstvenega raziskovanja, tj. ugotavljanja zakonitosti, ki nam omogočajo pojasnjevanje in napovedovanje opazovanih pojavov (Ferligoj, Leskošek in Kogovšek, 1995). Zanesljivost v širšem smislu pomeni, da bomo s ponavljanjem meritev istega pojava v enakih ali vsaj primerljivih okoliščinah dobili primerljive rezultate. Zanesljivost merjenja je tem večja, čim bolj so razlike v izmerjenih vrednostih posledica dejanskih sprememb merjenega pojava in čim manjši je vpliv slučajnih dejavnikov. V psihologiji je najbolj pereč problem, povezan z zanesljivostjo merjenja, vprašanje relativno majhnih vzorcev. Sedlmeier in Gigerenzer (1989) sta pred leti poročala, da je povprečna statistična moč¹ zaključevanja objavljenih študij v eni od psiholoških revij znašala komaj dobrih 40 %. S podobnim problemom se soočajo praktično vse znanstvene discipline, katerih raziskave so zasnovane na teoriji vzorčnega zaključevanja (prim. npr. Balding, 2006)

Smiseln odgovor na problem zanesljivosti posameznih študij, kliničnih poskusov oz. eksperimentov ponuja njihova integracija v obliki metaanalize. Metaanaliza omogoča večjo moč statističnega zaključevanja pri opazovanju določenega fenomena ter natančno oceno njegove variabilnosti (raztrosa) med študijami. Ideja o združevanju podatkov večih med seboj neodvisnih študij je stara že dobrih 400 let, temelje moderne metaanalize pa je pred slabimi tremi desetletji postavil Gene V. Glass (Smith in Glass, 1977), ko je na metodološko rigorozen način ovrgel smelo

¹ Moč statističnega testa se nanaša na verjetnost zavrnitve ničelne hipoteze, ko ta dejansko ne drži (Cohen, 1992). Statistična moč nekega testa je odvisna od treh parametrov: (i) izbrane ravni tveganja α , (ii) željene velikosti učinka ter (iii) velikosti vzorca.

Eysenckovo tezo o ničnosti učinka psihoterapije. Ustrezno izvedena metaanaliza (i) ponuja sistematične, hitre in zanesljive odgovore na raziskovalne domneve, (ii) zaradi večje količine podatkov povečuje moč statističnega zaključevanja, (iii) daje pregled nad metodologijo izvedbe posameznih poskusov ter nenazadnje (iv) omogoča velik prihranek sredstev na račun ponovitvenih poskusov.

Metaanaliza skozi čas

Metaanaliza ima dolgo preteklost, a razmeroma kratko zgodovino. Metodo ponovljenih merjenj pri merjenju istega pojava je v znanost vpeljal danski astronom Tycho Ottesen Brahe konec 16. stoletja (Plackett, 1958). Kepler je svoje tri slavne zakone, ki opisujejo gibanje planetov okoli Zemlje, osnoval ravno na osnovi njegovih dolgoletnih meritev. Brahe je bil prvi, ki je za zmanjševanje sistematične napake pri merjenju uporabil matematični koncept aritmetične sredine, ki se je v znanosti utrdil šele dobro stoletje kasneje. Drug pomemben miselni preskok v teoriji merjenja, neposredno povezan z razvojem metaanalize, je kombinacija meritev različnih opazovalcev, ki jo je vpeljal francoski matematik in astronom Pierre-Louis Moreau de Mauperuis (Plackett, 1958). Pri merjenju dolžin poldnevniške (meridianske) stopinje si je pomagal z večimi neodvisnimi opazovalci, meritve povprečil in tako empirično potrdil pravilnost Newtonove teorije o sploščenosti Zemlje. V veliki meri so bili prav astronomi tisti, ki so postavili temeljne kamne sodobni teoriji merjenja. George Biddell Airy je leta 1861 ugotovitve svojih stanovskih kolegov povzel v znanstveni monografiji z naslovom "On the algebraical and numerical theory of errors of observations and the combination of observations". Prvi resen poskus združevanja kliničnih rezultatov je na začetku 20. stoletja izvedel Karl Pearson z združitvijo podatkov različnih študij, ki so proučevale vpliv cepiva proti tifoidni mrzlici na različnih vzorcih angleških vojakov (Pearson, 1904). Medicina je potrebovala skoraj 50 let, da je odkrila Pearsonov prispevek (Olkin, 1995). Drug pomemben oče metaanalize je bil slavni britanski statistik in Darwinov naslednik Ronald A. Fisher. V eni od sklepnih monografij (Fisher, 1970) je takole povzel bistvo svojega pogleda na problem integracije rezultatov različnih neodvisnih študij pri merjenju istega pojava:

Pri testiranju statistične značilnosti večih neodvisnih testov se včasih zgodi, da malo oz. noben test ni posamezno statistično značilen, združeni pa dajo vtis, da so verjetnosti (zavrnitve ničelne hipoteze, op. a.) nižje, kot bi bile dobljene po naključju. (str. 99)

Za razliko od Pearsona, ki je združil surove korelacijske koeficiente posameznih študij, sta Fisher in Leonard Tippett naredila korak dlje ter neodvisno drug od drugega izpeljala inovativen postopek združevanja p -vrednosti pri testiranju več neodvisnih ničelnih hipotez (Rosenthal, 1991; Tippett, 1931). Medtem ko je Tippetov prispevek

utonil v zakladnico statistične zgodovine, se Fisherjev obrazec uporablja še danes. Fisher je pokazal, kako lahko m neodvisnih p -vrednosti združimo v enotno mero statistične značilnosti, ki se porazdeljuje po χ^2 porazdelitvi z $2m$ stopnjami prostosti (Fisher, 1970):

$$\chi_{2m}^2 = -2 \sum_{i=1}^m \log_e(p_i)$$

Eden od bolj slavnih primerov uporabe te enačbe v praksi je Gordonov (Gordon, Loveland in Cureton, 1952) poskus združevanja rezultatov študij inteligentnosti, ki sta jih opravila McNemar in Terman (1936).

Okno v svet je metaanalizi odprl Glass. Bolj kot ne zaradi pozitivne osebne izkušnje z lastno psihoterapijo se je spustil v ostro polemiko z eminentnim Eysenckom, zlasti z njegovo trditvijo o ničnosti učinka psihoterapije (Eysenck, 1952, 1965). Glassa štejejo za utemeljitelja sodobne metaanalize, je avtor skovanke metaanaliza ter nosilec nove paradigme v razvoju znanosti (Glass, 1976; Smith in Glass, 1977). Pred dobrimi 30 leti je takole zapisal:

Metaanaliza se nanaša na analizo analize. S terminom označujem statistično analizo velike zbirke rezultatov posameznih študij z namenom integracije novih spoznanj. Predstavlja močno alternativno dosedanji vzročni in pripovedni razlagi rezultatov in lahko služi kot podpora pri osmišljanju velike količine raziskovalnih podatkov. (Glass, 1976, str. 3)

Metaanaliza torej ni le suhoparen skupek statističnih obrazcev, ampak dodelan metodološki okvir za izkop novega znanja iz podatkov in njihovo osmišljanje (Schmidt, 1992). Istega leta je Robert Rosenthal (1976) objavil knjigo z naslovom "Experimenter effects in behavioral research", v kateri je predstavil koncept mer velikosti učinka in s tem sprožil močno kritiko klasične uporabe statističnih testov. Z uvedbo od velikosti vzorca neodvisnih mer razlik med rezultati merjenih spremenljivk je bilo tako dostopno tudi močno statistično orodje za primerjanje različnih študij med seboj. Eden najpomembnejših avtorjev s področja mer velikosti učinka je Jacob Cohen, ki je temelje kritike klasičnega testiranja statističnih domnev predstavil v članku s pomenljivim naslovom "The Earth is round ($p < .05$)" (Cohen, 1994). Leta 1977 je Glass skupaj s sodelavko Mary Lee Smith objavil članek, v katerem sta analizirala 375 neodvisnih študij s skupaj več kot 40.000 udeleženci, ki so z različnimi tehnikami in raziskovalnimi metodami proučevale učinke zdravljenja v različnih smereh psihoterapije, in Eysenckovo domnevo ovrgla (Smith in Glass, 1977). Eysenck je do konca svojega ustvarjalnega življenja ostal vnet nasprotnik takega pristopa k raziskovanju. Najbolj znan in največkrat citiran je njegov članek s provokativnim naslovom "Meta-analysis is an exercise in mega-silliness", v katerem je metaanalizo označil za nevredno metodo resnega znanstvenega dela (Eysenck, 1978).

Od vseh znanstvenih disciplin je metaanalizo najbolj unovčila biomedicina. Rečemo lahko, da je danes v biomedicini metaanaliza ena od najbolj uporabljenih metod zbiranja, analize in interpretacije raziskovalnih rezultatov (Egger, Ibrahim in Smith, 2002). Za razliko od drugih znanstvenih disciplin, je v biomedicini metaanaliza postala prava akademska industrija. Metaanaliza je postavila temelje t. i. znanstveno utemeljene medicine (angl. *evidence based medicine*). Znanstveno utemeljena medicina zajema postopek sistematičnega iskanja, ocenjevanja in nenazadnje tudi uporabe sodobnih raziskovalnih izsledkov kot temelj kliničnega odločanja ter pridobiva na popularnosti v številnih medicinskih disciplinah (Sackett, Rosenberg, Gray, Haynes in Richardson, 1996). Gre za razmeroma novo paradigmo, ki bo tudi medicini počasi omogočila uporabo pravega znanstvenega načina raziskovanja in njenega metodološkega aparata. Z namenom zagotavljanja kvalitetnih sistematičnih pregledov so po svetu ustanovili posebne raziskovalne centre, ki za potrebe medicine in njej sorodnih strok že več kot deset let opravljajo sistematične preglede in metaanalize. Najpomembnejši in najbolj organizirani so Cochranovi centri, poimenovani po angleškem epidemiologu Archiju Cochranu (Hill, 2000). Slovenija spada pod okrilje italijanske podružnice s sedežem v Milanu. Eden od najpomembnejših dosežkov združenja je Cochranova knjižnica (<http://www.cochrane.org/>), spletna zbirka sistematičnih pregledov in metaanaliz s področja biomedicine, med katerimi najdemo tudi precej zadetkov s področja psihologije in psihiatrije.

V zadnjem času so pomembne metodološke prispevke k metaanalizi dodali Nambury S. Raju, Larry V. Hedges, Ingram Olkin, John E. Hunter, Jacob Cohen, and Frank L. Schmidt. Z izjemo dveh, Hedgesa in Olkina, so (bili) vsi ostali avtorji eminentni raziskovalci na področju psihološke metodologije.

Metaanaliza je od prvih resnih poskusov pred 30 leti do danes postala močna veja teoretičnega, še bolj pa aplikativnega znanstvenega raziskovanja. V bibliografski zbirki PsycINFO je od leta 1865 do danes (10. 6. 2008) zbranih 1.168 zapisov, ki na kakršenkoli način omenjajo metaanalizo, od tega je 184 zapisov indeksiranih v specializirani zbirki PsycARTICLES. Googlov Učenjak (angl. *Google Scholar*), iskalnik po arhivih odprtega dostopa in spletnih mestih akademskih institucij nam ponudi 20.800 različnih zadetkov v povezavi z metaanalizo, medtem ko je v zbirki Medline² takih zadetkov kar 30.945.

Računsko jedro metaanalize

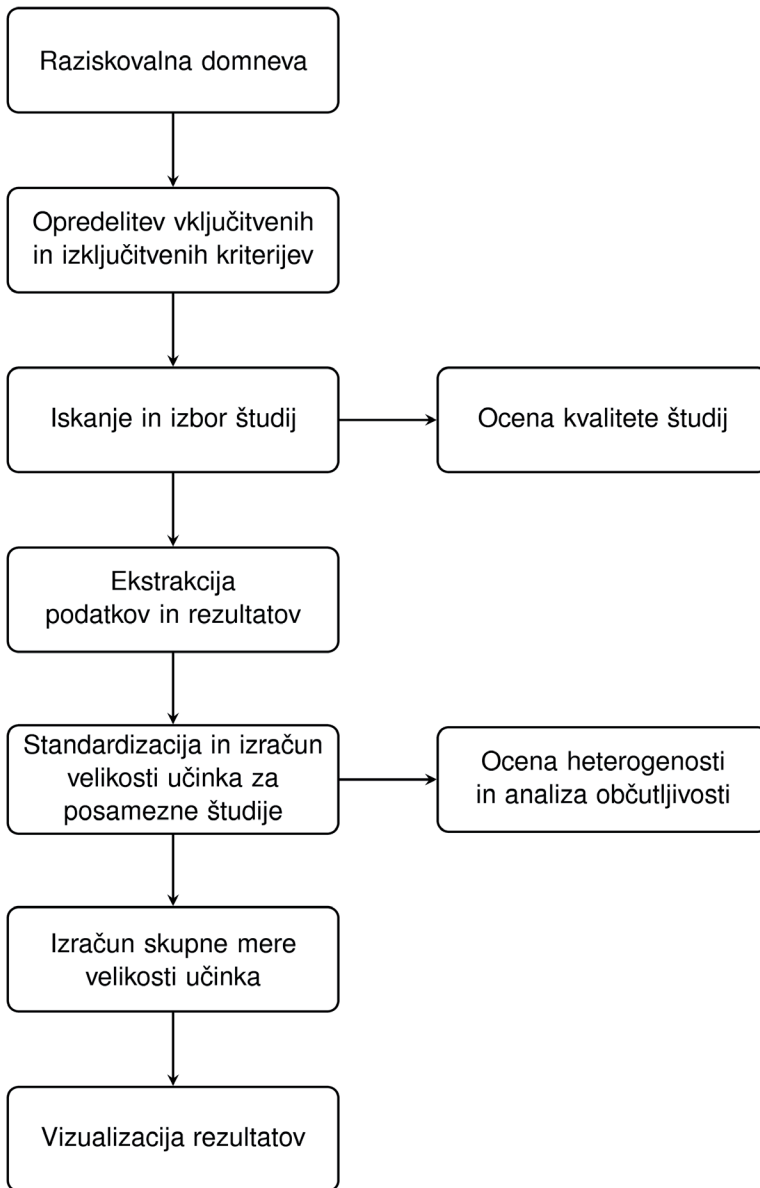
Prvi korak vsake raziskave je jasna opredelitev raziskovalnega problema. Tako kot pri vsaki eksperimentalni ali korelacijski raziskavi moramo tudi v primeru metaanalize natančno opredeliti ničelno domnevo. Jasno moramo opredeliti vključitvene

² Medline je bibliografska zbirka s področja biomedicine in znanosti o življenju in indeksira tudi precej revij s področja psihologije.

in izključitvene kriterije, ki nam bodo omogočali kasnejšo selekcijo med enotami, ki jih bomo vključili v našo raziskavo. Denimo, da želimo s pomočjo metaanalize ovrednotiti moč povezave med izraženostjo ekstravertnosti in potezo iskanja dražljajev v splošni populaciji. Primer vključitvenega kriterija bo v takem primeru npr. zajem vseh tistih študij, ki so kot merska pripomočka uporabili Eysenckov osebnostni vprašalnik EPQ-R za merjenje ekstravertnosti ter Zuckermanovo lestvico iskanja dražljajev SSS-VI, primer izključitvenega kriterija pa izpust študij, ki so testirale skupine udeležencev, za katere je znano, da imajo potezo iskanja dražljajev močnejše izraženo (npr. alpinisti, piloti, vojaki). Kriteriji lahko zajemajo tudi ostale parametre študij, npr. število udeležencev, tip eksperimentalne zasnove, leto in jezik objave študije itd. S pomočjo čimbolj učinkovitih iskalnih strategij nato iz vseh dostopnih bibliografskih in ostalih podatkovnih skladišč poiščemo vse zadetke, ki se nanašajo na predmet naše raziskave. V ožji izbor uvrstimo vse tiste zadetke, ki zadostijo vključitvenim kriterijem, ter izločimo zadetke, ki zadostijo izključitvenim kriterijem. Prvemu situ lahko dodamo še drugo sito, s katerim preverjamo kvaliteto posameznih študij. V ta namen so najbolj uporabni različni standardizirani postopki (npr. ocenjevalne liste), s katerimi ocenimo primernost posamezne študije (Wood, 2008). Faze ocenjevanja primernosti študij za vključitev v metaanalizo je smiselno izvesti paroma, z dvema neodvisnima ocenjevalcema. Sledi faza ekstrakcije podatkov, v kateri pripravimo podatkovno zbirko, ki za vsako študijo vključuje podatke o izmerjenih spremenljivkah ter za študijo specifične parametre. Slednji nam lahko v kasnejših korakih raziskovanja služijo kot sospremenljivke pri dodatnih analizah³. Surove podatke standardiziramo, s čimer dosežemo primerljivost rezultatov med študijami, ter izračunamo velikost učinka po posameznih študijah. S podporo namenskih računalniških programov nato izračunamo še oceno skupne velikosti učinka, katere stabilnost lahko potem preverjamo s sistematičnim izključevanjem posameznih študij (analiza občutljivosti). Povzetek opisanih korakov je prikazan na Sliki 1.

Glavni statistični koncept, s katerim operiramo v metaanalizi, je mera velikosti učinka (Hedges in Olkin, 1985). Dober teoretično-praktični uvod v problematiko mer velikosti učinka ponuja članek domačih avtorjev Cankarja in Bajca (2003). Mere velikosti učinka lahko taksonomsko razdelimo v dve veliki skupini. V prvo skupino uvrščamo mere standardiziranih razlik med proučevanimi skupinami, v drugo pa mere povezanosti. Nekateri avtorji tej grobi razvrstitvi dodajajo še intervale zaupanja (Hedges in Olkin, 1985). Pri standardiziranih razlikah gre običajno za opisovanje razdalj med aritmetičnimi sredinami proučevanih skupin v enotah standardnih odklonov, pri merah povezanosti pa za stopnjo povezanosti med učinkom in odvisno spremenljivko. Glede na tip spremenljivk je tako prva skupina mer uporabna zlasti na merskih ravneh, ki zadostijo vsaj ordinalni merski lestvici, druga skupina pa za

³ Klasični primer je npr. metaregresija. Odvisno spremenljivko predstavlja velikost učinka študije, prediktorske spremenljivke pa predstavljajo parametri študij (npr. starost, spol, socialno-ekonomski status udeležencev, tip eksperimentalnega načrta itd.).



Slika 1. Diagram poteka skozi glavne faze metaanalize.

nominalne (imenske) podatke. Med najbolj znane mere za ocenjevanje razlik spadajo Cohenov d , Glassov Δ in Hedgesova g statistika, med mere povezanosti pa statistike, ki računajo delež pojasnjene variance (npr. Pearsonov r koeficient korelacije, Cramerjev V koeficient povezanosti med nominalnima spremenljivkama itd.) ter ostale, v psihologiji manj znane mere povezanosti, izpeljane iz verjetnostnega računa (npr. razmerje obetov, relativno tveganje, pripisljivo tveganje itd.). Odličen pregled

različnih mer velikosti učinka ponuja Cohenova monografija (Cohen, 1988).

Heterogenost se nanaša na raztros posameznih ocen velikosti učinka preko študij (Whitehead, 2002). V grobem ločimo dve vrsti heterogenosti, vzorčno heterogenost in metodološko heterogenost. Prva se nanaša na razlike v vzorcih udeležencev, npr. razlike v spolu, starosti, socialno-ekonomskem položaju itd., druga pa na razlike v uporabljeni metodi merjenja proučevanega predmeta raziskave, na razlike v eksperimentalnih zasnovah in kvaliteti izvedbe posameznih študij. Oba vira heterogenosti imata za posledico statistično heterogenost, ki se nanaša na opazovane razlike med velikostmi učinkov posameznih študij. Določena stopnja klinične in metodološke heterogenosti med študijami vedno obstaja, v metaanalizi pa si prizadevamo, da je čim manjša.

Celotna teorija statističnega opisovanja in zaključevanja temelji na predpostavkah in modelih, zato moramo tudi pri metaanalizi opredeliti statistični model, na osnovi katerega bomo iz dosežkov po posameznih študijah sklepali na povprečno oceno velikosti učinka (Hedges in Olkin, 1985). V metaanalizi se najpogosteje uporabljata dva statistična modela (Hunter in Schmidt, 2000; Kisamore in Brannick, 2008): model stalnih učinkov (angl. *fixed effect model*) in model slučajnih učinkov (angl. *random effect model*). Model stalnih učinkov predpostavlja, da vključene študije ocenjujejo isti učinek oz. da so učinki posameznih študij vzorčne vrednosti iste populacije (Mantel in Haenszel, 1959). Na ta način upoštevamo le raztros znotraj posameznih študij. Predpostavka tega modela je torej, da vse vključene študije uporabljajo enako metodo merjenja, vzorci udeležencev pa so med seboj homogeni. Model slučajnih učinkov po drugi strani predpostavlja, da so vključene študije naključno vzorčene iz različnih populacij študij, ki imajo različne učinke (DerSimonian in Kacker, 2007; DerSimonian in Laird, 1986). Pri tem poleg raztrosa znotraj posameznih študij upoštevamo tudi raztros med študijami. Pri uporabi modelov moramo biti pazljivi, saj lahko modela na istih podatkih pripeljeta do popolnoma različnih rezultatov. V primeru, da so študije med seboj homogene, modela stalnih in slučajnih učinkov vrneta praktično primerljive rezultate. Teoretično gledano je prav heterogenost med študijami tista, ki definira izbiro statističnega modela. V primeru heterogenosti med študijami, ki ni posledica razlik nad populacijo vključenega univerzuma udeležencev v merjeni lastnosti, ampak je posledica raztrosa med posameznimi študijami, pa uporabimo model slučajnih učinkov. Za preverjanje predpostavke homogenosti posameznih študij lahko uporabimo različne testne statistike in grafične metode. Najbolj enostaven indikator statistične heterogenosti je slabo prekrivanje intervalov zaupanja, za numerično oceno pa uporabimo različne mere, izpeljane iz χ^2 in F preizkusa. Heterogenost študij sama po sebi ni nujno ovira. Resda je z vidika združevanja rezultatov moteča, vendar nam lahko služi tudi kot indikator vsebinskih razlik med študijami. Nekateri avtorji v tem okviru govorijo celo o dvojni naravi metaanalitičnih študij: njeni klasični, analitični obliki dodajajo še eksploratorni vidik, katerega namen je odkrivanje razlik in pojasnjevanje virov heterogenosti študij (Schmidt, 1992).

Običajno je tako, da podatkom najprej priredimo model s stalnimi učinki in preverimo njegovo prilaganje podatkom. Če empirični model statistično značilno

odstopa od pričakovanega, podatkom priredimo še model s slučajnimi učinki (Field, 2003). Model s slučajnimi učinki je v statističnem smislu bolj konzervativen in vrne širši interval zaupanja končne ocene velikosti učinka kot model s stalnimi učinki.

Vizualizacija rezultatov

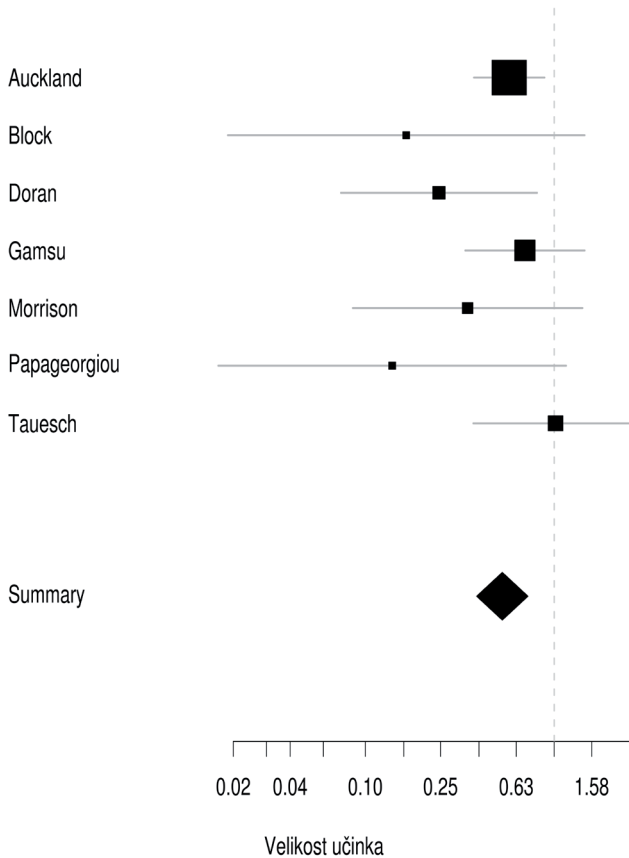
Vizualizacija podatkov se je z razmahom podatkovnega rudarjenja otesla priokusa nepotrebne okrasja in postala pomembna znanstvena disciplina. Vizualizacija podatkovnih struktur in rezultatov statističnih analiz je tako sestavni del ali pa celo končni cilj številnih statističnih orodij in metod. Metoda metaanalize se ponša z dvema vrstama grafik, po katerih je že ob bežnem prelistavanju znanstvene periodike tudi najbolj prepoznavna: drevesni diagram (angl. *forest plot*) je grafični prikaz vrednosti ocen velikosti učinkov posameznih študij ter njihove skupne ocene, lijakasti diagram (angl. *funnel plot*) pa dvorazsežni prikaz odnosa med velikostjo učinka študij in velikostjo njihovega raztrosa.

Drevesni diagram je ključni grafični povzetek rezultatov metaanalize (Slika 2). Idejo drevesnega diagrama je prvi predstavil Stephen Evans na osnovi razširjenega diagrama okvirja z ročaji (angl. *boxplot*; Lewis in Clarke, 2001). Velikost učinka posamezne študije predstavimo s kvadratom. Velikost kvadrata je odvisna od uteženosti študije, tj. od velikosti njenega vzorca. S horizontalno črto prikažemo interval zaupanja, v katerem z določeno stopnjo tveganja pričakujemo oceno. Običajno se uporablja 95-odstotni interval zaupanja. Skupno velikost učinka prikažemo z rombo. Interval zaupanja skupne ocene predstavljajo horizontalna oglišča romba. Ko imamo opravka z diskretnimi podatki in merami velikosti učinka, izpeljanimi iz verjetnosti dogodkov (npr. razmerja obetov), ocene velikosti učinka običajno logaritmiramo, s čimer zagotovimo simetričnost njihovih intervalov zaupanja (Agresti, 2002).

Lijakasti diagram je drugi standardni grafični prikaz v metaanalizi (Slika 3). Lijakasti diagram predstavlja odnos med velikostjo učinka študij in njihovimi ocenami raztrosa (Egger, Smith, Schneider in Minder, 1997). Uporablja se za odkrivanje sistematične heterogenosti (Song, Khan, Dinnes in Sutton, 2002). Pri popolni homogenosti je oblika diagrama simetrično lijakasta, asimetričnost pa nakazuje prisotnost heterogenosti, ki je lahko posledica različnih virov (omejenega izbora študij, vključenih v metaanalizo, slabe kakovosti študij, majhnih vzorcev, prave heterogenosti itd.).

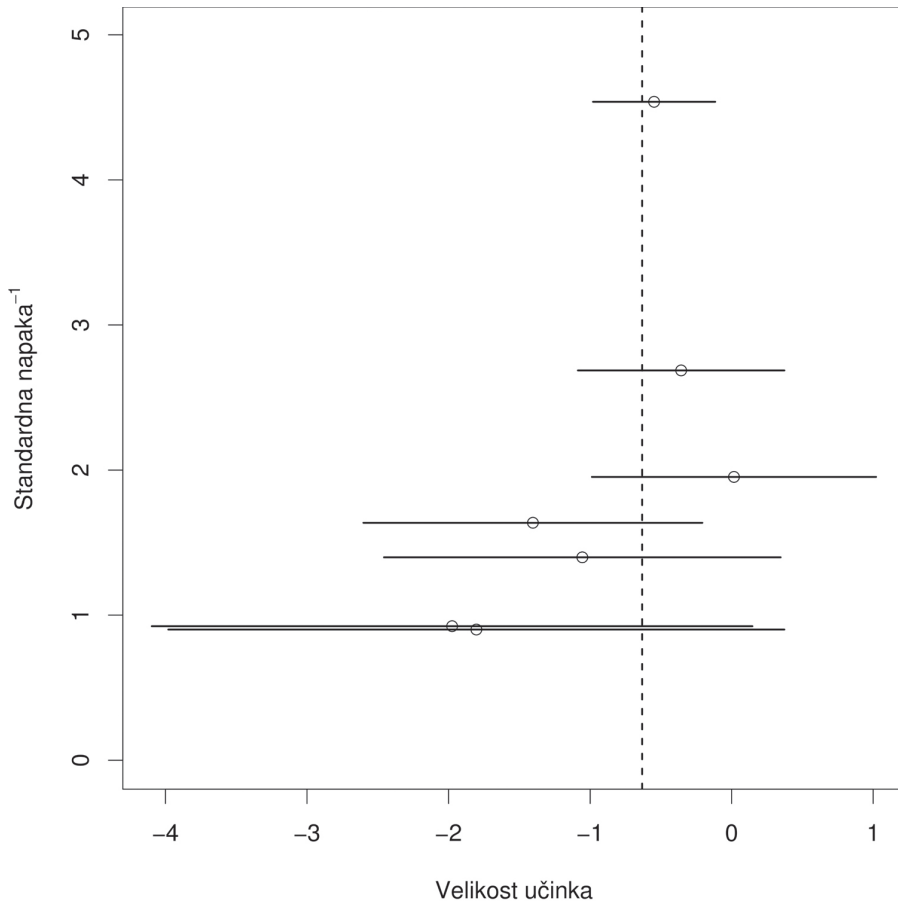
Kritika metaanalize

Po burnih začetnih odzivih Eysencka in raziskovalcev iz njegovega kroga je metaanaliza kot metoda tudi sama postala predmet znanstvenega proučevanja. Veliko podiplomskih študijskih programov, tako družboslovnih, humanističnih kot



Slika 2. Drevesni diagram. Diagram prikazuje velikosti učinkov za sedem različnih študij, ki so proučevale vpliv kortikosteroidne terapije na prezgodnji porod in neonatalno smrtnost. Surovi podatki so prosto dostopni na Cochranovem spletnem skladišču (<http://www.cochrane.org/>). Podatki so bili izvorno diskretni, mero velikosti učinka pa je predstavljalo razmerje obetov. Velikost učinka je prikazana v logaritemskih enotah. Velikost kvadratka je sorazmerna velikosti vzorca, na katerem je bila študija izvedena. Horizontalna črta označuje 95-odstotni interval zaupanja. Skupna ocena velikosti učinka je predstavljena z romбом, katerega levo in desno oglišče predstavljata 95-odstotni interval zaupanja skupne ocene. Skupna velikost učinka v našem primeru znaša 0,53, na osnovi česar lahko zaključimo, da je tveganje prezgodnjega poroda oz. neonatalne smrti pri nosečnicah s kortikosteroidnim zdravljenjem približno polkrat manjše kot pri kontrolni skupini. Bralec bo več informacij v zvezi z interpretacijo razmerja obetov našel v Agrestijevi monografiji (Agresti, 2002).

tudi naravoslovnih, v svojih predmetnikih že vključuje izbrana poglavja s področja metodologije metaanalize. Nekatera združenja organizirajo celo večdnevne študijske delavnice, na katerih študenti spoznajo uporabo metod metaanalize v različnih problematskih situacijah. Dandanes raziskovalci z različnih področij proučujejo uporabo



Slika 3. Lijakasti diagram. Diagram prikazuje velikost učinka študij v odnosu do mer njihovega raztrosa. Horizontalna črta označuje 95-odstotni interval zaupanja posameznih študij. Oblika diagrama je približno simetrična, kar nakazuje, da je homogenost študij zadovoljiva.

različnih mer velikosti učinka nad različnimi tipi podatkov, proučujejo veljavnost mer ocenjevanja heterogenosti med študijami in razvijajo nove računske metode za povzemanje skupnega učinka (npr. Bayesovi modeli). Na splošno lahko kritike metode metaanalize razčlenimo v treh sklopih.

Metaanaliza zahteva, da so podatki, s katerimi vstopamo v analizo, med seboj primerljivi. V metodoloških učbenikih je ta problem opisan z metaforo združevanja jabolk in pomaranč. Nehomogenost vzorcev, različni tipi eksperimentalnih zasnov in vsebinsko močno različna operacionalizacija merskih spremenljivk v posameznih študijah so glavni viri sistematičnih napak lahkovernih (meta)raziskovalcev.

Drug glavni problem je kvaliteta študij, vključenih v metaanalizo. Problem izbora kvalitetnega in reprezentativnega vzorca je ključnega pomena ne samo za

ustrezno statistično oceno proučevanega fenomena, pač pa tudi za generalizabilnost zaključkov. Stabilnost skupnega učinka, izračunanega v metaanalizi, je smiselno preveriti s pomočjo analize občutljivosti. Pri analizi občutljivosti gre za to, da opazujemo raztros skupne ocene velikosti učinka v odvisnosti od nabora študij, vključenih v metaanalizo, od vrste uporabljene metode za izračun skupne ocene, eksperimentalne zasnove itd. Pri dobro zasnovanih metaanalizah je tako pričakovati, da se z izločitvijo študij z majhnim številom udeležencev skupna ocena ne bo statistično značilno spremenila.

Uredniška politika znanstvenih časopisov je v veliki večini primerov taka, da raziskovalci praviloma lažje objavijo raziskavo, ki poroča o statistično značilnih rezultatih, kot pa raziskavo, ki do takih izsledkov ni prišla. Kljub temu, da obstajajo namenske revije, ki objavljajo prav take študije (npr. *Journal of Negative Results in BioMedicine*), jih večina pristane na smetišču zgodovine. Heterogenosti, ki iz tega izhaja, pravimo pristranost v objavljanju (angl. *publication bias*). Zato je ključnega pomena, da v metaanalizo poleg klasičnih znanstvenih člankov, ki jih indeksirajo mednarodne bibliografske zbirke (npr. PsycINFO, Medline), vključimo tudi pregledne članke, povzetke in programe konferenc in simpozijev, diplomske, magistrske in doktorske naloge, raziskovalna poročila in elaborate, neindeksirane strokovne članke, poročila vladnih in nevladnih organizacij in bibliografije (Rothstein, Sutton in Borenstein, 2005).

Programska oprema

Pestrost numeričnih postopkov za analizo zbranih podatkov na različnih tipih merskih lestvic narekuje razvoj specializirane programske opreme, ki tudi manj večjemu uporabniku omogoča kvalitetno in metodološko ustrezno izvedbo metaanalize. Statistično bolj podkovani raziskovalci lahko metaanalizo opravijo z uporabo splošnih statističnih programskih paketov (npr. SPSS, SAS) ali celo z uporabo elektronske preglednice (npr. Excel, OpenOffice Calc), medtem ko večina raziskovalcev raje poseže po namenski programski opremi. Izbor nekaterih programov je povzet v Tabeli 1. Med njimi najdemo tako komercialne kot tudi prostodostopne in odprtokodne programe. Glavna slabost vseh programov je razmeroma dolg čas ogrevanja, saj za osvojitve osnovnih principov manipulacije s podatki običajno potrebujemo dalj časa kot za samo izvedbo metaanalize. Z namenom enostavne, hitre in učinkovite podpore k metaanalitični metodologiji smo zato razvili spletni strežnik RMetaWeb, ki uporabniku ponuja možnost interaktivne online analize podatkov. Gre za prvo tovrstno spletno orodje in predstavlja močno alternativo obstoječi programski opremi. Numerično jedro spletnega strežnika predstavlja okolje R za statistično analizo in grafiko (R Development Core Team, 2008).

R sta pred dobrim desetletjem zasnovala Ross Ihaka in Robert Gentleman z Univerze v Aucklandu, Nova Zelandija (Ihaka in Gentleman, 1996). V desetih letih je R postal vodilni statistični programski paket in programski jezik. Gre za

Tabela 1. Programska podpora za meta analizo.

Program	URL naslov
Comprehensive Meta-Analysis*	http://www.meta-analysis.com/
EasyMA	http://www.spc.univ-lyon1.fr/easy.ma/dos/
EpiMeta	http://ftp.cdc.gov/pub/Software/epimeta/
Hepima	http://www.hsph.harvard.edu/faculty/spiegelman/tcs.html
Meta-Analysis 5.3	http://userpage.fu-berlin.de/~health/meta_e.htm
Meta-Analyst	http://www.medepi.net/meta/MetaAnalyst.html
Meta-Stat	http://edres.org/meta/metastat.htm
MetaWin*	http://www.metawinsoft.com/
MIX	http://www.mix-for-meta-analysis.info/
RevMan	http://www.cc-ims.net/RevMan
WEasyMA*	http://www.weasy.ma.com/
RMetaWeb	http://www2.arnes.si/~akastr1/

Opomba: Z zvezdico (*) so označeni plačljivi programi.

odprtokodno (in torej brezplačno) implementacijo predmetnega jezika S, ki so ga zasnovali v Bellovih laboratorijih. Bistvo okolja R predstavlja njegova nadgradljivost s programskimi paketi. Osnovna namestitvena distribucija vključuje le osnovni nabor paketov, ostale pakete pa si uporabnik namešča po potrebi preko omrežja zrcalnih strežnikov CRAN (angl. *The Comprehensive R Archive Network*). Okolje R je na voljo za različne izvedbe operacijskih sistemov (Unix, Linux, Mac OS, Windows) in ponuja zelo obsežno programsko podporo za delo s podatki, računske operacije in grafične prikaze. Njegove glavne prednosti v primerjavi z ostalimi, bolj razširjenimi in poznanimi statističnimi programi so predvsem: (i) visoka razvitost, razumljivost in preglednost programskega jezika, ki vključuje široko paleto podatkovnih struktur, zanke, rekurzivne klice in pogojno izvajanje programske kode; (ii) hitre računske operacije nad vektorji in matrikami; (iii) učinkovito shranjevanje in priklic podatkov; (iv) zmogljivi grafični podsistemi in paketi za vizualizacijo podatkov ter (v) neomejena razširljivost, ki omogoča integracijo z drugimi programskimi jeziki (Fortran, C, C++, Java, Perl, Python).

Poleg enostavnega dostopa do klasičnih orodij za analizo in vizualizacijo eno- in dvodimenzionalnih podatkov ponuja R zahtevnejšemu raziskovalcu s področja psihologije tudi celo paleto orodij za napredne statistične analize. Pred kratkim so bili razviti paketi, ki omogočajo preprost dostop do funkcij za analizo modelov latentnih potez, strukturnega modeliranja in korespondenčne analize. Metaanalizi sta namenjena dva, med seboj razmeroma podobna paketa: *rmeta* in *meta*. Funkcije slednjih dveh smo uporabili tudi za razvoj spletnega strežnika RMetaWeb (<http://www2.arnes.si/~akastr1/>).

Spletni strežnik RMetaWeb je v celoti razvit z uporabo prostih in odprtokodnih orodij. Dinamičen pretok med numeričnim delom in spletnim prikazom je implementiran s pomočjo paketa CGIwithR (Firth, 2003). Uporabnik datoteko s surovimi podatki najprej pretoči na spletni strežnik in označi strukturo vhodne datoteke. Trenutno podprti formati vhodnih podatkov so standardne tekstovne datoteke (tsv, csv), excelove (xls) ali SPSS (sav) datoteke. V drugem koraku izbere želeno mero velikosti učinka ter metodo združevanja rezultatov. Strežnik nato izpiše rezultate metaanalize za dane podatke ter ponudi možnost prenosa rezultatov v obliki tekstovne datoteke, ki si jo uporabnik lahko shrani za kasnejšo uporabo. V tretjem koraku sistem izriše drevesni in lijakasti diagram. Odvisno od namena uporabe lahko velikost slike poljubno spreminjamo. Uporabniku je na voljo tudi možnost prenosa diagramov v PNG in PDF zapisu. PNG je enoznačen slikovni format in je namenjen predvsem ogledu slik na spletu ali predstavitvam, medtem ko je slika v PDF formatu shranjena v vektorski obliki in jo lahko uporabimo neposredno za tisk.

V trenutni implementaciji RMetaWeb omogoča analizo zveznih in diskretnih spremenljivk preko različnih mer velikosti učinka za statistične modele s stalno in slučajno eksperimentalno zasnovo. Odvisno od izkušenosti in potreb, uporabnik kombinira različne mere velikosti učinka in metode za izračun njene skupne ocene. Pri diskretnih podatkih lahko uporabnik izbira med merami razmerja obetov, razlik v tveganju (oz. ogroženosti) in relativnim tveganjem, pri zveznih podatkih pa je trenutno na voljo Hedgesova g statistika. Za izračun skupne velikosti učinka so na voljo tri standardne metode: Mantel-Haenszelova, Petova in metoda inverzne variance.

V nadaljevanju bomo implementirali še algoritem za analizo občutljivosti posameznih študij ter ponudili interaktivne obrazce za pretvarjanje različnih mer velikosti učinka in izračunavanje njihovih intervalov zaupanja.

Zaključek

Newton je leta 1676 v pismu Hooku zapisal: "Če sem videl dlje, je to zaradi tega, ker sem stal na ramenih velikanov." Z uvedbo metaanalize (ali pa njenim ponovnim odkritjem) je psihologija za skoraj dvajset let prehitela razvoj danes zelo opevanega področja odkrivanja zakonitosti iz podatkov. Upali bi si celo trditi, da je metoda metaanalize eden od temeljni kamnov vseh sodobnih tehnologij znanja.

Metaanaliza sicer zahteva razmeroma velik vložek dela in premišljeno kombiniranje kvalitativne in kvantitativne analize, vendar po drugi strani omogoča sprotno in sistematično spremljanje najnovejših znanstvenih spoznanj, učinkovitejšo izrabo obstoječih podatkov, pomembno prispeva h kakovosti obstoječega znanja o določenem proučevanem fenomenu in nenazadnje nudi podporo pri gradnji novih raziskovalnih domnev.

Uporaba katerekoli statistične metode zahteva veliko znanja in izkušenj. Zavedati se moramo, da metaanaliza ni nadomestek raziskovalčeve ustvarjalnosti, pač pa le orodje, ki lahko močno pospeši in izboljša kvaliteto raziskovalnega dela.

Literatura

- Agresti, A. (2002). *Categorical data analysis* (2. izd.). Hoboken, NJ: Wiley.
- Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, 7(10), 781–791.
- Cankar, G. in Bajec, B. (2003). Velikost učinka kot dopolnilo testiranju statistične pomembnosti razlik [Effect size as a supplement to statistical significance testing]. *Psihološka obzorja*, 12(2), 97–112.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. izd.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997–1003.
- DerSimonian, R. in Kacker, R. (2007). Random-effect model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials*, 28(2), 105–114.
- DerSimonian, R. in Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–188.
- Egger, M., Ebrahim, S. in Smith, G. D. (2002). Where now for meta-analysis? *International Journal of Epidemiology*, 31(1), 1–5.
- Egger, M., Smith, G. D., Schneider, M. in Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629.
- Eysenck, H. J. (1952). The effect of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 16(5), 319–324.
- Eysenck, H. J. (1965). The effects of psychotherapy. *International Journal of Psychiatry*, 1, 97–142.
- Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, 33(5), 517.
- Ferligoj, A., Leskošek, K. in Kogovšek, T. (1995). *Zanesljivost in veljavnost merjenja [Reliability and validity of measurement]*. Ljubljana: Fakulteta za družbene vede.
- Field, A. P. (2003). The problems in using fixed-effect models of meta-analysis on real-world data. *Understanding Statistics*, 2(2), 105–124.
- Firth, D. (2003). CGIwithR: Facilities for processing web forms using R. *Journal of Statistical Software*, 8(10), 1–8.
- Fisher, R. A. (1970). *Statistical methods for research workers*. New York: MacMillan.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8.
- Gordon, M. H., Loveland, E. H. in Cureton, E. E. (1952). An extended table of chi-square for two degrees of freedom, for use in combining probabilities from independent samples. *Psychometrika*, 17(3), 311–316.
- Hedges, L. V. in Olkin, I. (1985). *Statistical methods for meta-analysis*. Boston, MA: Academic Press.

- Hill, G. B. (2000). Archie Cochrane and his legacy. An internal challenge to physicians' autonomy? *Journal of Clinical Epidemiology*, 53(12), 1189–1192.
- Hunter, J. E. in Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8(4), 275–292.
- Ihaka, R. in Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
- Kisamore, J. L. in Brannick, M. T. (2008). An illustration of the consequences of meta-analysis model choice. *Organizational Research Methods*, 11(1), 35–53.
- Lewis, S. in Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *British Medical Journal*, 322(7300), 1478–1480.
- Mantel, N. in Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719–748.
- McNemar, Q. in Terman, L. M. (1936). Sex differences in variational tendency. *Genetic Psychology Monographs*, 18(1), 31.
- Olkin, I. (1995). Statistical and theoretical considerations in meta-analysis. *Journal of Clinical Epidemiology*, 48(1), 133–146.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 2(2288), 1243–1246.
- Plackett, R. L. (1958). Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. *Biometrika*, 45(1–2), 130–135.
- R Development Core Team (2008). *R: A language and environment for statistical computing* [Računalniški software]. Dunaj, Avstrija. (ISBN 3-900051-07-0)
- Rosenthal, R. (1976). *Experimenter effect in behavioral research* (razširjena izd.). New York: Halsted Press.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Rothstein, H. F., Sutton, A. J. in Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustment*. London: Wiley.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B. in Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312(7023), 71–72.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47(10), 1173–1181.
- Sedlmeier, P. in Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105(2), 309–316.
- Smith, M. L. in Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32(9), 752–760.
- Song, F., Khan, K. S., Dinnes, J. in Sutton, A. J. (2002). Asymmetric funnel plots and publication bias in meta-analysis of diagnostic accuracy. *International Journal of Epidemiology*, 31(1), 88–95.
- Tippet, L. H. C. (1931). *The methods of statistics*. London: Williams & Norgate.
- Torgerson, C. (2003). *Systematic reviews*. London: Continuum.

- Wachter, K. W. (1988). Disturbed by meta-analysis. *Science*, 241(4872), 1407–1408.
- Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*. West Sussex: Wiley.
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park, CA: Sage.
- Wood, J. (2008). Methodology for dealing with duplicate study effects in a meta-analysis. *Organizational Research Methods*, 11(1), 79–95.