

## Opis i tumačenje metodoloških i statističkih tehnika u meta-analizama

### Describing and interpreting the methodological and statistical techniques in meta-analyses

Alfred A. Bartolucci

Odsjek za biostatistiku, Sveučilište Alabame u Birminghamu, Birmingham, Alabama, SAD  
Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama, USA

#### Sažetak

Pojam meta-analize udomaćio se kao termin koji označava tumačenje rezultata iz znanstvene literature i stvaranje konačnog zaključka o prihvatljivosti rezultata istraživanja kojima se ispituju slične intervencije, tehnike ili liječenja. Meta-analiza je postala neophodna u razumijevanju velike zbirke neobrađenih podataka ili literature koja je ponekad proturječna, nedosljedna i nejasna o nekoj temi te u razumijevanju prave važnosti statističkih rezultata kod obrade neke znanstvene teme kao što je, primjerice, učinkovitost. Ne treba samo razumjeti motivaciju za meta-analizu, već i tehnike kojima sumiramo rezultate različitih istraživanja i koje pružaju odgovarajuće statističko promatranje i tumačenje rezultata. Poseban izazov predstavljaju različiti ustroji istraživanja, različiti kriteriji prihvaćanja za uključanje u neko istraživanje, posebno u medicinsko-kliničko ispitivanje, te mogući različiti pristupi tumačenju rezultata, iako ovo zadnje nije prevladavajuće. Općenito govoreći, istraživanja koja obuhvaćaju ispitivanje uloge neke intervencije kao što je, primjerice, sniženje koncentracije kolesterola i njen utjecaj na kardiovaskularno zdravlje, obrađuju sličnu temu te ako se rezultati tih istraživanja ne podudaraju, uloga meta-analize je usporediti sve rezultate i donijeti krajnji zaključak o pozitivnom, negativnom ili nejasnom utjecaju smanjene koncentracije lipida na kardiovaskularno zdravlje. Prikaz pristupa meta-analizi je jako vrijedan budući da u većini slučajeva nudi objašnjenje statističkih pristupa, njihovog značenja, svrhe i krajnjeg utjecaja na tumačenje meta-analize. Stoga bi istraživač, kad ima na raspolaganju elemente meta-analize, trebao razumjeti prednosti i ograničenja tog načina tumačenja rezultata iz literature.

**Ključne riječi:** meta-analiza; statistika; pregled; sustavna pogreška

#### Abstract

The term, meta-analysis, has become a regular description for interpreting results from the scientific literature and hopefully reaching a definitive conclusion concerning the reasonableness of results of studies investigating similar interventions, techniques or treatments. It has become a necessity for understanding the sometimes huge repository of conflicting, inconsistent or inconclusive literature or raw data on a subject and understanding the true relevancy of the statistical results in addressing a scientific question such as efficacy. One has to understand, not only the motivation for a meta-analysis, but the techniques employed to help synthesize results from different studies giving them appropriate statistical scrutiny and interpreting the conclusions. A particular challenge is the differing designs of studies, differing eligibility criteria for entry into the study, especially in the medical clinical trial and possibly non similar approaches in interpreting the results, although this latter point is not too prevalent. For the most part studies investigating the role of a particular intervention such as lowering cholesterol and the impact on cardiovascular health address a similar question and should studies not all agree on the results, the role of meta-analysis is to hopefully put all the results together and reach a definitive conclusion about the positive, negative or inconclusive impact on lowering lipid levels and cardiovascular health. One should appreciate an overview of the approach to meta-analysis assuming mostly the understanding of the statistical approaches, their meaning, purpose and ultimate impact on the interpretation of the meta-analysis. Thus having been exposed to the elements of the meta-analysis one should understand the advantages and limitations of this approach to interpreting results from the literature.

**Key Words:** meta-analysis; statistics; review; bias

*Pristiglo: 8. siječnja 2009.*

*Prihvaćeno: 17. ožujka 2009.*

*Received: January 8, 2009*

*Accepted: March 17, 2009*

## Uvod

Meta-analiza je u svojem sadašnjem obliku statističke integracije podataka iz nekoliko istraživanja s istom temom prisutna već više od 30 godina. Medicina, industrija i temeljna znanost svjedoci su brojnih pokušaja raznih znanstvenika da sumiraju podatke iz različitih izvora unutar jedne znanstvene discipline s ciljem davanja konačne izjave o najnovijim znanstvenim spoznajama u toj disciplini. Slično tome, autori publikacija u uvodima svojih radova uvijek sumiraju ono za što vjeruju da je trenutno sveukupno stanje spoznaja unutar struke do trenutka prezentiranja svojih rezultata. Novi podaci i rezultati koje izlažu u svojoj publikaciji predstavljaju pokušaj ažuriranja napretka u tom polju. Na taj način smatraju da su izvršili djelomičnu meta-analizu sumirajući informacije iz prošlosti i dodavajući tome svoj prilog te na taj način obnavljaju bazu znanja. Oni zapravo nisu u potpunosti integrirali podatke iz prošlosti strogim statističkim postupkom, već su samo upotrijebili podatke iz arhive za opravdanje svog istraživanja koje je u biti nezavisno.

Na taj način meta-analiza postaje naknadni pokušaj prikupljanja trenutne baze znanja u obliku publikacija ili sirovih podataka. Ona predstavlja statističku sintezu svih informacija te donosi zaključak o najboljoj strategiji liječenja ili intervencije na temelju svih prijašnjih istraživanja. Prema riječima Glassa (1): „Meta-analiza se odnosi na analizu analiza...statističku analizu velike zbirke rezultata analiza iz pojedinih istraživanja sa svrhom integriranja rezultata“. U većini disciplina statistička analiza može obrađivati završne točke kao što su to, primjerice, veličina učinka (engl. *mean effects*), stopa odgovora (engl. *response rates*), omjeri izgleda (engl. *odds ratios*, OR), korelacije (engl. *correlations*) i slično. Ovdje želimo pokazati statistička pitanja meta-analize na primjeru, jer u ovom pregledu ne možemo proći sve završne točke. Stoga ćemo se koncentrirati na kontinuirane završne točke veličine učinka, imajući na umu ostale kontinuirane (npr. korelacije) i kategoričke (npr. omjer izgleda, stopa odgovora, relativni rizik (engl. *relative risk*)) završnim točkama. Započeti ćemo s obradom izvora pregleda literature za meta-analizu. U ostatku članka obraditi ćemo statističku terminologiju i pristupe izradi meta-analize. Nakon toga ćemo predstaviti tehnike istraživanja sustavnog odstupanja rezultata zbog pretežitog objavljivanja određenog tipa istraživanja (engl. *publication bias*) i zaključiti o prednostima i ograničenjima meta-analize. Naše će izlaganje uključiti glavne ciljeve meta-analize te će biti prilično detaljno, ali ne i iscrpno, zbog činjenice da je područje meta-analize prilično opsežno te obuhvaća alternativne statističke tehnike o kojima ovdje nećemo raspravljati, kao što su Bayesovi i neparametrijski testovi.

## Introduction

Meta-analysis in its present form of statistically integrating information from several studies all with a common underlying theme has been around for over 30 years. The medical, industrial and basic science fields have seen many attempts by many researchers to pull summary data together from various sources within a discipline with the goal of making some definitive statement about the state of the science in that discipline. Likewise authors of manuscripts in the background and rationale section of their paper always summarize what they believe to be the state of affairs up to the time of the presentation of their own results in that particular paper. The new data and results they present in their current publication is an attempt to update the progress in that field. Thus in a sense they have performed a partial meta-analysis of summarizing information from the past, presenting their added contribution and thus updating the knowledge base. They have not quite integrated past data in a rigorous statistical way with their new data, but have merely used the data history to justify their current research which pretty much stands on its own.

Thus meta-analysis is an after the fact attempt to pull together the current knowledge base whether it be publications or raw data and present a statistical synthesis of all the information and reach a conclusion as to the best treatment or intervention strategy based on all these past contributions. As stated by Glass (1) "Meta-analysis refers to the analysis of analyses...the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings". In most disciplines the statistical analysis can be endpoints such as mean effects, response rates, odds ratios, correlations, etc. We want to present the statistical issues of meta-analysis by example. We cannot possibly cover all the endpoints in this manuscript. We will thus concentrate on the continuous endpoint of mean effects with the understanding that the statistical concepts are transparent to other continuous endpoints such as correlations and to categorical endpoints such as odds ratios, relative risks and response rates. We begin by discussing the sources of literature review for a meta-analysis. The remainder of the manuscript will address the statistical terminology and approaches to conducting a meta-analysis. We then present techniques for investigating publication bias and conclude with some thoughts on the advantages and limitations of meta-analyses. Our discussions involve the major objectives of meta-analysis and are certainly detailed but not exhaustive as the meta-analytic field is quite extensive involving alternative statistical techniques not discussed here such as Bayesian and non parametric procedures as the available literature attests.

## Sustavni pregledni članci

Prirodno je započeti pitanjem poput „Smanjuju li statini rizik od raka?“ ili „Je li koncentracija vitamina C u plazmi odgovarajući biološki biljeg za uzimanje vitamina C?“. Mora postojati način pretraživanja literature kako bi se odredilo mogu li se sažeti objavljeni rezultati i donijeti krajnji zaključak. To se zove „sustavni pregledni članak“, odnosno pregled koji jasno opisuje korištene metode za pronalženje, procjenjivanje te, gdje se to može, kombiniranje relevantnih istraživanja. Sustavni pregledni članak može imati nekoliko ključnih elemenata koji uključuju: 1) pripremu sustavnog pretraživanja ili formuliranja istraživačkog pitanja, 2) sustavno istraživanje primarne literature, 3) odabir radova za pregled i 4) kritična prosudba odabrane literature. Najčešći izvori pretraživanja literature su: MEDLINE (baza podataka američke medicinske knjižnice), EMBASE (medicinska i farmakološka baza podataka izdavačke kuće Elsevier Publishing), CINAHL (zbirni pokazatelj sestrinske i srodne zdravstvene literature, engl. *Cumulative Index to Nursing and Allied Health Literature*), CANCELIT (baza podataka za pretraživanje literature o raku) i Cochrane Collaborative (međunarodna neprofitna i nezavisna organizacija koja daje obnovljene i točne informacije o rezultatima/učincima zdravstvene skrbi u cijelom svijetu. Ona stvara i distribuira sustavne pregledne članke o intervencijama u zdravstvenoj skrbi te podržava potragu za podacima temeljenim na dokazima u obliku kliničkih i drugih istraživanja.)

Svi su ti izvori imali važnu ulogu u prikupljanju informacija kako bi istraživači na jednom mjestu imali dostupne rezultate raznih istraživanja. Meta-analiza jest zapravo statistička sinteza, bilo kvantitativna ili kvalitativna, rezultata istraživanja uključenih u sustavni pregledni članak i tumačenje tih podataka. U mnogim područjima istraživanja statističke završne točke mogu biti veličina učinka, stopa odgovora, omjeri izgleda, korelacije i ostalo. Stoga bi istraživač trebao biti svjestan odabira istraživanja, metode prikupljanja podataka, načina kako se podaci trebaju statistički sažeti te kako se izvodi zaključak. Metode za uključivanje istraživanja u meta-analizu mogu biti opsežne i dovoljno je reći da ćemo, budući da je svrha ovog pregleda obrađivanje statističkih elemenata, pretpostaviti da istraživanja obrađuju sličnu tematiku, da imaju dobro postavljene kriterije prihvaćanja te da su im tehnike procjene završnih točaka mjerenja sustavno postavljene u cijelom istraživanju. Jedan kriterij uključivanja u meta-analizu u kliničkom okruženju koji moramo imati na umu jest da istraživanja koja uspoređuju učinak novog liječenja, onog koje ispituje, i učinak starog liječenja moraju biti ispravno randomizirana.

## Veličina učinka

Veličina učinka (engl. *effect size*, ES) u jednom istraživanju je mjera razlike između učinaka dvaju liječenja. Uzmimo

## Systematic Reviews

One naturally starts out with a question such as “Do statins reduce the risk of cancer?” or “Is plasma vitamin C an appropriate biomarker for vitamin C intake?” There has to be a resource to search the literature to determine if one can synthesize the published or reported results and reach a definitive conclusion. This is called the “systematic review” or a review that clearly describes the methods used to locate, appraise and, where appropriate, combine relevant studies. The systematic review can have several key elements which include: 1) preparation for systematic reviewing or formulating the research question, 2) systematic research of the primary literature, 3) selection of papers for review and, 4) critical appraisal of the selected literature. The most common literature search sources are MEDLINE (United States Library of Medicine database), EMBASE (medical and pharmacologic database by Elsevier publishing), CINAHL (cumulative index to nursing and allied health literature), CANCELIT (cancer literature research database) and the Cochrane Collaborative (an international nonprofit and independent organization which provides up-to-date, accurate information about the results/effects of healthcare worldwide. It produces and disseminates systematic reviews of healthcare interventions and in addition supports the search for evidence based information in the form of clinical trials and other studies of interventions.)

Thus all, these sources have played an important part in bringing information together so that researchers can be exposed to the availability of results from research directed at a common endpoint. The statistical synthesis, either quantitatively or qualitatively, of the results of the studies included in the systematic review and interpretation of the data is the actual meta-analysis. In most areas of research the statistical endpoints can be mean effects, response rates, odds ratios, correlations, etc. Thus, hopefully, one is aware of the selection of studies, the methods of data collection and how the results were statistically summarized and the conclusions derived. The methods for including studies in a meta-analysis can be extensive and suffice it to say, since our purpose is to address the statistical issues we will assume that the studies are addressing a common question, have well established eligibility criteria and the evaluation techniques for the endpoints are systematically established across the studies. One inclusion criteria in the meta-analysis in the clinical setting to be aware of is that the studies making comparisons of a treatment of interest to a control are properly randomized.

## Effect sizes

An effect size in a single study is the measure of the difference of two treatment effects. For example in a single study where one wants to compare the mean reduction

primjer istraživanja gdje se žele usporediti srednje vrijednosti smanjenja krvnog tlaka zbog novog antihiperzivnog liječenja, A i starog liječenja, B. ES iznosi:

$$ES = c(m) [(M_A - M_B) / S],$$

gdje je  $M_A$  srednja vrijednost sniženja krvnog tlaka kod liječenja A,  $M_B$  je srednja vrijednost sniženja krvnog tlaka kod liječenja B, a  $S$  je združena standardna devijacija (engl. *pooled standard deviation*) obaju uzoraka. Združena standardna devijacija definira se kao

$$S = \{ [(n_A - 1)(s_A)^2 + (n_B - 1)(s_B)^2] / (n_A + n_B - 2) \}^{1/2},$$

gdje  $s_A$  i  $s_B$  predstavljaju standardne devijacije (2) iz uzoraka A, odnosno B, a  $n_A$  i  $n_B$  predstavljaju veličine uzoraka A, odnosno B. Vrijednost  $c(m)$  je korekcijski faktor (3) za sustavnu pogrešku (engl. *bias*) u određivanju stvarnog ES zbog malog uzorka. Definira se kao

$$c(m) = 1 - 3 / (4m - 1),$$

gdje je  $m = n_A + n_B - 2$ . Može se vidjeti da se s povećanjem veličine uzorka,  $c(m)$  primiče vrijednosti 1. Varijanca istraživanja ES se ponekad određuje kao (3)

$$\text{var}(ES) = [(n_A + n_B) / n_A n_B] + (ES)^2 / (2T - 3.94),$$

gdje je  $T = n_A + n_B$ . Približna konstanta 3,94 je korekcijski faktor (engl. *correction factor*) (3) u ovakvoj vrsti izračuna. Pretpostavimo sada da imamo mnoga istraživanja,  $k > 6$ , koja uspoređuju neki oblik liječenja A sa starim liječenjem. Ovdje želimo napomenuti da se u meta-analizi ne pretpostavlja da je liječenje A u identičnom obliku u svim studijama. Primjerice, u jednom istraživanju to može biti liječenje A vježbanjem, u drugom je liječenje A s vitaminskim dodatkom prehrani, u trećem je samo liječenje A itd. Kontrolno liječenje u svakom istraživanju također ne mora biti identično, dokle god se ne radi o liječenju A. Odnosno, kontrolno liječenje u jednom istraživanju može biti placebo, u drugom to može biti samo vježbanje, u trećem jednostavno izostanak bilo kakve intervencije (to se može napraviti ukoliko je etično) i slično. Poanta je da se u meta-analizi postavlja pitanje pokazuje li liječenje A u bilo kojem obliku prednosti kod snižavanja krvnog tlaka u usporedbi s bilo kojim liječenjem koje nije A. Uz pretpostavku da imamo istraživanja  $k$  koja želimo sumirati, meta-analiza kombinira veličinu učinka  $k$   $ES_1, ES_2, \dots, ES_k$  te daje srednju vrijednost veličine učinka  $ES_{\text{mean}}$

$$ES_{\text{mean}} = \sum_{i=1, k} w_i ES_i / \sum_{i=1, k} w_i \text{ za } i=1, \dots, k.$$

Vrijednosti  $w_i$  za  $i = 1, \dots, k$  predstavljaju težinu istraživanja. U meta-analizi nisu sva istraživanja jednake težine.

in blood pressure of a new anti hypertensive treatment, A, and the control treatment, B. The sample effect size (ES) is

$$ES = c(m) [(M_A - M_B) / S],$$

where  $M_A$  is the mean reduction in blood pressure on treatment, A, and  $M_B$  is the mean reduction in blood pressure on treatment, B and  $S$  is the pooled standard deviation from the both samples. The pooled standard deviation is defined as

$$S = \{ [(n_A - 1)(s_A)^2 + (n_B - 1)(s_B)^2] / (n_A + n_B - 2) \}^{1/2},$$

where  $s_A$  and  $s_B$  are the standard deviations (2) from samples A and B, respectively, and  $n_A$  and  $n_B$  are the sample sizes for A and B, respectively. The value  $c(m)$  is a correction factor (3) for the bias in estimating the true ES due to small sample sizes. It is defined as

$$c(m) = 1 - 3 / (4m - 1),$$

where  $m = n_A + n_B - 2$ . One can see that as the sample size increases  $c(m)$  converges to the value 1. The variance of the study ES is sometimes estimated as (3)

$$\text{var}(ES) = [(n_A + n_B) / n_A n_B] + (ES)^2 / (2T - 3.94),$$

where  $T = n_A + n_B$ . The approximate constant 3.94 is a correction factor (3) in this type of calculation.

Now let us assume we have many studies,  $k > 6$ , comparing some form of treatment A with a control. Please note that in meta-analyses we are not assuming that the treatment A is in the same form across all studies. For example in one study it may be A with exercise and in another A with a vitamin supplement, in another just A alone etc. The control in each may not be the same as long it is not A. That is to say, the control in one study may be a placebo tablet, in another it may be exercise alone, in another no intervention at all (if ethically this can be done), etc. The point being, in a meta-analysis you are asking the question whether receiving A in any form shows superiority in reducing blood pressure when it is compared to something other than A. Thus assuming we have  $k$  studies that we want to synthesize, a meta-analysis combines the  $k$  effect sizes,  $ES_1, ES_2, \dots, ES_k$  resulting in a mean effect size,  $ES_{\text{mean}}$

$$ES_{\text{mean}} = \sum_{i=1, k} w_i ES_i / \sum_{i=1, k} w_i \text{ for } i=1, \dots, k.$$

Note the values,  $w_i$  for  $i = 1, \dots, k$ . These are weights. In a meta-analysis, not all studies should be weighted

Primjerice, istraživanje s 500 ispitanika treba se vrednovati više od istraživanja s 80 ispitanika zbog više informacija koje nudi. Težina se može odrediti i na drugi način, kao što je opseg (duljina) praćenja u uzdužnim longitudinalnim istraživanjima. Istraživanje u kojem je završna točka ovisna o vremenu kao, primjerice, u slučaju hoće li ispitanik preživjeti ili gdje se ispitanici prate 10 godina, treba vrednovati više nego ono u kojem se ispitanici prate samo 5 godina. Stvarni izračuni, kojima se određuje težina istraživanja, mogu biti prilično složeni. Uobičajena tehnika jest postaviti težinu kao inverznu varijancu veličine učinka (3) koji smo već definirali, tj.,

$$w_i = 1/\text{varijanca}(ES_i), i=1, \dots, k.$$

Jedna je teza da će veća istraživanja imati višu preciznost i time manju varijaciju te sukladno tome i veću težinu. Drugi način za jednostavan izračun težine bio bi omjer ispitanika u određenom istraživanju i svih ispitanika koji su uključeni u meta-analizu. Primjerice, ako je  $N$  ukupan broj ispitanika svih istraživanja meta-analize, a  $N_i$  predstavlja ukupan broj ispitanika istraživanja  $i$ ,  $i = 1, \dots, k$ , onda je  $N = N_1 + N_2 + \dots + N_k$ . Tada definiramo  $w_i = N_i / N$  za  $i = 1, \dots, k$ , kao težinu istraživanja  $i$ . Jedna od pretpostavki koja se često postavlja u meta-analizi jest da se varijacija između istraživanja ne razlikuje. Stoga težine definirane gornjom formulom ne vrijede za varijaciju između istraživanja (4). To se naziva meta-analiza prema modelu fiksnih učinaka. Kako bi izračunali slučajnost između istraživanja ili model slučajnog učinka koristimo

$$w_i = 1 / [\text{varijanca}(ES_i + \tau^2)], i=1, \dots, k,$$

gdje  $\tau^2$  predstavlja varijaciju unutar istraživanja. Izračun je kompliciran te ga ovdje nije nužno prikazati. Može se primijetiti da se svi naši izračuni mogu jednostavno rabiti, a većina njih se radi rutinski uz statistički program. Dakle, sada imamo sve elemente za testiranje ukupnog učinka novog liječenja A u usporedbi sa starim liječenjem B za kombinirana istraživanja  $k$ . Postavimo  $\delta$  kao broj ispitanika kombiniranog  $ES_{\text{mean}}$ . Nulta hipoteza glasi da ne postoji razlika između liječenja A i liječenja B u kombiniranim istraživanjima ili  $H_0: \delta = 0$  nasuprot alternativnoj hipotezi,  $H_A: \delta \neq 0$ . 95%-ni interval pouzdanosti (5) na srednju vrijednost broja ispitanika iznosi:

$$[ES_{\text{mean}} - 1,96\tau, ES_{\text{mean}} + 1,96\tau].$$

Ako taj interval pokriva i 0 tada nema statistički značajnog učinka u korist liječenja A.

Kako bi zornije prikazali prethodno opisan koncept pogledajmo sljedeći primjer. Pretpostavimo da imamo devet istraživanja s veličinom učinka iz usporedbe liječenja A i

equally. For example, a study containing 500 subjects should be weighted more than a study containing 80 subjects due to the added information. Weights can be determined in other ways such as the amount of follow up in longitudinal studies. A study in which the endpoint is time dependent such as survival and where subjects are followed for 10 years should be weighted more than a study in which subjects are only followed for 5 years. The actual calculations for determining weights can be quite complicated. One common technique is to allow the weight to be the inverse variance of the effect size (3) which we defined above i.e.,

$$w_i = 1/\text{variance}(ES_i), i=1, \dots, k.$$

One rationale being that larger studies will have more precision and thus smaller variation and therefore greater weight. Another alternative choice for the simple weight would be the proportion of subjects in that study that contribute to the meta-analysis. For example, if  $N$  is the total number of subjects from all the studies contributing to the meta-analysis and  $N_i$  is the total number in study  $i$ ,  $i = 1, \dots, k$ , then  $N = N_1 + N_2 + \dots + N_k$ . We then define  $w_i = N_i / N$  for  $i = 1, \dots, k$ , as the weight of the  $i$ th study. One of the assumptions often made in a meta-analysis is that the between studies variation do not differ. Thus the weights defined above do not account for the between studies variation (4). This is called a fixed effects meta-analysis. To allow for the randomness between the studies or a random effects model we let

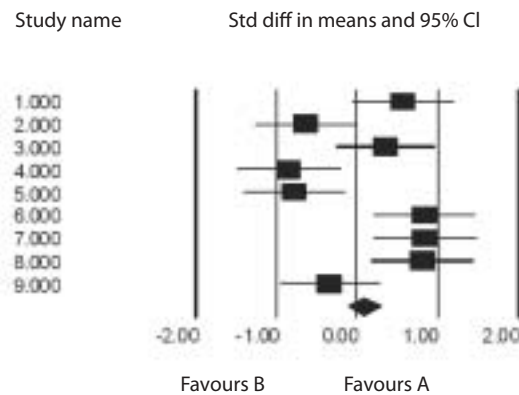
$$w_i = 1 / [\text{variance}(ES_i + \tau^2)], i=1, \dots, k,$$

where  $\tau^2$  is the among study variation. The computation is complex and need not be shown to make our point. One can see that all our calculations can be quite involved computationally and most calculations are routinely done using statistical software. However, we now have all the elements to test the overall effect of the new treatment, A, versus the control, B, for the combined  $k$  studies. Suppose that  $\delta$  is the population value of the combined  $ES_{\text{mean}}$ . The null hypothesis is that there is no difference between A and B in the combined analysis or  $H_0: \delta = 0$  versus the alternative hypothesis,  $H_A: \delta \neq 0$ . A 95% confidence interval (5) on the population mean is:

$$[ES_{\text{mean}} - 1,96\tau, ES_{\text{mean}} + 1,96\tau].$$

If this interval covers 0 then there is no significant effect due to A.

To demonstrate the concepts described above, let us consider the following example. Suppose we have nine studies with effect sizes comparing treatments A and B;



**SLIKA 1.** Trakasti grafikon veličine učinka (ES) s 95%-nim intervalom pouzdanosti.

**FIGURE 1.** Forest plot of ES with their 95% confidence intervals.

B: 0,574, -0,642, 0,363, -0,840, -0,783, 0,833, 0,842, 0,803, -0,345. Kao što se može vidjeti pet je brojeva pozitivnih u korist liječenja A i četiri su negativna u korist liječenja B.  $ES_{\text{mean}} = 0,086$ ,  $t = 0,109$  za model fiksnih učinaka i  $\tau = 0,243$  za model slučajnih učinaka; 95%-ni interval pouzdanosti (95% CI) za fiksne učinke je  $CI = [(-0,127)-0,299]$ , a za slučajne učinke  $CI = [(-0,387)-0,563]$ . U oba slučaja interval pokriva nulu, što znači da iz ovih devet istraživanja proizlazi kako liječenje A nema učinka. Također se može primijetiti da je CI za slučajne učinke širi od CI za fiksne učinke, budući da uzima u obzir dodanu varijaciju između istraživanja koju smo pripisali heterogenosti ili inherentnim razlikama između istraživanja. Jasno je da je to zbog toga što je oko polovina istraživanja pokazala uspješnost liječenja A, a druga su zaključila da je liječenje B uspješnije.

Na slici 1. prikazan je trakasti grafikon (engl. *forest plot*) koji prikazuje veličine učinaka i njihove 95%-ne intervale pouzdanosti. Ukupna veličina učinka prikazana je u obliku romba na dnu grafikona.

## Rasprava o heterogenosti

Sljedeći je zadatak odrediti postoji li statistički značajna heterogenost među istraživanjima. Statistička metoda koju ćemo pritom rabiti zove se Q-statistika (engl. *Q statistic*) kojom se određuje stupanj heterogenosti. Ovdje to nismo izračunali. Dovoljno je reći sljedeće: ako je Q blizu nule, nema heterogenosti. Q ima hi-kvadratnu raspodjelu (engl. *chi square distribution*) (6) na  $k - 1$  stupnjeva slobode (engl. *degree of freedom*, df) ili broj istraživanja minus je-

0.574, -0.642, 0.363, -0.840, -0.783, 0.833, 0.842, 0.803, -0.345. As one can see, 5 are positive, favoring A and 4 are negative favoring B. The  $ES_{\text{mean}} = 0.086$ ,  $t = 0.109$  for the fixed effects model and  $\tau = 0.243$  for the random effects model. The fixed effects 95% confidence interval (CI) =  $(-0.127-0.299)$  and the random effects 95% CI =  $(-0.387-0.563)$ . In both cases the interval covers 0 and there is no benefit to A from the 9 studies. One can also see that the random effects CI is wider than the fixed effects CI as it takes into account the added variation between the studies which we attribute to heterogeneity or some inherent differences among the studies. Clearly this is so as about half of the studies showed some superiority due to treatment A and the others showed an advantage due to treatment B.

Figure 1 is a Forest plot which graphs the effects sizes and their individual 95% confidence intervals. The resulting overall effect size is demonstrated by the means diamond at the base of the plot.

## Discussion of heterogeneity

The next task is to determine if there is significant heterogeneity among the studies. The statistic that we invoke is called the Q statistic and is used to determine the degree of heterogeneity. We do not compute it here. Suffice it to say, if Q is close to 0 then there is no heterogeneity. The Q has a chi square distribution (6) on  $k-1$  degrees of freedom (df) or the number of studies minus one. Like other statistics we calculate Q and compare it to a critical value for the Chi square distribution. The null hypothesis is that

dan ( $k - 1$ ). Kao kod drugih statističkih testova, izračunali smo  $Q$  i usporedili ga s kritičnom vrijednosti za hi-kvadratnu raspodjelu. Nulta hipoteza glasi da u kombiniranoj analizi nema heterogenosti ili  $H_0: Q = 0$ ; nasuprot alternativnoj hipotezi,  $H_A: Q \neq 0$  koja pretpostavlja da postoji statistički značajna heterogenost. Kritična vrijednost devet istraživanja za hi-kvadratnu raspodjelu od  $9 - 1 = 8$  df na razini značajnosti od 0,05 iznosi 16,919. Izračunata vrijednost  $Q$  temeljem naših podataka je 38,897. Vidljivo je da je izračunata vrijednost  $Q$  više nego dvostruko mnogo viša od kritične vrijednosti. Stoga odbacujemo nultu hipotezu i određujemo da postoji statistički značajna razlika među našim istraživanjima. To ne mijenja naš zaključak da nema statistički značajnog učinka za liječenje A. Međutim, to nas upozorava na nužnost određivanja izvora heterogenosti. Neki mogući izvori heterogenosti su kliničke razlike između ispitivanja kao što su odabir bolesnika, osnovni stupanj jačine bolesti, obrada neželjenih ishoda (toksičnost), osobine bolesnika itd. Također se mogu istražiti i metodološke razlike između istraživanja, kao što su mehanizmi randomizacije, udio ispitanika koji su odustali/bili isključeni tijekom praćenja (engl. *extent of withdrawals and lost to follow up*). U nekim slučajevima heterogenost može biti rezultat slučajnosti ili uzroka koji se ne mogu lagano otkriti. Izračun heterogenosti može biti jedan od najproblematičnijih aspekata meta-analize. To je važan aspekt, budući da može utjecati na donošenje odluke o tome koji će se statistički model odabrati (fiksni ili slučajni učinci). Ako se pronađe statistički značajna heterogenost, onda se mogu pronaći i potencijalne moderatorske varijable kako bi se objasnila ta varijabilnost. To će možda zahtijevati da se meta-analiza koncentrira na podskupine istraživanja koja su sama po sebi homogena. Slijedimo ideju homogenosti korak dalje. Ispitajmo indeks  $I^2$  (6) kojim računamo opseg heterogenosti iz zbirke veličina učinka.

Indeks  $I^2$  se jednostavno može tumačiti kao postotak heterogenosti u sustavu ili općenito kao iznos ukupnih varijacija koje su se pojavile zbog varijance među istraživanjima,  $\tau^2$ . U našem primjeru je  $I^2 = 79,433\%$  ili  $79,4\%$  varijance ili, drugim riječima, heterogenost se pojavila zbog  $\tau^2$  - varijance između istraživanja.

### Sustavna greška objavljivanja

Ljevkasti grafikoni (engl. *funnel plot*), grafikoni s procjenama učinka istraživanja na osi x i uzlazno sortiranom veličinom uzorka na osi y, mogu biti korisni za procjenu valjanosti meta-analize. To se često naziva sustavna greška objavljivanja (engl. *publication bias*). Ljevkasti grafikon temeljen je na činjenici da će s porastom veličine uzorka porasti i preciznost određivanja učinka ispitivanog liječenja. Rezultati manjih istraživanja raspršiti će se na dnu grafikona te će se ta širina sužavati prema vrhu grafikona gdje se nalaze veća istraživanja. Kada ne postoji sustavna greška

there is no heterogeneity in the combined analysis or  $H_0: Q=0$  versus the alternative hypothesis,  $H_A: Q \neq 0$  implying there is significant heterogeneity. The critical value of the 9 studies for a chi square distribution on  $9-1=8$  df at an alpha level of 0.05 is 16.919. The calculated value of  $Q$  from our data is 38.897. Clearly the calculated  $Q$  is much greater than the critical value – over twice as much. Therefore we reject the null hypothesis and determine there is significant heterogeneity among our studies. This does not change our conclusion that there is no significant effect due to A. However, it does alert us that we should try to determine the source of the heterogeneity. Some possible sources of heterogeneity may be the clinical differences among the studies such as patient selection, baseline disease severity, management of inter current outcomes (toxicity), patient characteristics or others. One may also investigate the methodological differences between the studies such as mechanism of randomization, extent of withdrawals and lost to follow up or heterogeneity still can be due to chance alone and the source not easily detected. Quantifying heterogeneity can be one of the most troublesome aspects of meta-analysis. It is important because it can affect the decision about the statistical model to be selected (fixed or random effects). If significant heterogeneity is found then potential moderator variables can be found to explain this variability. It may require concentrating the meta-analysis on a subset of studies that are homogeneous within themselves. Let's pursue this idea of heterogeneity one step further. Let's examine the  $I^2$  index (6) which quantifies the extent of heterogeneity from a collection of effect sizes.

The  $I^2$  index can easily be interpreted as the percentage of heterogeneity in the system or basically the amount of the total variation accounted for by the between studies variance,  $\tau^2$ . In our example  $I^2 = 79.433\%$  or  $79.4\%$  of the variance or heterogeneity is due to  $\tau^2$  or the between studies variance.

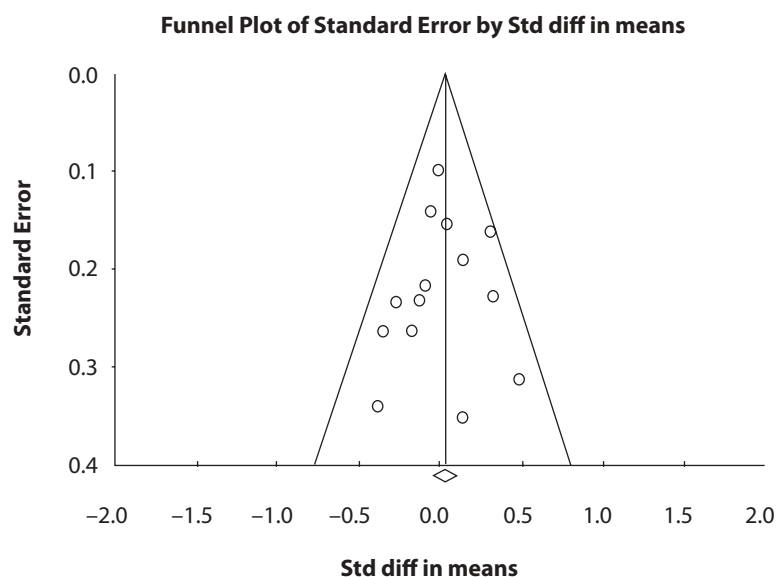
### Publication bias

Funnel plots, plots of the studies' effect estimates on the horizontal axis against sample size in ascending order on the vertical axis, may be useful to assess the validity of meta-analyses. This is often referred to as "publication bias". The funnel plot is based on the fact that precision in estimating the underlying treatment effect will increase as the sample size of the included studies increases. Results from small studies will scatter widely at the bottom of the graph, with the spread narrowing among larger studies to the top of the graph. In the absence of bias the plot will resemble a symmetrical inverted funnel. Conversely, if there is bias, funnel plots will often be skewed and asymmetrical. Funnel plots are usually constructed as plotting the ES vs. the sample size or the ES vs. the SE.

objavljivanja, grafikon slično obrnutom ljevku. Ljevka se grafovi obično rade kada se grafički prikazuje ES nasuprot veličini uzorka ili ES nasuprot standardnoj pogrešci (engl. *standard error, SE*).

U analizi kontinuiranih podataka možemo rabiti pristup linearne regresije (8) za mjerenje asimetrije ljevkastog grafikona kod ES. ES se izvodi iz preciznosti procjene koja se određuje kao inverzija standardne pogreške (jednažba regresije:  $ES = a + b \times \text{preciznost}$ , gdje je  $a$  odsječak, a  $b$  nagib). Budući da preciznost u velikoj mjeri ovisi o veličini uzorka, manja istraživanja će na osi x biti blizu nuli i obrnuto za veća istraživanja. Svrha je ne odbaciti nultu hipotezu da pravac prolazi kroz ishodište, tj. da je odsječak = 0. Za naših devet istraživanja ljevka grafikon (nije prikazan) je simetričan, no plosnat jer je bilo premalo istraživanja čiji se rezultati mogu dobro prikazati na grafikonu. Međutim, simetrija je očuvana kao što se vidi iz rezultata regresije. Ispitivanje nulte hipoteze da odsječak = 0 nije statistički značajno na razini značajnosti 0,05, te ukazuje da pravac prolazi kroz ishodište. Stvarna vrijednost P je 0,7747 što znači da postoje jaki dokazi u podacima koji ne podržavaju odbacivanje nulte hipoteze da je odsječak = 0. Stoga ne sustavna greška objavljivanja. Slika 2 prikazuje ljevka grafikon s 14 istraživanja koja uspoređuju dvije hipertenzivne terapije. Na osi x nalaze se ES standardne razlike (engl. *standard difference*) izražene u srednjim vrijednostima kao što je prethodno opisano, a na osi y je standardna pogreška (engl. *standard error, SE*).

In continuous data analysis we can use a linear regression approach (8) to measure funnel plot asymmetry on the ES. The ES is regressed against the estimate's precision, the latter being defined as the inverse of the standard error (regression equation:  $ES = a + b \times \text{precision}$ , where  $a$  is the intercept and  $b$  is the slope). Since precision depends largely on sample size, small trials will be close to zero on the horizontal axis and vice versa for larger trials. The purpose is to not reject the null hypothesis that the line goes through the origin, i.e. intercept = 0. For our nine studies the funnel plot (not shown here) is symmetrical but rather flat as there are too few studies to result in a good plot. However, the symmetry is preserved as shown by the regression results. The test of the null hypothesis that the intercept equal to zero is not significant at the 0.05 alpha level indicates that the line goes through the origin. The actual p-value is 0.7747 indicating that there is strong evidence in the data to not support rejection of the null hypothesis of zero intercept. Thus there is no publication bias. Figure 2 is a funnel plot of 14 studies comparing two anti hypertensive therapies. The horizontal axis is the ES of the standard difference in means that we described above and the vertical axis is the standard error. One can see the symmetry. The studies with larger sample size or most precision are the dots or points located at the top of the graph and the studies with decreasing sample size or less precision (larger standard error) scatter towards the bottom of the plot. The diamond at the bottom of



**SLIKA 2.** Ljevka grafikon od 14 istraživanja s veličinom učinka (engl. *Std diff in means*) na osi x te standardnom pogreškom na osi y.

**FIGURE 2.** Funnel Plot of 14 studies with the effect size (*Std diff in means*) on the horizontal and the standard error on the vertical.



Simetrija se može lako vidjeti. Istraživanja s većim uzorkom ili višom preciznošću prikazana su kao točke pri vrhu grafikona, a istraživanja s manjim uzorkom ili manjom preciznošću (višom standardnom pogreškom) raspršena su prema dnu grafikona. Oblik romba na dnu grafa predstavlja srednju vrijednost učinka cjelokupne meta-analize u usporedbi dvaju liječenja i može se vidjeti da je blizu nule. Statističkim testom za odsječak = 0 ili hipotezom da pravac prolazi ishodištem dobivena je  $P = 0,730$ . Pravac se vidi kao središnja okomita linija grafikona. Dvije linije pod kutom ocrtavaju oblik točaka i možemo vidjeti oblik lijevka na grafikonu. Detaljnije rasprave o ljevkastom grafikonu i testu simetrije mogu se pronaći u literaturi (8,9) koja se bavi meta-analizama.

## Zaključak

Pokušali smo dati pregled bitnih statističkih pristupa meta-analizi. Prednosti meta-analize nalazimo u mogućnosti statističkog obrađivanja proturječnih pitanja koja nisu jasno obrađena u pojedinim istraživanjima. Objavljivanjem meta-analiza često se u literaturi predstavlja mogućnost integriranja rezultata iz različitih izvora i obrađivanje relevantnih pitanja. Sigurno bi svatko htio iskoristiti prednosti obilja objavljenih informacija te ih sažeti na racionalan način. Kad pojedinac poželi integrirati informacije iz različitih publikacija i baza podataka, zasigurno će doći do proturječnosti (10). Meta-analiza je jako podložna takvim proturječnostima. Međutim, trebaju se poduzeti sve mjere kako se sustavne pogreške ne bi pojavljivale, a to se može postići istraživanjem tema poput heterogenosti i sustavne greške objavljanja ili da se te pogreške barem izlože i objasne. Ograničenja meta-analize očigledno proizlaze iz odabira istraživanja, izbora relevantnih ishoda, metoda analize, tumačenja heterogenosti te generalizacije i primjene rezultata. Statističke metode koje su nam pri ruci zasigurno su odgovarajuće za obradu tih tema. Postoje kvalitetni internetski izvori koji nude smjernice za provođenje meta-analize. To su CONDORT, QUORUM i MOOSE. Njih se može upisati kao ključne riječi u pretraživanju pri traženju izvora. The Cochrane Collaborative, baza koju smo već spomenuli, nudi izvanredne smjernice. Međutim, treba imati na umu da meta-analiza nije zamjena za dobro dizajnirana opsežna randomizirana istraživanja (engl. *large scale randomized studies*) niti opravdanje za provođenje malih istraživanja slabe snage (engl. *small underpowered studies*). Ona je sredstvo koje uz primjerenu upotrebu pomaže u donošenju razumne i održive odluke na temelju postojećih znanstvenih informacija.

the plot is the means diamond for the overall meta-analysis comparing the two treatments and one sees it is close to zero. The statistical test for the intercept = 0 or the line goes through the origin is  $P = 0.730$ . The line is seen as the middle vertical line on the graph. The two angled lines give an outline of the pattern of points and one can see the funnel shape of the plot. Detailed discussions of funnel plots and test of symmetry can be found throughout the literature (8,9) pertaining to meta-analyses.

## Conclusion

We have attempted here to give an overview of the major statistical approaches to meta-analysis. Advantages of meta-analyses are seen from the ability to address a controversial issue statistically that may not have been conclusively addressed in single studies. The ability to integrate results from diverse sources and address the relevant issues have been demonstrated many times in the literature through the publication of meta-analyses. One would want to certainly take advantage of the wealth of information published and have a rational way of synthesizing that information. There are certainly controversies (10) involved when one attempts to integrate information from various published works or different databases. Meta-analysis is prone to that controversy. However, every attempt is made to guard against bias through examination of topics such as heterogeneity and publication bias and to at least expose them and explain them. Limitations obviously result from selection of studies, choice of relevant outcome, methods of analysis, interpretation of heterogeneity and generalization and application of results. The statistical tools at hand are certainly adequate for addressing these issues. There are good online sources which provide guidelines for conducting a meta-analysis. These include CONDORT, QUORUM and MOOSE. These can be entered as key words to locate the resource. The Cochrane Collaborative mentioned above provides excellent guidelines as well. However, one should keep in mind that meta-analyses should not be a replacement for well designed large scale randomized studies (10) nor a justification for conducting small underpowered studies. It is a tool when properly utilized helps one to arrive at a reasonable and defensible decision from the scientific information already presented.

**Adresa za dopisivanje:**

Alfred A. Bartolucci  
Department of Biostatistics  
University of Alabama at Birmingham  
1665 University Boulevard  
Birmingham, Alabama 35294-0022  
USA  
e-pošta: Albartol@uab.edu

**Corresponding author:**

Alfred A. Bartolucci  
Department of Biostatistics  
University of Alabama at Birmingham  
1665 University Boulevard  
Birmingham, Alabama 35294-0022  
USA  
e-mail: Albartol@uab.edu

**Literatura/References:**

1. Glass GV. Primary, secondary, and meta-analysis of research. *Educational Researcher* 1976;5:3-8.
2. McHugh ML. Standard error: meaning and interpretation. *Biochemia Medica* 2008;18:7-13.
3. Hedges LV, Olkin O. *Statistical methods for meta-analysis*. San Diego (CA): Academic Press; 1985.
4. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;7:177-88.
5. Simundic AM. Confidence interval. *Biochemia Medica* 2008;18(2): 154-61.
6. Triola MM, Triola MT. *Biostatistics for the Biological and Health Sciences*. Boston (MA): Pearson Education, Inc.;2006.
7. Higgins JPT, Thompson SG. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* 2002;21:1539-58.
8. Egger M, Davey G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *British Medical Journal* 1997;315:629-34.
9. Bartolucci AA, Howard G. Meta-analysis of data from the six primary prevention trials of cardiovascular events using aspirin. *American Journal of Cardiology* 2006;10:746-50.
10. Bartolucci AA. The significance of clinical trials and the role of meta-analysis. *Journal of Surgical Oncology* 1999;72(3):121-3.