Advances in Meta-Analysis: Examples from Internal Medicine to Neurology

Massimiliano Copetti^a Andrea Fontana^a Giusi Graziano^{b, d} Federica Veneziani^e
Federica Siena^f Marco Scardapane^c Giuseppe Lucisano^c Fabio Pellegrini^{a, c}

^aUnit of Biostatistics, Casa Sollievo della Sofferenza, San Giovanni Rotondo, ^bLaboratory of Lipid Metabolism and Cancer, DTP, and ^cUnit of Biostatistics, DCPE, Consorzio Mario Negri Sud, Santa Maria Imbaro, ^dNational Cancer Research Center, IRCCS Oncologico Giovanni Paolo II, ^eNeurodegenerative Disease Unit, Department of Neurosciences and Sense Organs, University of Bari, Bari, and ^fDepartment of Clinical and Research Neurology, Azienda Ospedaliera Pia Fondazione Panico, Tricase, Italy

Key Words

Meta-analysis · Data pooling · Advanced techniques · Generalized linear mixed models · Exact methods · Bayesian approaches

Abstract

Objective: We review the state of the art in meta-analysis and data pooling following the evolution of the statistical models employed. **Methods:** Starting from a classic definition of meta-analysis of published data, a set of apparent antinomies which characterized the development of the meta-analytic tools are reconciled in dichotomies where the second term represents a possible generalization of the first one. Particular attention is given to the generalized linear mixed models as an overall framework for meta-analysis. Bayesian meta-analysis is discussed as a further possibility of generalization for sensitivity analysis and the use of priors as a data augmentation approach. **Results:** We provide relevant examples to underline how the need for adequate methods to solve practical issues in specific areas of research have guided the development of advanced methods in meta-analysis. **Conclusions:** We show how all the advances in meta-analysis naturally merge into the unified framework of generalized linear mixed models and reconcile apparently conflicting approaches. All these complex models can be easily implemented with the standard commercial software available.

© 2013 S. Karger AG, Basel

Background

Quite a long time has passed since the methodological debate on uses and misuses of meta-analysis was dominating the scene of research [1–4]. Either as a statistical tool for combining evidence or as a methodological framework, meta-analysis is now widely accepted. This development, which occurred over the last decades, has been impressive [5] (fig. 1) and primarily driven by the needs of specific fields of research. We provide here a thorough review of advances in the methodology of meta-analysis, particularly stressing the clinical and research needs which somehow determined them.

Biostatistics, in its essence as a ‘discourse on the method’, is always challenged by specific problems [6], and its development often mirrors the necessity to solve practical issues in science. It is broadly accepted that the first attempt of meta-analysis goes back to 1904 when Karl Pearson was concerned with the pooling of existing evidence in terms of a set of correlation coefficients obtained from independent experiments [7].

Meta-analysis is a statistical technique to combine evidence of different findings obtained by similar experiments conducted on the same topic. When referring to a finding, we usually intend a treatment or an intervention effect on a given outcome. Logically this can be generalized to any form of exposure: risk factors, alleles, geno-

Fig. 1. Graph showing the estimated yearly relative increase of meta-analyses published in neurology for the years 1992–2012 according to Ovid MEDLINE. The year 1992 is set as reference year equal to 1. (Based on a highly comprehensive and low specific strategy search.)

types, etc. For the sake of simplicity the term ‘treatment effect’ or simply ‘effect’ will be used throughout this text.

The standard definition of meta-analysis reported above is clear and sound. Apparently it does not seem to possess many nuances and contrasts. Behind this concept a straightforward weighted mean is hidden, in which the weight for each study is formally represented by the inverse of the within-study variance (i.e. study precision). The process of forming an idea by giving each piece of evidence a different weight is something very intuitive. Nonetheless, a set of antinomies or dichotomies has always been at the core of the meaning and the possibilities of meta-analysis. The main purpose of this article is to show how apparent antinomies have resolved to clear dichotomies which can be easily handled, and the formalization of which has simply broadened the scope of meta-analysis.

Dichotomy 1: Fixed- versus Random-Effects Meta-Analysis

Can we always combine our available evidence? The answer to this concern has so far represented the very first dichotomy, which was based on the presence or absence of heterogeneity of effects, that is, whether the evidence to be combined belonged or not to the same underlying population. Cochran’s Q test provides a formal test for

heterogeneity assessment [8], although criticism about its low power to detect presence of heterogeneity makes a straightforward and intuitive measure (I^2) of the degree of heterogeneity preferable [9]. In the presence of heterogeneity, evidence can be combined using the inverse of within-study plus between-study variance as weights, i.e. using a random-effects model. This model yields an estimate of heterogeneity (namely, between-study variance) which is included in the process of weighting. The introduction of linear mixed models along with a seminal paper by DerSimonian and Laird [10] provided the theoretical background to this approach [11]. At the beginning, the purpose of resorting to random-effects models mainly consisted in relaxing the assumption that all the combined evidence stemmed from the same underlying population. Thereafter, focus was directed to a different concern: to explain heterogeneity by subgroup analysis and meta-regression, which simply consisted in exploring the sources of heterogeneity itself. Meta-regression, in particular, corresponds to the classical regression analysis though at study-level, through which we try to explain the between-study variability (say, the heterogeneity of treatment effect) according to study characteristics.

Therefore, fixed-effects meta-analysis is merely a particular case of random-effects meta-analysis in which between-study heterogeneity is assumed to be zero. However, it is now accepted that when heterogeneity exists, the concern should not be to obtain an overall combined

estimate of the effect, which might be in most cases clinically meaningless, but to explore the potential sources of heterogeneity itself [12]. To this purpose, contrary to some groundless common practices, there is no need to perform a meta-regression when no heterogeneity is present, and, on the other hand, researchers performing meta-regression to investigate sources of heterogeneity should protect themselves against the risk of false-positive findings by using appropriate statistical testing [13].

Dichotomy 2: Randomized Controlled Clinical Trials versus Observational Studies Meta-Analysis

Are study designs equally valuable for pooling in meta-analysis? Is there any difference in combining evidence from randomized controlled clinical trials (RCTs) versus evidence from observational studies such as case-control studies?

The spread of meta-analysis started with the need of collecting evidence from several small underpowered RCTs, the results of which needed to be combined to reach conclusive evidence on a given topic in view of the complexities of designing and conducting adequately powered RCTs. Conclusive evidence was therefore only a matter of adequately powered trials, and bias due to confounding, of course, was not an issue for randomized studies. On the contrary, the lack of randomization had always been seen as a threat to meaningfully pooling evidence in observational (i.e. non-experimental) studies, even though reliable methods to control for study-level confounding had been adopted. The above-mentioned debate [1–4] did not seem to reconcile the opposing points of view. In replying to Shapiro's [1] criticism ('I propose that the meta-analysis of published non-experimental data should be abandoned'), Petitti [2] claimed that: 'then all of the non-experimental epidemiology fails for the same reason'; while Greenland [3] brilliantly argued: 'A major problem with his arguments is that they constitute a basis for banning observational epidemiology'. Shapiro's [4] rejoinder, however, termed meta-analysis as bad science, and many epidemiologists were accused of being 'willing to equate data aggregation with truth'. From this moment on, it seemed to the scientific community to be time to set up rules necessary to better resemble the approach to meta-analysis of RCTs. In 2000, the MOOSE statement (Meta-analysis of Observational Studies in Epidemiology) was released to this purpose [14]: inherent unmeasured bias as well as non-homogeneity in study design were the overall concerns for meta-

analysis of case-control studies or other non-experimental studies. Few years later, genetics would again stress a formal methodological framework to rely on for pooling evidence outside the familiar ground of RCTs [15]. As in the past for RCTs, there was a need to increase power due to the small effects sizes to detect (genetics often deals with relative risks not higher than 1.20). Certainly, the gain in power immaterial of control of bias would confirm Shapiro's misgiving about conferring 'false validity to a spurious result' [1]. The inappropriate use of meta-analysis of observational studies rises from the lack of possibility of controlling for unmeasured confounding at study level, and meta-analysis can never be qualitatively superior to single studies which might still suffer from residual confounding. To this non-superiority, moreover, meta-analysis adds, by pooling the available evidence, a high power which can only entail false validity, as Greenland [3] clearly acknowledges: 'Shapiro rightly warns that the omnipresence of bias and confounding makes a mockery of the narrow confidence limits one can obtain from meta-analysis'. However, the recent development of 'instrumental variable' approaches devised to mimic random assignment in the absence of randomization [16, 17], through which both measured and unmeasured confounding can be now taken into account, make the use of meta-analyses of non-experimental data much more reliable and put them into a broader perspective. Nevertheless, the approach must not be purely statistical. Sophisticated statistical techniques to pool data are very useful but, when used without a sufficient clinical or epidemiological background, might obscure differences whose nature is biological. An unwary use of the methodological tools can sometimes contribute to the spread and publication of heterogeneous results. The caveat of equating data aggregation with truth [4] should always be considered.

Dichotomy 3: Aggregate Data versus Individual Participant Data Meta-Analysis

At its beginning the practice of meta-analysis implicitly referred to combining evidence collected in the form of published data, namely data which were reported as a summary measure of effect and its variability (e.g. an odds ratio or a mean difference along with their confidence intervals). This is known as aggregate data meta-analysis, which is to be distinguished from individual participant data meta-analysis, which is when the original datasets for all participants in each study are available. It

is intuitive that this practice, which is now quite common, especially in genetics, has been a dream for many years. Very rarely individual data were directly accessible or made available by researchers; furthermore, in the past, statistical models were unable to handle data in this form until generalized linear mixed models were developed. Indeed, with (generalized) linear mixed models [11, 18] we pool individual data in a one-step process of estimation, allowing for possible between-studies heterogeneity of effect, which simply corresponds to analyzing data from a multi-centre trial where between-centres heterogeneity is accounted for.

Nowadays, generalizations of this approach have increased the flexibility of these models. These models can now handle continuous, binary, ordinal and count data endpoints, to pool mixed data – that is both individual and aggregate – to account for confounding at individual and aggregate levels, and to pool designs of different nature as clustered versus non-clustered design [18, 19].

Dichotomy 4: Frequentist versus Bayesian Approaches to Meta-Analysis

It is well known to any scientist that the discipline of statistics possesses two main and often opposing paradigms. Frequentist and Bayesian statistics are basically in contrast over what is said to be the form of population quantities. They are fixed unknown quantities for frequentists, whereas Bayesians attribute to them a probability distribution. Inferentially the core idea relies on Bayes theorem: that is, the actual result of an experiment is given by the posterior probability distribution, which corresponds to the product of a prior probability distribution (i.e. what we think we know about a given topic) and the likelihood (i.e. the evidence provided by our experiment). As a consequence, though often criticized for being subjective, the role of prior knowledge on a given subject, which Bayesian statisticians formalize mathematically with a probability distribution, belongs to the updating process of knowledge. On the contrary, according to the frequentist approach, the basis of investigation relies solely on the likelihood.

Frequentist statistics has always appeared as straightforward, more intuitive and less computationally demanding, let alone the fact that any subjectivity inherent in the Bayesian paradigm seemed to contrast with the definition of science itself. However, the recent contribution of Greenland [20] to this debate clarified how close the two worlds and how false all the charges against Bayesian

statistics are. Misconceptions on the questionability of the assumptions of Bayesian statistics as well as on the dependency on computer-intensive methods for estimation have been made explicit. Indeed, the assumptions of frequentist statistics are as subjective as those of Bayesian one [20]. The same misconceptions also concern the apparent antinomy of Bayesian versus frequentist meta-analysis.

A Bayesian meta-analysis is equivalent to a frequentist meta-analysis with the additional constraint that on all parameters to be estimated (usually the overall treatment effect and the between-study heterogeneity) a prior probability distribution must be specified. If we use non-informative priors, Bayesian meta-analysis yields the same results as classical frequentist meta-analysis, since the posterior combined estimate of treatment effect will be based only on the likelihood function. The scope of Bayesian meta-analysis, therefore, goes far beyond the classical frequentist approach, according to the ability of the investigator to use prior probabilities as a tool of sensitivity analysis in order to explore and quantify different hypotheses [21].

Interestingly, as a further practical advance, Greenland [20] showed how a Bayesian prior probability within a meta-analysis can be considered as an additional study to be included along with the set of studies available which represents the likelihood. This approach is known as data equivalents or data augmentation. Bayes theorem itself is in fact a simple product of terms where the posterior probability represents an update of the prior probability via the likelihood. By exploiting exchangeability within this product of terms a sudden turn to the possibility of meta-analysis is entailed: a prior could represent not only a study or the previous cumulative evidence, but also a possible study which we do not know (i.e. which is yet to come) and that could alter, in a sensitivity analysis fashion, the final combined estimate of the effect. The data augmentation or data equivalents approaches are known to be common practice in many contexts [22]. Using a similar approach, Bayesian meta-analysis also allows an easy handling of complex study design, as clarified above [18].

A further strength of Bayesian meta-analysis methods is that they employ exact statistics. This topic will be covered in the next section.

Overall, despite its wrongly supposed subjectivity, Bayesian meta-analysis has been fundamental in solving some unanswered questions as well as in disentangling delicate matters for which different analytical methods seemed not to provide a conclusive answer, as in the case of meta-analyses conflicting with mega-trials [21].

Dichotomy 5: Asymptotic versus Exact Methods for Non-Continuous Endpoints

Mainstream meta-analysis, like many other branches of biostatistics, has been mostly based on the natural framework of asymptotic normal theory. However, methodological research developed alternative models based on more appropriate methods which overcome several drawbacks that researchers had to face when dealing with binary, categorical or count endpoints (i.e. non-continuous), and necessarily had to resort to standard statistical methods relying on asymptotic normal theory. A definitive answer to these unsolved issues has been given by meta-analysis based on exact methods. Along with Bayesian statistics which per se uses exact methods, a general framework based on generalized linear mixed models was recently provided by Stijnen et al. [23]. First of all, their approach includes previous standard approaches as special cases, can be implemented with straightforward commercial software and overcomes all the concerns about the inclusion of studies with zero events in both arms as well as avoiding any arbitrary zero cell continuity correction. The reason is intuitive – within an exact model, for example the Fisher exact test in a 2×2 contingency table, a zero occurrence is mathematically treatable. This approach somehow reconciles the many debates that arose with the increasing attention given to safety data. In particular, the case of rosiglitazone is considered a paradigmatic example which we will address in the next sections [24]. When rare events are at stake, meta-analysis of efficacy data can suffer the same concerns as safety data. Of course, meta-analysis of safety normally deals with rare events and has often yielded conflicting evidence. According to the existing methods used, combined evidence is seldom consistent [24]. It has been shown that even slight differences in the zero cell correction techniques might produce different results. Formulas of asymptotic standard methods for meta-analysis commonly refer to relative measures of risks (e.g. odds ratios, relative risks, incidence rate ratios), which, with the exception of Peto's method based on the comparison between observed versus expected events, cannot be estimated when a zero event or count is present. On the other hand, generalized linear mixed models proposed by Stijnen et al. [23] are based on a bivariate or multivariate exact model, which means that combined estimate is modelled in each single arm as an 'absolute' measure of risk (i.e. logits, log risks, log incidence rates) and the relative measures of risks are defined by appropriate contrasts between arms as the difference of the pooled absolute risks. Exact methods cor-

rectly handle inferences for studies with rare events endpoints by using appropriate exact confidence intervals, do not introduce distortions for arbitrary zero cell corrections, and allow the inclusion of all the available evidence without any exclusion of studies due to the presence of zero cells in both arms.

Generalized Linear Mixed Models as an Overall Framework: Meta-Analysis of Rates, Multi-Arm, Multiple Outcomes Meta-Analysis and Network Meta-Analysis, Surrogate Endpoints and Dose-Response Relationships

We have shown in the previous sections how apparently opposing or conflicting approaches can be seen as belonging to a general framework where the second term of each single dichotomy is a generalization of the first one and represents a broader view of the classical meta-analysis methods. A bivariate approach restricted to one single arm provides a natural framework for meta-analysing rates as well as (which is rarely performed) mean values. Therefore, any meta-analysis of prevalence or incidence is a special case of a bivariate approach to meta-analysis, where no effect (i.e. difference) at all is under investigation, but simply an overall summary measure. By extending a bivariate approach for two-arm comparisons to multi-arms, we have a meta-analysis for multi-arm trials, where between-arm correlation must be accounted for. When a relationship exists between study arms, a dose-response meta-analysis is the natural generalization [25, 26]. Moreover, this structure applies also to meta-analysis of genetic studies, in which the number of alleles at risk formally corresponds to treatment dose levels [27]. On the other hand, when we have to deal with multiple outcomes, the same structure applies as if we were addressing correlation between arms [11]. Although the correlation to take into account in this case is between outcomes and not between arms, the concept of multiple outcomes and multi-arms are formally exchangeable. Furthermore, when addressing multiple outcomes, not only do we have to extract (if possible) from each single study information about within-study correlations between arms or outcomes, but between-outcomes correlation becomes an important issue [28].

In all cases sparse data are a threat or zero events need to be corrected by adding the usual quantity of 0.5 or, by other methods, of course, exact bivariate or multivariate methods should be used instead [23]. Linear and generalized linear mixed models have been used as well for par-

ticular areas of meta-analysis, often named complex evidence synthesis: the bivariate form of random-effects meta-analysis has been used to assess the relationship between treatment effect and baseline risk [11], the sensitivity and specificity of diagnostic tools [29], as well as the joint synthesis of correlated outcomes and the assessment of surrogacy in clinical trials endpoints [28, 30].

As for meta-analysis of studies of diagnostic test accuracy, the approach based on mixed models has shown to be very promising in neurology, and, as soon as the scope of individual participant data meta-analysis is understood, its usefulness will be greater. Following this approach, for instance, Foerster et al. [31] performed a meta-analysis which estimated the diagnostic test accuracy measures of diffusion tensor imaging for the diagnosis of amyotrophic lateral sclerosis.

Mixed treatment indirect and direct comparisons represent the last advance in applying generalized linear mixed models to meta-analysis [32]. The key aspect is that each single study represents only a subset of all possible treatment arms, and, a straightforward interpretation of a missing data issue provides the framework for indirect comparisons. This is also known under the name of network meta-analysis. The core idea relies on the assumption of missingness at random of specific arms which guarantees that randomization is preserved. The importance of indirect comparisons allowed by a network meta-analysis can be easily understood when the real head-to-head comparisons are not available due to specific preferences or pressure on the use of some comparators. A network meta-analysis, through its geometry (fig. 2), makes a network of RCTs easily interpretable and comparable. The amount of existing evidence for specific direct treatment comparisons can be straightforwardly captured, and inferences can also be obtained from indirect comparisons. In figure 2, circles represent treatments and connecting lines represent direct comparisons with the relevance stressed by the line thickness. A network geometry further highlights the lack of head-to-head comparisons which somehow have been avoided. Particular concerns refer to heterogeneity and incoherence [32], where incoherence represents the possible discrepancies in results obtained by direct or indirect evidence.

Publication Bias

A particular concern in meta-analysis is publication bias, which refers to the differential propensity to publish study results according to their positive or negative find-

Fig. 2. Representation of a hypothetical network geometry. Circles refer to different treatments, while connecting lines represent direct drug comparisons. The thickness of the connecting lines is proportional to the amount of available evidence of specific comparisons.

ings [33]. Negative results have never been appealing. Eureka ('I have found') is the keyword in science. Nobody would be so enthusiastic to say 'I have not found'. Therefore, it is evident how negative findings, especially when coming from small-sized studies, can receive little attention from journals; it is also clear that researchers themselves can be less inclined to submit a paper with negative results.

A graphical inspection of a funnel plot, which reports a measure of effect in the x-axis and a measure of study size in the y-axis, is the most common form to address publication bias. The size of a study can be expressed as sample size, standard error or precision (which corresponds to the inverse of standard error itself). Usually, under the hypothesis of complete availability of evidence on a given effect, it is common to see that such a plot has the symmetrical form of a funnel. Therefore, asymmetry of the funnel plot means that publication bias is present. To this purpose many formal tests for publication bias assessment have been provided, none of them being completely satisfactory [34].

In our view, the most intriguing approach has been provided by the 'trim and fill' method [33], which resembles a Bayesian-flavoured sensitivity analysis such as the data augmentation approaches proposed by Greenland [20] and Greenland and Christensen [22]. Trim and fill methods practically formulate a conjecture about the missing studies, and formally reassess the combined estimate of treatment effect once the hypothetical missing studies have been included.

When present, publication bias is always a threat to a meta-analysis: in most cases it makes the pooling of the

data invalid since not all the possible evidence is taken into account. Sensitivity analyses using Trim and Fill can directly assess the potential impact of missing studies on the overall conclusions of the meta-analysis.

Overall, it is well known that tests for publication bias have low statistical power and can only be used in sensitivity analysis. On the other hand, an important drawback of publication bias detection is that some tests can produce significant results even though publication bias is not present. That is, all those tests which base their significance on the ability of detecting asymmetry in a funnel plot can be heavily influenced by the presence of heterogeneity. In these situations it is mandatory to distinguish between actual publication bias and the presence of heterogeneity, which makes the available tests wrongly significant and misleading. A recent solution to this problem has been proposed by Peters et al. [35]. Fortunately, as registration prior to their initiation is now required for clinical trials and a register of observational studies is going to be a natural consequence of it; it is likely that publication bias will soon become a less threatening concern for meta-analysis.

Examples of Advances in Meta-Analysis in the Specific Areas

The first example is taken from a recent publication on the effects of anti-platelet therapy for patients with chronic kidney disease [36]. Here the main purpose was to obtain an overall estimate of treatment effect on mortality and cardiovascular and bleeding outcomes. Most of the outcomes had a rare events distribution, and some of the trials reported zero events for both arms. Standard asymptotic methods would fail to include all available evidence and could be sensitive to zero cell continuity correction. On the other hand, the approach of Stijnen et al. [23], based on exact methods, would overcome those concerns but could be implemented only when crude data reported as 2×2 tables were available. Unfortunately, an important study to be included, the CURE trial [36, 37], only reported relative risks and confidence intervals. A trade-off between allowing for possible bias using standard methods and resorting to exact frequentist methods though discarding the crucial information given by the CURE trial was not convincing. The authors satisfactorily attempted a third way [20]. The approach of data equivalents proposed by Greenland allowed the use of Bayesian statistics in a very easy fashion which permitted both the handling of crude data whenever available and,

therefore, taking into account the nature of rare events and including the CURE trial in its aggregate form of a relative risk, considering it as a prior probability of the combined effect. A general Bayesian approach which includes this as an alternative has been proposed by Sutton et al. [18].

The debate on the safety of rosiglitazone serves as a second example. Rosiglitazone is a drug for treating hyperglycaemia in patients with type 2 diabetes. Concerns arose regarding its cardiovascular safety in 2007 following the publication of a meta-analysis which proved a significantly higher risk for myocardial infarction and a borderline significant higher risk for cardiovascular mortality. In 2010 the European Medicines Agency recommended the suspension of the drug from the European market, while in the USA, although recommendations were made to take the drug off the market, it is still present but subject to important restrictions. This controversial issue of cardiovascular safety [38] recently received a thorough review from a methodological perspective [24]. Although we can admit that aspects such as underreporting or inappropriate adjudication of events can be seriously misleading for meta-analysis of trials which were not formally designed to specifically address cardiovascular safety, it is surprising to see how some oversights affect this review. First, there is no reference to the paper by Stijnen et al. [23] published at least 2 years before, whose frequentist exact method, as described above, can straightforwardly handle trials with zero events in both arms, avoid arbitrary continuity corrections and yield correct inferences. Second, the strengths of Bayesian exact methods are underestimated due to a presumed weakness based on the possible disagreement in specifying the priors. Priors are not necessarily of concern: frequentist exact methods and Bayesian ones yield the same results when non-informative priors are used. Here, therefore, a final answer could have been provided by appropriately addressing the meta-analysis via Bayesian or frequentist exact methods. Furthermore, in our view, the meta-analysis of rosiglitazone would be a natural setting for a thorough Bayesian reassessment in the methodological wake of magnesium trials in myocardial infarction [21].

The third example refers to the issue of surrogate endpoints for clinical trials in multiple sclerosis [39]. The extent of surrogacy in clinical trials is typically controversial and needs an overall assessment which should necessarily refer to more than one study [40]. The first attempts of meta-analytic approaches to validate surrogate outcomes suffered from addressing the problem only at trial level. Formally, a bivariate meta-analytic approach based

on linear mixed models has been provided to assess the degree of surrogacy both at individual level and at trial level [30]. Indeed, causation should be proven at individual level, otherwise ecological bias can still be a threat. The paper by Sormani et al. [39] is an elegant attempt to provide a solution to this controversial issue in multiple sclerosis. Magnetic resonance active lesions as a possible surrogate for relapses are investigated in two clinical trials both separately and as an individual data meta-analysis according to the four Prentice criteria. However, the novelty of the meta-analytic method proposed by Buyse et al. [30] would allow going beyond the classical Prentice criteria. Two easily interpretable measures of squared correlation between surrogate and true endpoint are given both at trial and at individual level, even though the model formulation can become complex and a sufficient number of trials is required for implementation.

Conclusions

We have shown how all the advances in meta-analysis naturally merge into the unified framework of (generalized) linear mixed models and reconcile apparently con-

flicting approaches. Bayesian meta-analysis allows further generalization via sensitivity analysis and the use of priors as a data augmentation. All these complex models can be easily implemented with the standard commercial software available, and syntax codes are provided in most of the original articles [11, 18, 21, 23, 26, 29, 30]. Nevertheless, researchers must keep in mind that the purpose of a meta-analysis is to assess the effect size of a biological phenomenon. Meta-analyses have progressed beyond combining studies in order to increase statistical power as the main focus. Data pooling can now address complex research questions that aim to distinguish between true and artefactual effects, once the presence of heterogeneity has been detected. Meta-analyses can be used to generate new hypotheses such as the identification of a particular subpopulation or subpopulations where a treatment may be of benefit. Therefore, a new study in the relevant subpopulation could be designed with adequate statistical power to provide a conclusive answer.

Disclosure Statement

The authors do not have any conflict of interest.

References

- Shapiro S: Meta-analysis/shmeta-analysis. *Am J Epidemiol* 1994;140:771–778.
- Petitti DB: Of babies and bathwater. *Am J Epidemiol* 1994;140:779–782.
- Greenland S: Can meta-analysis be salvaged? *Am J Epidemiol* 1994;140:783–787.
- Shapiro S: Is there is or is there ain't no baby? Dr. Shapiro replies to Drs. Petitti and Greenland. *Am J Epidemiol* 1994;140:788–791.
- Sutton AJ, Higgins JP: Recent developments in meta-analysis. *Stat Med* 2008;27:625–650.
- Solomon PJ: Some statistics in bioinformatics: the fifth armitage lecture. *Stat Med* 2009;28:2833–2856.
- Pearson K: Report on certain enteric fever inoculation statistics. *Br Med J* 1904;2:1243–1246.
- Cochran WG: The combination of estimates from different experiments. *Biometrics* 1954;10:101–129.
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. *Br Med J* 2003;327:557–560.
- DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–188.
- van Houwelingen HC, Arends LR, Stijnen T: Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21:589–624.
- Thompson SG: Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994;309:1351–1355.
- Higgins JP, Thompson SG: Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23:1663–1682.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB: Meta-analysis of observational studies in epidemiology – a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA* 2000;283:2008–2012.
- Minelli C, Thompson JR, Abrams KR, Thakkinian A, Attia J: The quality of meta-analyses of genetic association studies: a review with recommendations. *Am J Epidemiol* 2009;170:1333–1343.
- Bennett DA: An introduction to instrumental variables analysis. Part 1. *Neuroepidemiology* 2010;35:237–240.
- Bennett DA: An introduction to instrumental variables. Part 2. Mendelian randomisation. *Neuroepidemiology* 2010;35:307–310.
- Sutton AJ, Kendrick D, Coupland CA: Meta-analysis of individual- and aggregate-level data. *Stat Med* 2008;27:651–669.
- Bacci S, Di Paola R, Menzaghi C, Di Fulvio P, Di Silvestre S, Pellegrini F, Baratta R, Marucci A, Mastroianno S, Fini G, Formoso G, Consoli A, Perticone F, Frittitta L, Pandolfi A, Trischitta V: ENPP1 Q121 variant, increased pulse pressure and reduced insulin signaling, and nitric oxide synthase activity in endothelial cells. *Arterioscler Thromb Vasc Biol* 2009;29:1678–1683.
- Greenland S: Bayesian perspectives for epidemiological research. I. Foundations and basic methods. *Int J Epidemiol* 2006;35:765–775.
- Higgins JPT, Spiegelhalter DJ: Being skeptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. *Int J Epidemiol* 2000;31:96–104.
- Greenland S, Christensen R: Data augmentation for Bayesian and semi-Bayesian analyses of conditional-logistic and proportional-hazards regression. *Stat Med* 2001;20:2421–2428.
- Stijnen T, Hamza TH, Ozdemir P: Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med* 2010;29:3046–3067.

- 24 Lane PW: Meta-analysis of incidence of rare events. *Stat Methods Med Res* 2013;22:117–132.
- 25 Greenland S, Longnecker MP: Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* 1992;135:1301–1309.
- 26 Rota M, Bellocco R, Scotti L, Tramacere I, Jenab M, Corrao G, La Vecchia C, Boffetta P, Bagnardi V: Random-effects meta-regression models for studying nonlinear dose-response relationship, with an application to alcohol and esophageal squamous cell carcinoma. *Stat Med* 2010;29:2679–2687.
- 27 Bagos PG: A unification of multivariate methods for meta-analysis of genetic association studies. *Stat Appl Genet Mol Biol* 2008;7:article31.
- 28 Kirkham JJ, Riley RD, Williamson PR: A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Stat Med* 2012;31:2179–2195.
- 29 Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PMM, Zwinderman AH: Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–990.
- 30 Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H: The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1:49–67.
- 31 Foerster BR, Dwamena BA, Petrou M, Carlos RC, Callaghan BC, Pomper MG: Diagnostic accuracy using diffusion tensor imaging in the diagnosis of ALS: a meta-analysis. *Acad Radiol* 2012;19:1075–1086.
- 32 Mills EJ, Thorlund K, Ioannidis JP: Demystifying trial networks and network meta-analysis. *BMJ* 2013;346:f2914.
- 33 Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR: Empirical assessment of effect of publication bias on meta-analyses. *Br Med J* 2000;320:1574–1577.
- 34 Moreno SG, Sutton AJ, Ades AE, Stanley TD, Abrams KR, Peters JL, Cooper NJ: Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med Res Methodol* 2009;9:2.
- 35 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L: Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol* 2008;61:991–996.
- 36 Palmer SC, Di Micco L, Razavian M, Craig JC, Perkovic V, Pellegrini F, Copetti M, Graziano G, Tognoni G, Jardine M, Webster A, Nicolucci A, Zoungas S, Strippoli GF: Effects of antiplatelet therapy on mortality and cardiovascular and bleeding outcomes in persons with chronic kidney disease: a systematic review and meta-analysis. *Ann Intern Med* 2012;156:445–459.
- 37 Eltai M, Tonelli M, Mann JF, Sitkei E, Lewis BS, Hawken S, CURE Trial Investigators: Renal function and outcomes in acute coronary syndrome: impact of clopidogrel. *Eur J Cardiovasc Prev Rehabil* 2007;14:312–318.
- 38 Nissen SE, Wolski K: Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007;356:2457–2471.
- 39 Sormani MP, Stubinski B, Cornelisse P, Rocak S, Li D, De Stefano N: Magnetic resonance active lesions as individual-level surrogate for relapses in multiple sclerosis. *Mult Scler* 2011;17:541–549.
- 40 Strippoli GF, Craig JC: Hypothesis versus association: the optimal hemoglobin target debate. *Am J Kidney Dis* 2005;46:970–973.