Meta-analysis: pitfalls and hints

T. Greco[1,2], A. Zangrillo[1], G. Biondi-Zoccai[3,4], G. Landoni[1]

[1]Department of Anesthesia and Intensive Care, San Raffaele Scientific Institute, Milan, Italy; [2]Section of Medical Statistics and Biometry Giulio A. Maccacaro, Department of Occupational and Environmental Health, University of Milan, Milan, Italy; [3]Department of Medico-Surgical Sciences and Biotechnologies, Sapienza University of Rome, Latina, Italy; [4]Meta-analysis and Evidence based medicine Training in Cardiology (METCARDIO), Ospedaletti, Italy

ABSTRACT

The present work is an overview of the main pitfalls which may occur when a researcher performs a meta-analysis. The main goal is to help clinicians evaluate published research results. Organizing and carrying out a meta-analysis is hard work, but the findings can be significant. Meta-analysis is a powerful tool to cumulate and summarize the knowledge in a research field, and to identify the overall measure of a treatment's effect by combining several conclusions. However, it is a controversial tool, because even small violations of certain rules can lead to misleading conclusions. In fact, several decisions made when designing and performing a meta-analysis require personal judgment and expertise, thus creating personal biases or expectations that may influence the result. Meta-analysis' conclusions should be interpreted in the light of various checks, discussed in this work, which can inform the readers of the likely reliability of the conclusions. Specifically, we explore the principal steps (from writing a prospective protocol of analysis to results' interpretation) in order to minimize the risk of conducting a mediocre meta-analysis and to support researchers to accurately evaluate the published findings.

Keywords: *meta-analysis, systematic review, limits, difficulties, recommendations.*

INTRODUCTION

Many of the groups… are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved (Karl Pearson, 1904).

Meta-analysis refers to the analysis of analyses. I use it to refer to the statistical analysis of a large collection of results from individual studies for the purpose of integrating findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding literature (Gene Glass; 1976).

Corresponding author:
Teresa Greco, MSc
Department of Anesthesia and Intensive Care
San Raffaele Scientific Institute
Via Olgettina 60, 20132 Milan, Italy
e-mail: greco.teresa@hotmail.it

Karl Pearson (1) was probably the first medical researcher to report the use of formal techniques to combine data from different studies when examining the preventive effect of serum inoculations against enteric fever.

All individual estimates were presented for the first time in a table, together with the pooled estimate. However, a method for uncertainty estimation had not yet been identified. Although such techniques would be widely ignored in medicine for many years to come (2), social sciences, especially psychology and educational research, showed particular interest in them. Indeed, in 1976 the psychologist Gene Glass (3) coined the term "meta-analysis" in a paper entitled "Primary, Secondary and Meta-analysis of Research", to help make sense of the growing amount of data in literature. Since the 80s, the amount of information generated

by meta-analyses grew constantly, up to the point of becoming overwhelming. A PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) search of the word "meta-analysis", in the title or in the abstract, yielded 39,840 hints (update at December 31th, 2012), 7,665 (19%) of them only in the year 2012 (*Figure 1*).

Meta-analysis is a powerful tool to cumulate and summarize the knowledge in a research field through statistical instruments, and to identify the overall measure of a treatment's effect by combining several individual results (4). However, it is a controversial tool, because several conditions are critical and even small violations of these can lead to misleading conclusions. In fact, several decisions made when designing and performing a meta-analysis require personal judgment and expertise, thus creating personal biases or expectations that may influence the result (5, 6).

As statistical means of reviewing primary studies, meta-analyses have inherent advantages as well as limitations (7). Pooling data through meta-analysis can create problems, such as non linear correlations, multifactorial rather than unifactorial effects, limited coverage, or inhomogeneous data that fails to connect with the hypothesis. Despite these problems, the meta-analysis method is very useful: it establishes whether scientific findings are consistent and if they can be generalized across populations, it identifies patterns among studies, sources of disagreement among results, and other interesting relationships that may emerge in the context of multiple studies. This short article introduces the basic critical issues in performing meta-analysis with the aim of helping clinicians assess the merits of published results.

Meta-analysis' protocol registration

It is important to write a prospective analysis' protocol, which specifies the objectives and methods of the meta-analysis. Having a protocol can help restrict the risk of biased post hoc decisions in methods, such as selective outcome reporting. The PRISMA (Preferred Reporting Items Systematic Reviews and Meta-Analysis) guidelines (8) recommend the prior registration of the protocol of any systematic
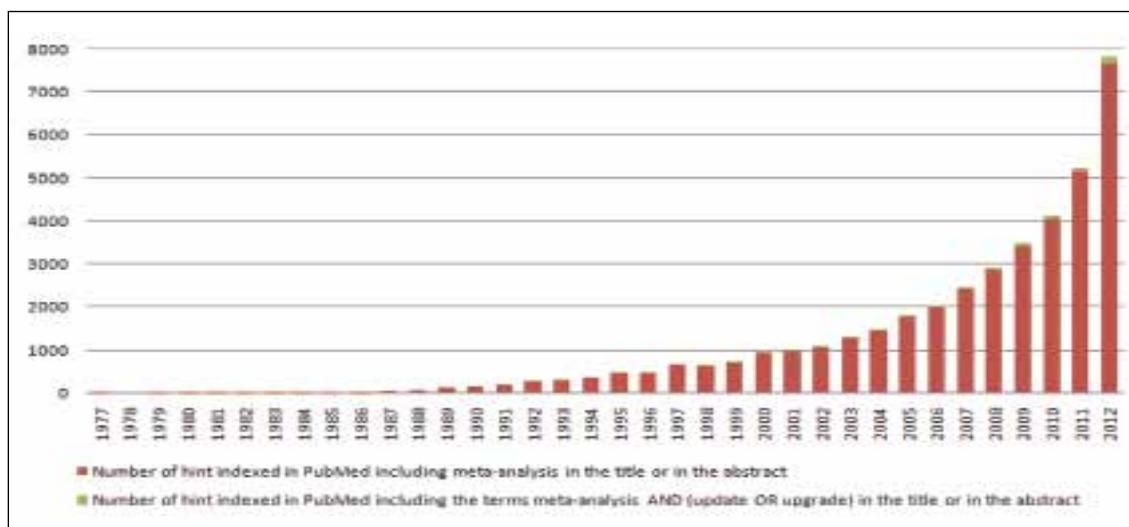
Figure 1 - *Amount of information generated by meta-analyses. PubMed search of the words "meta-analysis" in the published literature.*

review and meta-analysis, requiring that this protocol should be made accessible before any hands-on work is done. The prior registration (i.e. through PROSPERO - International prospective register of systematic reviews - http://www.crd.york.ac.uk/Prospero/) should prevent "the risk of multiple reviews addressing the same question, reduce publication bias, and provide greater transparency when updating systematic reviews". It is also true that meta-analyses are published only after passing through at least two steps: peer reviews and an editorial decision. These filters may be sufficient to decide whether a meta-analysis is good and novel enough to deserve publication. Takkouche B et al. (9) stated that an additional committee or register does not increase the quality of what is published but it only increases bureaucracy.

Rigorous meta-analyses undertaken according to standard principles (pre-specified protocol, comparable definitions of key outcomes, quality control of data, and inclusion of all information available) will ultimately lead to more reliable evidence on the efficacy and safety of interventions than either retrospective meta-analysis (10).

Identification and selection of studies

The first reason to criticize the meta-analytic method is that it provides evidence extracted and integrated from a number of primary studies, not from a random sampling; thus, results cannot lead to test relations such as causality (11). However, meta-analysis may lead to support or rejection of the generalization of primary evidence, and may contribute to direct future research in a field. Moreover, meta-analysis results can improve understanding, but sometimes they may not be very helpful in clinical practice. In this context, the definition of the scientific start-point (population and intervention) is crucial: the clinical question can either be broad or very narrow. Broad inclusion criteria could increase the heterogeneity between studies, making it difficult to apply the results to specific patients; narrow inclusion criteria make it hard to find pertinent studies and to generalize the results in clinical practice. Hence, the researcher should find the right compromise, focusing on the benefits for the patient. One of the aims of meta-analysis is to take into account all the available evidence from multiple independent sources to evaluate an hypothesis (6). However, meta-analysis usually includes only a small fraction of the published information, often derived from a small range of methodological designs (i.e. meta-analysis restricted to randomized clinical trials or to English languages). It is also true that with limited resources it is impossible to identify all the evidence available in the literature. Systematic reviews, in contrast to traditional narrative reviews, require an objective and a reproducible search of a series of sources to identify as many relevant studies as possible (12). The search strategy should be comprehensive and sensitive; searching more than one computerized database is strongly recommended. Commonly searched databases are: MEDLINE (http://www.nlm.nih.gov/pubs/factsheets/medline.html), including PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), The Cochrane Central Register of Controlled Trials (CENTRAL) (http://www.mrw.interscience.wiley.com/cochrane), and EMBASE (http://www.embase.com). These databases are available to individuals free of charge, on a subscription or on a 'pay-as-you-go' basis. They can also be available free of charge through national provisions, professional organization or site-wide licenses at institutions such as universities or hospitals. There are also regional electronic bibliographic databases that include publications in local languages (12). Additional studies can be identified employing the "backward snowballing" (i.e.

scanning of references of retrieved articles and pertinent reviews) or investigating the "grey literature", namely the literature that is not formally published in sources such as books or journal articles (i.e. personal communications, conferences, abstracts, etc).

Authors often provide supplementary data, not included in the original publications or relative to unpublished studies. Decisions regarding what primary evidence to include in a meta-analysis depend on evidence availability. Practical problems, regarding access to primary data, include studies published in languages foreign to the researcher and evidence available only confidentially or in the "gray literature" of congress and dissertations. Similar issues are faced by analysts who want to perform a meta-analysis with individual patient data (which has several advantages over analysis on aggregate-level data (13)), since patient-level data is often confidential or protected by corporate interests. Moreover, many other biases linked to study selection may influence the estimates and the interpretation of findings: citation bias, time-lag bias and multiple publications bias (12).

To overcome these biases, several tools are available. For example, the sensitivity analysis can spot bias by exploring the robustness of the findings under different assumptions.

Quality of included studies
The conclusions of a meta-analysis depend strongly on the quality of the studies identified to estimate the pooled effect (14). The internal validity may be affected by errors and incorrect evaluations during all the phases of a clinical trial (selection, performance, attrition, detection bias (15)), so the assessment of the risk of study bias is a central step when one carries out a meta-analysis. The quality of randomized clinical trials should be evaluated with regard to randomization, adequate blinding and explanation for dropouts and withdrawals, which addresses the issues of both internal validity (minimization of bias) and external validity (ability to generalize results) (16).

The information gained from quality assessment is fundamental to determine the strength of inferences and to assign grades to recommendations generated within a review. The main problem during the quality assessment process is the inconsistent base for judgment: if the studies were re-examined, the same trained investigator might alter category assignments (6). The investigator may also be influenced (consciously or unconsciously) by other unstated aspects of the studies, such as the prestige of the journal or the identity of the authors (6).

The published work can and should explain how the reviewers made these judgments, but the fact remains that these approaches can suffer from substantial subjectivity. Indeed, it is strongly recommended that reviewers use a set of specific rules to assign a quality category, aiming for transparency and reproducibility.

Publication bias
The biggest potential source of type I error (increase of false positive results) in meta-analysis is probably publication bias (14).

This occurs when, in clinical literature, statistically significant "positive" results have either a better chance of being published, are published earlier or in journals with higher impact factors, and are more likely to be cited by others (17). The graphical representation to evaluate the presence of publication bias is the funnel plot. In a funnel plot, effect size is plotted versus a measure of its precision, such as sample size. If no publication bias were present, we would expect that the effect size of each included study to be symmetrically distributed around the underlying true effect size, with more random variation of this value in smaller studies. Asymmetry or gaps in the

plot are suggestive of bias, most often due to studies which are smaller, non-significant or have an effect in the opposite direction from that expected, having a lower chance of being published (14). Therefore, it is important to note that conclusions exclusively based on published studies can be misleading. Methods as Trim and Fill (18) allow estimation of an adjusted meta-analysis in the presence of publication bias.

Small-study effect

The small-study effect occurs when small studies have systematically different effects from the large ones.

It has often been suggested that small trials tend to report larger treatment benefits than larger trials (19, 20).

Such small-study effects can result from a combination of lower methodological quality of small trials or publication bias (small studies with negative effects are unpublished or less accessible than larger studies) or other reporting biases (15). However, this effect could also reflect clinical heterogeneity, if small trials were more careful in selecting patients, so that a favorable outcome of the experimental treatment might be expected (21). Researchers that are worried about the influence of small-study effects on the results of a meta-analysis in which there is evidence of between-study heterogeneity ($I^2 > 0$) should compare the fixed- and random-effects estimates of the treatment effect. If the estimates are similar, then any small-study effect has little effect. If the random-effects estimate is more beneficial, researchers should consider whether it is reasonable to conclude that the treatment was more effective in smaller studies. This is because the weight given to each included study through the random effect model is less influenced by the sample size than that given by means of the fixed effects model. In the eventuality that the small-study effect is present, the researcher should consider ana-

lyzing only large studies (if these tend to be conducted with more methodological stringency (22)). One must note that if there is no evidence of heterogeneity between studies, the fixed- and random-effects estimates will be identical, so there will be an actual difficulty in identifying the small-study effect (12).

Data analysis

The degree of heterogeneity is another important limitation, and the random effects model should be used during the data analysis phase to incorporate in the treatment effect the identifiable or non-variability between-studies (23). It is fundamental to observe that exploring heterogeneity in a meta-analysis should start at the stage of protocol writing, by identifying a priori which factors are likely to influence the treatment effect. Visual inspection of the meta-analysis plots may show whether the results of a subgroup of studies have the same overall direction of the treatment effect. One should pay attention to meta-analysis in which results have a discordant treatment effect for groups of studies and no explanation of variance has been done. Sources of variation should be identified and their impact on effect size should be quantified using statistical tests and methods, such as analysis of variance (ANOVA) or weighted meta-regression (14).

Actually, when high heterogeneity is evident, individual data should be not pooled and definitive conclusions should be drawn when more studies become available. Moreover, meta-analysis makes it possible to look at events that were too rare in the original studies to show a statistically significant difference. However, analyzing rare events represents a problem because small changes in data can determine important changes in the results and this instability can be exaggerated by the use of relative measures of effect instead of absolute

224

ones. To overcome this problem several methods have been proposed (12,24,25). Another problem that affects meta-analysis carried out with aggregate data, is the ecological fallacy that arises when the averages of the patient's features fail to properly reflect the individual-level association (26). The best scenario is when data at an individual-level is available, but it is equally true that there is resistance from authors to allow ready access to their own dataset containing individual patient data. Very often aggregate data is the only information offered.

Finally, it is essential to spend a few moments discussing the common problem that occurs when one wants to perform multiple subgroup analysis, according to multiple baseline characteristics, and then examine the significance of effects not set a priori into the protocol.

Testing effects suggested by data and not planned a priori considerably increase the risk of false-positive results (27). To minimize this error it is important to identify the effects to test before data collection and analysis (5); otherwise, one may adjust the p-value according to the number of analysis performed. In general, post hoc analysis should be deemed exploratory and not conclusive.

CONCLUSION

Important decisions in a systematic review are often based on understanding the medical domain and not the underlying methodology. The clinical question must be relevant to clinicians and the outcomes must be important for patients. Efforts are made to avoid bias by including relevant research, using adequate statistical methodology and interpreting results based on the context and available evidence. Published reports should include quality criteria and

should describe the selected tools and their reliability and validity. The synthesis of the evidence should reflect the a priori analytic plan including quality criteria, regardless of statistical significance or the direction of the effect. Published reviews should also include justifications of all post hoc decisions to synthesize evidence. Organizing and carrying out a meta-analysis is hard work, but the findings can be significant. In the best-case scenario, by revealing the magnitude of effect sizes associated with prior research, meta-analysis can suggest how future studies might be best designed to maximize their individual power.

On the other hand, low-powered analysis based on a small number of studies can still provide useful insights by revealing publication bias through a funnel plot or highlighting a deficiency in a particular topic that deserves further attention. Meta-analysis represents a powerful way to summarize data and effectively increase sample size to provide a more valid pooled estimate.

However, the results of a meta-analysis should be interpreted in the light of the various checks previously discussed in this work, which can inform the readers of the likely reliability of the conclusions.

REFERENCES

1. Pearson K. Report on Certain Enteric Fever Inoculation Statistics. Br Med J 1904; 2: 1243-6.
2. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? Int J Epidemiol 2002; 31: 1-5.
3. Glass GV. Primary, secondary and meta-analysis of research. Educ Res 1976; 5: 3-8.
4. Biondi-Zoccai G, Landoni G, Modena MG. A journey into clinical evidence: from case reports to mixed treatment comparisons. HSR Proceedings in Intensive Care and Cardiovascular Anesthesia 2011; 3: 93-96
5. Walker E, Hernandez AV, Kattan MW. Meta-analysis: Its strengths and limitations. Cleve Clin J Med 2008; 75: 431-9.
6. Stegenga J. Is meta-analysis the platinum standard of evidence? Stud Hist Philos Biol Biomed Sci 2011; 42: 497-507.
7. Biondi-Zoccai G, Lotrionte M, Landoni G, Modena MG. The rough guide to systematic reviews and meta-analysis. HSR Proceedings in Intensive Care and Cardiovascular Anesthesia 2011; 3: 161-173.
8. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA state-

ment for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339: b2700.

9. Takkouche B, Norman G. Meta-analysis protocol registration: sed quis custodiet ipsos custodes? (but who will guard the guardians?).Epidemiology 2010; 21: 614-5.

10. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet 1998; 351: 47-52.

11. Timothy MD, Ryan WT. Do we really understand a research topic? funding answers thought meta-analysis. University of Technology, Sydney, Faculty of Business. Available on http://www.academia.edu/2385330/Do_We_Really_Understand_a_Research_Topic_Finding_Answers_Through_Meta-Analysis (Accessed February 2013)

12. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 (updated March 2011). The Cochrane Collaboration. Available on http://www.cochrane.org/handbook/chapter-6-searching-studies (Accessed at March 2013)

13. Lyman GH, Kuderer NM. The strengths and limitations of meta-analysis based on aggregate data. BMC Med Res Methodol 2005; 25: 14.

14. Harrison F. Getting started with meta-analysis. Methods in Ecology and Evolution 2011, 2: 1-10.

15. Egger M, Smith GD, Altman DG. Systematic Reviews in Health Care: Meta-Analysis in Context. London: BMJ Publishing Group 2001: 69-86.

16. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analysis? Lancet 1998; 352: 609-13.

17. Dubben HH, Beck-Bornholdt HP. Systematic review of publication bias in studies on publication bias. BMJ 2005; 331: 433-4.

18. Rothstein HR, Sutton AJ, Dr. Michael Borenstein Director Associate Professor Lecturer PI4Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Chapter 8. The Trim and Fill Method. New York John Wiley & Sons, Ltd 2005.

19. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000; 53: 1119-29.

20. Sterne JA, Egger M, Smith GD. Review Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001; 323: 101-5.

21. Nüesch E, Trelle S, Reichenbach S, et al. Small study effects in meta-analysis of osteoarthritis trials: meta-epidemiological study. BMJ 2010; 341: c3515.

22. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. Crit Care Med 2009; 37: 3114-9.

23. Ng TT, McGory ML, Ko CY, Maggard MA.Meta-analysis in surgery: methods and limitations. Arch Surg 2006; 141: 1125-30; discussion 1131.

24. Cai T, Parast L, Ryan L. Meta-analysis for rare events. Stat Med 2010; 29: 2078-89.

25. Lane PW. Meta-analysis of incidence of rare events. Stat Methods Med Res 2013; 22: 117-32.

26. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 1989; 18: 269-74.

27. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine-reporting of subgroup analysis in clinical trials. N Engl J Med 2007; 357: 2189-94.