JP Journal of Biostatistics

Published: December 2016
http://dx.doi.org/10.17654/BO013020181
Volume 13, Number 2, 2016, Pages 181-193 **ISSN: 0973-5143**

AN OVERVIEW OF APPROACH TOWARDS THE USE OF STATISTICAL MODELS IN META-ANALYSIS

Shaffi Ahamed Shaik

Department of Family and Community Medicine
College of Medicine
King Saud University
P. O. Box 2925
Riyadh 11461, Saudi Arabia
e-mail: shaffi786@yahoo.com

Abstract

This paper overviews the use of statistical models, fixed as well as random, in meta-analysis. Fifteen proponents and ten opponents have been observed through a qualitative analysis. We found that fixed and random models are unable to explain all the relevant variables, and hence recommend the use of a mixed model.

Introduction

Meta-analysis is the use of quantitative statistical procedures to integrate the results of individual studies to advance a research area. The term meta-analysis was first coined by Glass in [1], although techniques for combining studies were proposed as early as the 1930s by Fisher [2]. The work by Glass prompted a growth in the amount of research devoted to improving meta-analysis methodology, which in turn, lead to increased use of the methodology, particularly in epidemiology, psychology and education.

Received: September 28, 2016; Revised: October 20, 2016; Accepted: November 2, 2016
Keywords and phrases: meta-analysis, random effect model, fixed effect model and mixed model.

Despite its widespread use, the controversy in the application of fixed effects and random effects model persist.

What is meta-analysis?

The basic procedure for conducting a meta-analysis, as proposed by Glass, involves calculating a dimensionless outcome measure for each study, called the effect size. The estimated effect sizes are then combined in some way to obtain an estimate of the overall effect sizes across studies. For example, the estimated effect size (d_i) for studies which compare the average performance of a treatment with a control is often the standardized difference between the treatment means for that study

$$d_i = \frac{\bar{Y}_i^E - \bar{Y}_i^C}{S_i},$$

where S_i is the within group standard deviation for the *i*th study. If the studies involve testing for an association between a predictor and a dichotomous outcome, the effect size is usually given by the odds ratio or relative risk.

The choice of effect size estimator, and the methods used to combine the effects, although central to meta-analytic methodology, but we are concerned with the two different ways of conceptualizing the calculated effects. The two approaches (the random-effects approach and the fixed-effects approach) differ in the way in which they treat the between-study variation in the estimated effects.

The term between-study variation refers to the variation in the outcome measure (effect-size) due to differences in the component studies. Some factors which may differ between medical studies are treatment duration, study population, entry criteria, drug dosage or completeness of follow-up. A basic premise upon which combining results of different studies carried out is that, each study provides information about a common treatment effect. Differences between the studies will therefore affect the inferences which can be made from the meta-analysis.

Simple fixed effects model

The simplest fixed effects model assumes that all studies have the same (constant but unknown) effect size and that effect size estimates for individual studies differ due to sampling variability. The assumption of a single effect size across studies can be tested using a chi-square test for homogeneity [3]. More complex fixed effects models allow the treatment effects to be functions of the between-study variables such as drug dosage or duration of treatment. The basic assumption of the fixed effects approach is that between-study variation in treatment effects can be accounted for by knowable differences between the component studies. Regression models can be used to model the treatment effects in terms of these differences [4]. In this situation, the model goodness of fit can be used to test the homogeneity assumption. If the assumption of homogeneity is not rejected, the component effect sizes can be combined to obtain a meaningful estimate of the overall effect size. If however, the effects of the component studies are heterogeneous, the interpretation of the overall effect is more complicated. For example, if one half of the studies exhibited a substantial positive population effect size and the other studies had a substantial negative effect, it would be misleading to characterize the overall effect size as zero.

Random effects model

Random effects model for meta-analysis was first proposed by Hedges in 1983 [5]. The motivation for random effects model was that in many studies, it was not possible to explain a reasonable amount of the between-study variation in terms of a finite set of fixed effects. The random effects conceptualization treats between-study variation in treatment effects as if it arises from essentially random (or at least unquantifiable) differences between the studies. The random effects model assumes that the population values of the effect size are sampled from a distribution of effect-size parameters. The observed variability in sample estimates of effect size is partly due to the variability in the underlying population parameters and partly due to the sampling error of the estimator about the parameter value. The component studies are therefore assumed to be representative (if not a

random sample) of some universe of studies, each with different population effect sizes. The purpose of the statistical analysis is to understand the distribution of the true (population) treatment effects, which usually reduces to estimating the mean and variance of these effects.

Comparison of fixed effect and random effect model

A specific example

In order to explain on some of the points made above it is useful to compare a commonly used fixed effects model (the Pete-modified Mantel-Hansel Method (MH)) with a commonly used random effects model (the DerSimonian and Laird modified Cochran method (DL)).

The MH method calculates the overall treatment effect by weighting the individual treatment effects by the inverse of the within-study (sampling) variance. That is, the weights are based on the sample size used in each study. In contrast, to calculate the weighted average of rate differences in the DL model both between-study and within study variability are used.

In the DL model, if the between-study variance is relatively large this will dominate the weights and tend to weight all studies (large and small) about equally. If heterogeneity is low, the DL method like the MH method, weights the studies according to sample size. That is, if the meta-analysis is investigating studies that do not demonstrate between-study heterogeneity both methods yield similar results.

In general, the difference in the choice of weights means that the DL method is more conservative and yields wider confidence intervals than the MH method. That is, in the MH model only the within-study variability contributes to the standard error of the estimate of the overall treatment effect. In random effects models both within-study variability and between-study variability contribute to the standard error of the estimate of the mean of the treatment effect distribution. Consequently the MH model yields a smaller standard error and frequently obtains a statistically significant overall treatment effect when the DL model may not.

The objective of this study is to quantify the number of proponents and opponents of random effects model by listing out the statements for and against random-effects model.

Methods

To accomplish the objectives of this study a detailed overview of literature has been done by using MEDLINE and have collected from reputed journals approximately 50 articles related to these topics. As most of the debate, discussion and arguments were taken place in the mid 1980's the literature search was restricted up to 1994, as most of the experts opinions have work published within this period.

Qualitative Analysis

About 35 articles were identified with the proponents and opponents of application of random effects model. Among these articles it was found that 15 articles and their authors as proponents of random effects model and 10 as opponents based on their arguments. Some of the interesting arguments given by these authors have listed below.

The following were some of the statements (arguments) "for" the use of random effects models:

S. No.	Statements	Name of the proponent
1	"*Heterogeneity of treatment effects across studies is common and should be incorporated into the analysis. The random effects model incorporates this heterogeneity, however small in the analysis of the overall efficacy of the treatment. The method estimates the magnitudes of the heterogeneity and assigns a greater variability to the estimate of overall treatment effect to account for heterogeneity. The un weighted statistic which assigns an equal weight to each study may not be appropriate for testing homogeneity when differences in sample sizes and/or underlying proportions across studies are large*".	… Dersimonian and Laird [12]

2	"*Even though* 2 *methods* (*fixed-effect & random-effect*) *yielded similar results the random-effect method is preferred because a difference in absolute risk is more meaningful to clinicians than a RR or OR, and incorporating heterogeneity into the calculation of the overall variance* ***makes more sense from the statistical standpoint***".	… Hine et al. [13]
3	"*Let me propose another question for which the random effects model might be appropriate. For the clinicians who asked yesterday, "For my next patient what do I do*? *The O-E approaches the pooling that takes into account only variation within studies is irrelevant. If the clinicians can identify those studies that are closest to his practices then those are the studies he should concentrate on, not the average across heterogeneous studies. And if the clinician can't identify the study that represents something close to his or her practice, then it's the whole wide range of studies that must take into account. This next patient could be from a population at one end of the scale of effects to the other. So, pooling may be informative for answering the question that was posed yesterday, but it had better take into account the heterogeneity that time and again we've seen existent.* ***The C.I. for answering that question must be based on study-to-study as well as on within-study variability****"*.	… Fleiss [14]
4	"*We must bear in mind*; *however that acceptance of the hypothesis of homogeneity is often weak and does not leave us confident of a constant treatment effect unless all the individual studies have substantial sample sizes. Regardless of the acceptance or rejection of the homogeneity hypothesis, one should consider carefully the size of t, the among-study standard deviation, relative to the estimated size of the effect. At the very least, D & L method by explicitly incorporating variability among studies may provide a more realistic approach to combination of studies unless one is prepared to limit inference to the particular studies at hand. If there is heterogeneity that cannot be explained as a function of patient populations, protocols etc., then the* ***random effects model is more appealing philosophically****. The D and L approach gives the appropriate C.I's in this setting*".	… Berlin et al. [6]

5	"*We believe that the question of heterogeneity should be carefully examined in any meta-analysis. If heterogeneity is present but no explanation is found then a random effects model should be used. For random-effects the unbiased estimate of treatment effect* θ, *has smaller variance and is therefore preferable*".	… Whitehead and Whitehead [8]
6	"*When the research question concerns whether the treatment will have an effect on the average or whether exposure to a hypothesized risk factor will cause disease, on the model then the model of studies being random is the appropriate one. Here, we implicitly assume that there is a population of studies from which those included in the meta-analysis were sampled. It anticipates the possibility of future studies being conducted or even previously unknown studies uncovered*".	… Bailey [7]

The following were some of the arguments "against" the use of random effects models:

S. No.	Statements	Name of the proponent
1	"*When I do an overview of many trails, I'll calculate the p-value based on the variation that one would expect within each trail. This is the sort of conventional p-value calculation. I would expect the size of the effect in different trails actually to be somewhat heterogeneous. But I'll not let the random differences between different trails contribute to my final* *p-value or contribute to my final estimate of the magnitude of the effect or to the confidence intervals that I'll put about it. You see there are two approaches to this. One is called a random-effect analysis and the other is what I called the standard p-value analysis* (*or fixed effects analysis*). *The random effects analysis says,* '*look we have got a lot of different trail results here, what's the mean and what's the scatter of the different trail results*?' *I think that this is actually wholly wrong as an approach to overviews and trails. I think it does answer a question. But it's a very abstruse and uninteresting question. It's trying to say* '*what would happen if we choose another treatment at random from the universe of treatments that we could choose another*	… Peto [15]

	population at random from the universe of populations.' I think this is not an important question. The question of interest which I try to address by the standard p-value approach, is saying, and given the studies that people actually chose to do, have we observed more deaths in the treated groups that we would have expected just by the play of chance? I think that is the appropriate analysis, and ***that the random effects analysis is wrong; not statisfacally wrong, but commonsensically wrong. It's asking the wrong the question***".	
2	"*Several cautious apply to the use of random effect models:* 1. *In situations in which addition of a random effect to the model yields materially important changes in inferences, the degree of heterogeneity present will often* (*if not usually*) *be* ***so large as to nullify the value of the summary estimates*** (*with or without the random effect*). *Such a situation is indicative of the need to further explore sources of conflict among the study results.* 2. *Specific distributional forms for the* ***random effects have no empiric, epidemiologic, or biologic justification in typical applications****. Therefore, use of methods that specify the random-effect distribution should be accompanied by checks on the assumed distributional form.* 3. ***The summary estimate obtained from a random-effect model has no population-specific interpretation****, but instead represents the mean of a distribution that generates effects. Unlike a standardized rate ratio, it does not correspond to an average effect in a population.* *In essence, then a random-effect model exchanges a questionable homogeneity assumption for a fictitious random distribution of effects. The advantage purchased by this exchange is that the S.E.'s and C. Limits for any resulting estimates can more accurately reflect unaccounted for sources of variation is study results than can estimates from fixed-effect models; a drawback is that some simplicity of inter-predation is lost.* ***In any case, when residual heterogeneity is small relative to study-specific variance, essentially the same conclusions should be arrived at using either a fixed-effect or random-effect approach***".	… Greenland [16]

3	"*The fact that a common protocol is being followed in a co-operative trail does not necessarily mean that the patients are homogeneous. It is surprising how commonly large co-operative trails are recorded as pooled results without an* ***effort to examine between-center variance and without using a method to take heterogeneity into account such as the D&L procedure****"*.	… Chalmers [17]
4	"*We can use the random effects model to consider the question what is the probability that the experimental treatment is superior to the control at fourth center*? ***And the resulting estimate, would be an overestimate if the study had been chosen to be similar rather than at random***".	… Whitehead and Whitehead [8]
5	"*When the question concerns whether treatment has produced an* ***effect, on the average, or whether exposure has caused disease, on the average in the studies at hand, then the model of studies being random is not appropriate****. This question assumes that only the studies included in the meta-analysis are of interest and there is no interest in generalizing the result to the other studies"*.	… Bailey [7]
6	"*Standard Errors in the fixed effects model are unaffected by any artificial heterogeneity of results and so the only concern in artificial heterogeneity of results and so the only concern in using this model is the bias in estimation. For the random effects model, the estimation of the between-trail variance is essential both for the overview estimate of effect and also in ascertaining its S.E. Therefore an exaggerated treatment effect estimates for the variance leads to a weighted overview estimate that places more weight on small trails. Since trails that stop early, tend to be smaller and have exaggerated treatment effect estimates, the bias may be substantial in the overview estimate of the mean effect, m in the random effects model"*.	… Hughes et al. [18]
7	"*The random effect model implies that these studies are a random sample of some universe of studies. Thus any estimation would include a between study as well as a within component of variation. While some measure of between study variability may be useful, one can easily get into an awkward situation if this random effects approach is taken formally. Yet, combined with a between study component of*	… Demets [19]

	variation, the results may not allow a claim of a statistically difference. ***One can also question whether there really is a universe of trails. For these reasons, the fixed effect and the conditional inference seems most suitable***. *Certainly few, if any, multicenter trails even add a between center component in their references*".	
8	"*The random-effects method is no panacea for heterogeneity. Formal interpretation relies on the peculiar premise that the trails done are representative of some hypothetical population of trails, and on the unrealistic assumption that the heterogeneity between studies can be represented by a single variance. The results are also often strongly dependent on the inclusion or exclusion of small trails which may themselves reflect publications bias.* ***The random-effects methods may therefore give undue weight to small studies emphasizing poor evidence at the expense of good***".	… Thompson and Pocock [20]
9	"*The random-effects conceptualization is very appealing because it is simple and because it reflects the empirical fact that between-study differences may arise from so many causes that it may be difficult to characterize them as* '*systematic*'. *In spite of this appeal, there is something about random effects models that defies common sense. In the random-effects conceptualization, between-study difference in treatment effects is treated as totally non-systematic. Yet everyone who actually conducts clinical trials recognizes that different trails are systematically and purposefully different. Thus I feel uncomfortable with the random effects notion even though I am sympathetic to it, because I feel uncomfortable about the notion that we should treat observable difference among patient populations and study protocols as if they were purely random events*".	… Hedges [11]

Conclusions

In conclusion, the qualitative analysis of this overview about fixed and random effects models indicates that the meta-analysis should aim to explain any observed heterogeneity in treatment effects, rather than model these differences as a random effect without giving importance to the cause. It is

very obvious that we will not be able to explain all of the variation in terms of fixed effects. It may therefore be reasonable to use a *mixed model*, which would enable us to explain some of the variation in terms of fixed effects while allowing a random component to model the remainder of the between-study variation. It is also interesting to note that the references which give a more balanced approach to random-effects modeling are mainly articles written from the late 1980's onwards. The strongest opinions either for or against the approach were mainly expressed in earlier articles. Hence, it can infer that time and experience might have modified the approach towards implementing the random effects in a meta-analysis.

Acknowledgement

The authors thank the anonymous referees for their valuable suggestions which led to the improvement of the manuscript.

References

[1] G. V. Glass, Primary, secondary, and meta-analysis of research, Educational Researcher 5 (1976), 3-8.

[2] R. A. Fisher, Statistical Methods for Research Workers, 4th ed., Oliver and Boyd, London, 1932.

[3] L. V. Hedges, Estimation of effect size from a series of independent experiments, Psychological Bulletin 92 (1982), 490-499.

[4] G. V. Glass, Integrating findings: the meta-analysis of research, Review of Research in Education 5 (1977), 351-379.

[5] L. V. Hedges, A random effects model for effect sizes, Psychological Bulletin 93 (1983), 388-395.

[6] J. A. Berlin, N. M. Laird, H. S. Sacks and T. C. Chalmers, A comparison of statistical methods for combining event rates from clinical trials, Stat. Med. 8(2) (1989), 141-151.

[7] Kent R. Bailey, Inter-study differences: how should they influence the interpretation and analysis of results?, Stat. Med. 6 (1987), 351-358.

[8] A. Whitehead and J. Whitehead, A general parametric approach to the meta-analysis of randomized clinical trials, Stat. Med. 10(11) (1991), 1665-1677.

[9] T. D. Sector and S. G. Thompson, The potential and limitations of meta-analysis, J. Epidemio. Commun. Health 45 (1991), 89-92.

[10] S. Greenland and A. Slavan, Bias in the one-step method for pooling study results, Stat. Med. 9 (1990), 247-252.

[11] L. V. Hedges, Commentary, Stat. Med. 6 (1987), 381-385.

[12] R. Dersimonian and N. Laird, Meta-analysis in clinical trails, Controlled Clinical Trails 7 (1986), 177-188.

[13] Louis K. Hine, Nan M. Laird, Peg Hewitt and Thomas C. Chalmers, Meta-analysis of empirical long-term antiarrhythmic therapy after myocardial infarction, JAMA 262(21) (1989), 3037-3040.

[14] Joseph Fleiss, Discussion, Stat. Medicine 6 (1987), 359-360.

[15] Mr. Peto, Discussion, Stat. Med. 6 (1987), 241-244.

[16] Sander Greenland, Quantitative methods in the review of epidemiologic literature, Epidemiologic Reviews 9 (1987), 1-31.

[17] Thomas C. Chalmers, Problems induced by meta-analyses, Stat. Med. 10(6) (1991), 971-980.

[18] M. D. Hughes, L. S. Freedman and S. J. Pocock, The impact of stopping rules on heterogeneity of results in overviews of clinical trials, Biometrics 48(1) (1992), 41-53.

[19] David L. Demets, Methods for combining randomized clinical trails: strengths and limitations, Stat. Med. 6 (1987), 341-348.

[20] S. G. Thompson and S. J. Pocock, Can meta-analyses be trusted?, The Lancet 338(8775) (1991), 1127-1130.

[21] Stephen B. Thacker, Meta-analysis. A quantitative approach to research integration, JAMA 259(11) (1988), 1685-1689.

[22] Richard J. Light, Accumulating evidence from independent studies: what we can win and what we can lose, Stat. Med. 6 (1987), 221-228.

[23] William E. Boden, Meta-analysis in clinical trials reporting: has a tool become a weapon?, Amer. J. Cardiology 69(6) (1992), 681-686 (Editorial).

[24] John B. Carlin, Meta-analysis for 2×2 tables: a Bayesian approach, Stat. Med. 11 (1992), 141-158.

[25] Colin B. Begg and L. Pilote, A model for incorporating historical controls into a meta-analysis, Biometrics 47 (1991), 899-906.

[26] Joseph L. Fleiss and A. J. Gross, Meta-analysis in epidemiology with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique, J. Clin. Epid. 44(2) (1991), 127-139.

[27] Nathan L. Pace, Prevention of Succinylcholine Myalgias: a meta-analysis, Anesthesia and Analgesia 70(5) (1990), 477-483.

[28] Allan S. Detsky, C. David Naylor, Keith O'Rourke, Allison J. McGeer and Kristan A. L'Abbé, Incorporating variations in the quality of individual randomized trials into meta-analysis, J. Clin. Epid. 45(3) (1992), 255-265.

[29] Kristan A. L'Abbé, Allan S. Detsky and Keith O'Rourke, Meta-analysis in clinical research, Annals of Internal Medicine 107(2) (1987), 224-233.

[30] Judith E. Schmid, Gary G. Koch and Lisa M. LaVange, An overview of statistical issues and methods of meta-analysis, J. Biopharm. Stat. 1(1) (1991), 103-120.

[31] Z. B. Gerbarg and R. I. Horwitz, Resolving conflicting clinical trails: guidelines for meta-analysis, J. Clin. Epid. 41(5) (1988), 503-509.

[32] Charles H. Hennekens, Julie E. Buring and Patricia R. Hebert, Implications of overviews of randomized trials, Stat. Med. 6(3) (1987), 397-402.

[33] S. Yusuf, R. Simon and S. Ellenberg, eds., Proceedings of the Workshop on Methodological Issues in Overviews of Randomness Clinical Trails, May 1986, Stat. Med. 6 (1987), 217-403.

[34] L. V. Hedges and I. Olkin, Statistical Methods for Meta-analysis, Academic Press, Orlando, 1985.

[35] J. E. Hunter and F. L. Schmidt, Methods of Meta-analysis: Correcting Error and Bias in Research Findings, Sage Publications, Newbury Park, 1990.