Journal of Data Science **9**(2011), 389-397

Perspectives on Pooled Data Analysis: the Case for an Integrated Approach

Demissie Alemayehu
Pfizer, Inc.

Abstract: In the absence of definitive trials on the safety and efficacy of drugs, a systematic and careful synthesis of available data may provide critical information to help decision making by policy makers, medical professionals, patients and other stakeholders. However, uncritical and unbalanced use of pooled data to inform decision about important healthcare issues may have consequences that adversely impact public health, stifle innovation, and confound medical science. In this paper, we highlight current methodological issues and discuss advantages and disadvantages of alternative meta-analytic techniques. It is argued that results from pooled data analysis would have maximal reliability and usefulness in decision making if used in a holistic framework that includes presentation of data in light of all available knowledge and effective collaboration among academia, industry, regulatory bodies and other stakeholders.

Key words: Drug safety assessment, heterogeneity, meta-analysis, reporting bias.

1. Introduction

Randomized controlled trials (RCTs) are the gold standard for evidence-based decision making about the safety and efficacy of medicinal products. Accordingly, regulatory agencies and healthcare providers rely on data from such studies as the basis for approval and use for new drugs, biologics and medical devices. However, studies used for initial drug approval are typically limited in size and, hence, lack the strength to suggest safety signals for rare events or benefits in special populations.

To complement data from RCTs, it is now customary to rely on post-marketing observational and other epidemiological studies and pooled data analysis. In particular, safety evaluation concerning a new drug generally involves continuous monitoring of spontaneous reporting databases, despite the known limitations of data from such sources (Evans, 2002). Pharmacovigilance databases capture only

cases that have been reported, and as a result reporting bias complicates determination of causality in a comparative setting. Further the information captured is invariably incomplete, lacking critical information about severity of adverse events, causal relationships, concomitant medications and other confounding factors. There is often the potential of multiple reporting of the same case, generally making application of standard statistical procedures impractical. While there is considerable progress in the management, analysis and reporting of such data (see, e.g., DuMouchel, 1999; and Bate, *et al.*, 2002), much methodological work remains to be developed to bridge the gap between phrmacovigilance studies and RCTs..

On the other hand, the field of meta-analysis has received increasing attention both in the detection of potential safety signals and determination of the generalizablility of results from RCTs conducted in the overall patient population to special subgroups. Meta-analysis is the technique of synthesizing research results from previous separate but related studies by using various statistical methods. This is normally done in a systematic manner, which involves transparent processes for literature search, study selection, measure of effect size identification, analytical approaches and reporting of results.

With the growing awareness of the importance of the method, numerous guidelines, monographs and original research papers have been published on various aspects of the subject in recent years. For relevant literature see, for example, Sutton and Higgins (2008).

Like any other statistical endeavor, the conduct, analysis and reporting of meta-analysis, however, requires a careful evaluation of the underlying assumptions and the limitations of the approach in establishing causality. Indeed, uncritical and lopsided use of pooled data to inform decision about important healthcare issues may have consequences that adversely impact public health, stifle innovation, and confound medical science.

A recent meta-analysis conducted and reported by Nissen and Wolski (2007) relating to the cardiovascular effects of treatment with rosiglitazone for type 2 diabetes, illustrates both the strengths and limitations synthesizing information from different studies. In that study, the authors evaluated data from trials involving rosiglitazone, and concluded that the drug "... was associated with a significant increase in the risk of myocardial infarction and with an increase in the risk of death from cardiovascular causes that had borderline significance". The results of the paper were immediately picked up by the lay media, and fed to the general public often with no context or factual representation. The conclusion of the paper also became the focus of considerable discussion among patients, healthcare providers, policy makers and medical researchers. To date, several papers and commentaries have been published in major journals about

the strengths and weaknesses of the paper, and the work continues to generate mixed opinions among medical and statistical researchers (see, e.g., Drazen, et al, 2007; Tian, *et al.*, 2009; and Dahabreh, 2008).

In this paper, we highlight current methodological issues, discuss advantages and disadvantages of commonly used techniques, and suggest steps to be taken to maximize the reliability and usefulness conclusions from pooled data analysis in decision making. Particular attention is paid to the risk of undue dependence on models with heavy assumptions, the impacts of data extraction errors and publication bias, the need to present data with fair balance through effective strategies for dissemination of findings, and the importance of collaboration among stakeholders through mutual trust and constructive engagement.

2. Curtailing Bias

Perhaps the greatest challenge in meta-analysis is the handling of bias, the sources of which might be within or outside of the control of the researcher (Sterne, Gavaghan and Egger, 2000; and Sterne, Egger and Davey-Smith, 2001). Following well-established protocols developed to ensure the quality of meta-analysis, most sources of bias may be avoided. For example, bias that arises during the selection of studies for inclusion may be reduced by establishing criteria a priori based on study characteristics, accessibility of published results, and other quality considerations.

However, even with the most stringent steps taken, it is almost impossible to eliminate certain sources of bias, especially those that are beyond the control of the investigator. For example, the investigator has little recourse to assess the degree of, let alone eliminate, publication and reporting biases. Available methods, including the funnel plot and Egger's test, have limited capability to give definitive results about existence of publication bias. Even when there is evidence of publication bias, there are no reliable techniques to quantify the magnitude of bias or to adjust the estimated effect sizes for the associated bias. The recent drive to post study results in publicly accessible repositories would certainly help prevent the impact of biases arising from the apparent tendency to publish positive results over neutral or negative findings. However, since most analyses would continue to depend on data from legacy trials, analysts and users should exercise great caution in interpreting findings from such studies.

Other biases could arise through post-hoc manipulation of data and subjective choice of analytical methods dictated by interim study findings This may be minimized by pre-specifying the analytical strategy in a meta-analysis plan, which should also clearly state the hypothesis of interest and the steps to be taken for study selection and assessment of model assumptions.

Since most meta-analysis activities involve the synthesis of summary statistics

or aggregate data, concern is often expressed about the lack of patient-level data to take into account potential confounding factors. However, when the studies are homogenous, this issue may not have much relevance, especially if the pooled analysis relates to the target population studied in the individual RCTs. At the study level, randomization is expected, on the average, to ensure balance in patient characteristics among treatment groups across all possible allocations; and, therefore, it may be implausible to attribute an apparent treatment effect to imbalance in prognostic factors.

3. Methodological Issues

Despite the sizeable methodological advance made over the years, the analysis of pooled data is characterized by sundry issues and controversies about the choice of analytical techniques. One major issue that is central in the discussions surrounding the choice is the handling of heterogeneity or the degree of variability among studies contributing data for synthesis.

An approach that is conceptually appealing to clinicians for its simplicity is the strategy based on fixed effects models. Implicit in this approach is the assumption that studies are comparable with regard to such relevant attributes as methods, patients and measurements; and that a major source of variability that requires statistical manipulation is the within-study variation (Mantel and Haenszel, 1959; and Yusuf, *et al.*, 1985). By contrast, random effects models (DerSimonian and Laird, 1986) take into account both between-study and within-study variability, and as a result tend to be more "conservative" than fixed effects models. In practice, it is customary to use the fixed effects models when there is no evidence of heterogeneity, based on statistical testing, and to resort to random effects analysis, otherwise.

Neither method is, however, satisfactory in most practical applications. For example, a recent study by Schmidt and Hayes (2009) showed that fixed effects confidence intervals around mean effect sizes are, on average, 52% narrower than their actual width. In particular, nominal 95% fixed effects confidence intervals were found to have a level as low as 56%, on average. On the other hand, random effects models may suffer from poor estimation of among-study variance when there is little information, leading to overly conservative inferential results (Ziegler, Koch and Victor, 2007). Further, the normality assumption for underlying effects may not be justified. When the response variable is binary, rare events pose additional methodological challenge, since the common large-sample procedures may not give reliable results. In a recent article, Bradburn, *et al.* (2007) performed a fairly thorough comparison of various procedures using simulated data, and concluded that some of the most commonly used meta-analytic procedures tended to be biased when data were sparse. In another study, Tian,

et al. (2009) introduced an efficient method without artificial continuity corrections, and showed that exclusion of studies with no events, as was done in Nissen and Wolski (2007), could influence meta-analytic results.

When studies are not homogenous, the fixed and random effects approaches can give very contradictory answers. However, the problem of determining the existence of significant heterogeneity in meta-analysis is a difficult one and an area of ongoing research in the statistical literature. The traditional techniques, including Cochran's Q test and the I2 index, despite their widespread use, have lackluster performance in most practical situations. For example, the power of the former is low when the number of studies is small, and tends to be liberal if number of studies is large. The latter, which is often used as a complement to the Q test, has similar problems of power with a small number of studies (Huedo-Medina, 2006).

Meta-regression is an approach that has received increasing attention in the recent literature for handling heterogeneity by incorporating study characteristics in a regression setting. However, use of averages or study-level quantities in place of person-specific quantities can lead to biased results and incorrect conclusions (McIntosh, 1996). Further, as pointed out earlier, the need to control for confounding factors in a post-hoc manner may not be justified, especially if the population under study is the basis of the original randomization.

Bayesian meta-analysis has also gained considerable traction, in part because of the advance made in overcoming computational difficulties in the implementation of the procedure. Bayesian hierarchical models allow incorporation of study heterogeneity into the analysis of association, and permit inference relating to individual studies, on the basis of the combined information (Tweedie, *et al.*, 1996). Bayesian approaches, however, tend to give wide credible intervals, are not readily understandable to non-statisticians, may pose ambiguity in the choice of priors, and can also be computationally cumbersome, especially in applications involving random-effects models (Lambert, 2005).

Cumulative meta-analysis has also been proposed as an alternative strategy, especially when there is a desire to build on and leverage accumulating information over time. While its use is often advocated in a Bayesian setting, the potential for false positives in a frequentist framework renders its use to be limited for illustrative purposes only (Berkey, *et al.*, 1996).

4. Toward a More Integrated Strategy

As the preceding sections suggest, meta-analysis, despite its considerable utility to provide added precision and power in the evaluation of treatment effects, can also suffer from drawbacks that potentially may invalidate the study findings. Even in the best possible scenario, discovering association between treatment and

an outcome of interest does not necessarily translate into a determination of causation. Therefore, it is of utmost importance to view results of pooled data analysis in light of the totality of all available information and to communicate them in the proper context. The ultimate goal should be the promotion of the advancement of knowledge about the safety and efficacy of drugs, without causing unnecessary confusion and misunderstanding that may inevitably cause more harm than benefit to public health.

In tandem with meta-analysis, the researcher should perform a through review of the available literature for similar reports of the findings of the meta-analysis, including data from animal studies, trials on the same drug class, and pharmacological observations. A thorough investigation should be carried out to understand any potential biological basis, including pharmacogenomic evidence, to suggest a causal link.

Available observational data should be carefully analyzed to complement the findings of the meta-analyses based on RCTs, while recognizing event rates in clinical data may not be reflective of real-world observations. In particular epidemiological assessments should be conducted to evaluate expected event rates in comparable demographic groups in the general population. Pharmacovigilance databases should also be interrogated, using state-of-the-art data mining tools to see if the findings of the meta-analysis correlate to disproportional reporting based on spontaneous data.

The assessment of causality should also involve evaluation of potential dose-response to determine whether larger exposures tend to be associated with higher rates of disease. It may also be advisable to ascertain if cause and effect relationship has already been established for a similar exposure-disease combination. The temporal nature of the finding needs to be investigated by looking to see whether the cause precedes the effect. If the analysis is targeted at subgroups that were not the basis of the original randomization, it would be appropriate to assess the potential for effects of confounding factors.

A critical component of the integrated approach should be the formulation of a careful communication strategy that is aimed at presenting data with the proper context and fair balance. While calling attention to any untoward effect of a medicinal product should be a responsibility shared by all concerned, including patients, healthcare professionals, pharmaceutical companies and regulatory bodies, the responsibility must also be exercised with prudence, balancing the urge for transparency against scientific rigor and public safety. This could be carried out more effectively through enhanced collaboration among all stakeholders, including industry, academia and regulatory bodies, that is based on mutual trust and a common understanding of the ultimate goal, namely the protection of public safety and the advancement of medical science.

5. Conclusion

The synthesis of information in a meta-analytic framework plays a critical and complementary role in filling the knowledge gap that is created by the shortfalls of RCTs designed to assess the safety and efficacy of drugs for regulatory approval or promotional purposes. An essential requirement for optimal use of the approach is a thorough understanding of the limitations of the standard analytical methods, including the steps to be taken to minimize the various sources of bias and control study heterogeneity. It is essential that the approach be integrated in a more holistic strategy that includes the presentation of results in light of the totality of available data and the crafting of a communication scheme that balances the need for transparency with scientific rigor. Such an integrated approach would ensure the edifying dissemination of findings to patients, health-care providers and regulatory agencies, and a more effective protection of public safety and promotion of medical science.

References

Bate, A. Lindquist, M., Edwards, I.R. and Orre, R. (2002). A data mining approach for signal detection and analysis. *Drug Safety* **25**, 393-397.

Berkey, C.S., Mosteller, F., Lau, J. and Antman, E.M. (1996). Uncertainty of the time of first significance in random effects cumulative meta-analysis. *Controlled Clinical Trials* **17**, 357-371.

Bradburn, M.J., Deeks, J.J., Berlin J.A. and Localio, A.R. (2007). Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* **26**, 53-77.

Dahabreh, I.J. (2008). Meta-analysis of rare events: an update and sensitivity analysis of cardiovascular events in randomized trials of rosiglitazone. *Clinical Trials* **5**, 116-120.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177-188.

Drazen, J.M., Morrissey, S. and Curfman, G.D. (2007). Rosiglitazone–continued uncertainty about safety. *The New England Journal of Medicine* **357**, 63-64.

DuMouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *American Statistician* **53**, 177-202.

Evans, S.J. (2000). Pharmacovigilance: A science or fielding emergencies? *Statistics in Medicine* **19**, 199-209.

Huedo-Medina, T.B., Sánchez-Meca, J., Marín-Martínez, F. and Botella, J. (2006). Assessing heterogeneity in meta-aanalysis: Q statistic or I^2 index? *Psychological Methods* **11**, 193-206.

Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R. and Jones, D.R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* **24**, 2401-2428.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institution* **22**, 719-748.

McIntosh, M.W. (1996). The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine* **15**, 1713-1728.

Nissen, S.E. and Wolski, K. (2007). Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes. *The New England Journal of Medicine* **356**, 2457-2471.

Schmidt, F.L., Oh, I.S. and Hayes, T.L. (2009) Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology* **62**, 97-128.

Sterne, J.A.C., Gavaghan, D. and Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in literature. *Journal of Clinical Epidemiology* **53**, 1119-1129.

Sterne, J.A.C. and Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology* **54**, 1046-1055.

Sterne, J.A.C., Egger, M. and Davey-Smith, G. (2001). Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal* **323**, 101-105.

Sutton, A, J. and Higgins, J.P.T. (2008). Recent developments in meta-analysis. *Statistics in Medicine* **27**, 625-650.

Tian, L., Scottb, D.J., Biggerstaff, B.J. and Mengersenc K.L. (2009). Exact and efficient inference procedure for meta-analysis and its application to

the analysis of independent 2 x 2 tables with all available data but without artificial continuity correction. *Biostatistics* **10**, 275-281.

Tweedie, R.L., Scott, D.J., Biggerstaff, B.J. and Mengersen, K.L. (1996). Bayesian meta-analysis, with application to studies of ETS and lung cancer. *Lung Cancer* **14**(suppl. 1), S171-194.

Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985). Beta blockade during and after myocardial infarction. An overview of randomized trials. *Progress in Cardiovascular Disease* **27**, 335-371.

Ziegler, S., Koch, A. and Victor, N. (2001). Deficits and remedy of the standard random effects methods in meta-analysis. *Methods of Information in Medicine* **40**, 148-155.

Received June 2, 2010; accepted July 31, 2010.

Demissie Alemayehu
Pfizer, Inc.
235 East 42nd Street, Pfizer, Inc., New York, NY 10017
alemad@pfizer.com