Citation:

Koch, A. & Röhmel, J. (2003). Meta-analysis of randomized clinical trials in the evaluation of medical treatments – a partly regulatory perspective. In R. Schulze, H. Holling & D. Böhning (Eds.), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 99-112). Hogrefe & Huber.

7

Meta-Analysis of Randomized Clinical Trials in the Evaluation of Medical Treatments – A Partly Regulatory Perspective

Armin Koch
Joachim Röhmel

Federal Institute for Drugs and Medical Devices
Berlin

Summary

Meta-analysis of randomized clinical trials plays an important role in summarizing available evidence with respect to the comparison of different drugs for the same indication. In contrast, up to now meta-analysis is of only minor importance in the process of new drug application despite the fact that also in this situation a summary evaluation of available evidence from a, although limited, number of independent clinical trials is necessary. The main reason is, in our opinion, that presented meta-analyses often are not completely convincing because objectives are not appropriately chosen and conduct or presentation are not sufficiently detailed so that the reader can assess provided evidence. This chapter is intended to clarify why some meta-analyses have higher credibility than others and provide some guidance to how the credibility of meta-analyses can be increased. Presented ideas are, hopefully, not only in the regulatory setting of importance.

7.1 INTRODUCTION

Meta-analysis has been defined to be a quantitative and systematic summary of a collection of separate studies for the purpose of obtaining information that can not be derived from any of the studies alone (Boissel et al., 1988). With this definition, meta-analysis implicitly is also a technique that should lead to reproducible results and that can be distinguished from the classical review or overview, where results from various studies might be collected and qualitatively weighted by an expert in the field.

Originally invented in the social sciences, meta-analysis has found widespread use in clinical research during the last two decades and the per-year number of published meta-analysis is still increasing. However, only in rare cases has the discussion about the appropriateness of biostatistical methodology in medical research been as intensive as was the case with meta-analysis. From the very beginning meta-analysis has split up the community into clear proponents and those who completely dislike this type of analysis.

Feinstein (1995) named meta-analysis a synonym for "statistical alchemy for the 21st century", and others expressed their doubts on the credibility of results "proven" by means of meta-analysis. It has repeatedly been emphasized that pivotal trials should be designed to stand on their own and that in consequence meta-analyses should not be necessary ("If a treatment has an effect so recondite and obscure as to require meta-analysis to establish it I would not be happy to have it used on me" (Eysenck, 1994, p. 792)). And also empirical comparisons of the results from meta-analyses with results from large randomized clinical trials (Villar, Carroli, & Belizan, 1995) or critical expert reading of meta-analyses do not support the hypothesis that a meta-analysis can replace randomized clinical trials ("In my own review of selected meta-analyses, problems were so frequent and so serious, including bias on part of the meta-analyst, that it was difficult to trust in the overall 'best estimates' that the method often produces" (Bailar, 1997, p. 560)).

A positive view on meta-analysis is best summarized by a citation from a recent paper by Resch (1996), who wrote:

> I disagree, however, that a meta-analysis should exclusively be viewed as "hypothesis generating". This proposal denies the fact that, however biased, a high-quality meta-analysis quantitatively summarizes the existing evidence. What could be a better basis for a clinician's treatment decision at the time it must be made? (p. 621)

Meta-analyses – being retrospective and non-experimental investigations – are in a strict sense observational studies (Victor, 1995). Comparing, however, the evidence gained from a prospective observational study and a meta-analysis based on randomized clinical trials, clarifies that this can not be the whole truth, as the latter are based on what has often been termed "best available evidence", and at least on the study level distribution of covariates is controlled.

Approval of a new drug by national or European agencies is one of the most sensitive areas of evaluation of knowledge provided by clinical trials: If in a certain indication there is not yet a standard treatment, the approval of a new drug defines this standard and all future developments will be validated against this standard. In addition, an acceptable benefit/risk-ratio is in general taken for granted whenever a new drug is licensed.

International guidelines on statistical principles in clinical trials (ICH Topic E9) have made a valuable contribution to clarifying methodological principles for clinical trials in the regulatory setting. In this document reference is made to the use of meta-analysis or pooled analyses in general, and subsequently various meta-analyses have been presented in new drug applications. Not in all cases have these analyses been appropriate from a regulatory viewpoint, and the need for further clarification became obvious. This is also reflected in the current attempt to harmonize the opinions of the European regulatory authorities in a Points to Consider document, which is intended to provide better guidance for the pharmaceutical industry.

This chapter is not intended to summarize the discussion on the development of the Points to Consider document. Instead, ICH-E9 statements on meta-analysis are briefly reviewed and given recommendations are summarized. Instances are mentioned where these recommendations might need additional clarification or give the impression that the view on meta-analysis in the regulatory setting is very narrow. From a scientific viewpoint arguments why credibility of meta-analysis is in some instances greater than in others are given, and factors that can influence credibility are named. Some sample situations are discussed for illustration. A more appropriate use of meta-analysis will hopefully help the technique find greater acceptance in the regulatory setting.

7.2 QUOTES FROM "THE GUIDELINE"

In chapters II (Considerations for overall clinical development) and chapter VII (Reporting) direct or indirect reference is made to techniques of summarizing results from more than one clinical trial. The exact wordings are:

> Interpretation and assessment of the evidence from the total programme of trials involves synthesis of the evidence from the individual trials. This is facilitated by ensuring that common standards are adopted for a number of features of the trials such as dictionaries of medical terms, definition and timing of the main measurements, handling of protocol deviations and so on. A statistical summary, overview or meta-analysis may be informative when medical decisions are addressed in more than one trial. Where possible this should be envisaged in the plan so that the relevant trials are clearly identified and any necessary common features of their designs are specified in advance. Other major statistical issues (if any) that are expected to affect a number of trials in a common plan should be addressed in the plan. (Section 2.1.1: Development plan)

An overall summary and synthesis of the evidence on safety and efficacy from all the reported clinical trials is required for a marketing application [...]. This may be accompanied, when appropriate, by a statistical combination of results. [...] addressing the key questions of efficacy by considering the results of the relevant (usually controlled) trials and highlighting the degree to which they reinforce or contradict each other; [...]. During the design of a clinical programme careful attention should be paid to the uniform definition and collection of measurements which will facilitate subsequent interpretation of the series of trials, particularly if they are likely to be combined across trials. A common dictionary for recording the details of medication, medical history and adverse events should be selected and used. A common definition of the primary and secondary variables is nearly always worthwhile, and essential for meta-analysis. The manner of measuring key efficacy variables, the timing of assessments relative to randomization/entry, the handling of protocol violators and deviators and perhaps the definition of prognostic factors, should all be kept compatible unless there are valid reasons not to do so. Any statistical procedures used to combine data across trials should be described in detail. Attention should be paid to the possibility of bias associated with the selection of trials, to the homogeneity of their results, and to the proper modeling of the various sources of variation. The sensitivity of conclusions to the assumptions and selections made should be explored. (Section 7.2: Summarizing the clinical database)

Individual trials should always be large enough to satisfy their objectives. Additional valuable information may also be gained by summarizing a series of clinical trials which address essentially identical key efficacy questions. The main results of such a set of trials should be presented in an identical form to permit comparisons, usually in tables or graphs, which focus on estimates plus confidence limits. The use of meta-analytic techniques to combine these estimates is often a useful addition, because it allows a more precise overall estimate of the size of the treatment effects to be generated, and provides a complete and concise summary of the results of the trials. Under exceptional circumstances a meta analytic approach may also be the most appropriate way, or the only way, of providing sufficient overall evidence of efficacy via an overall hypothesis test. When used for this purpose the meta-analysis should have its own prospectively written protocol. (Section 7.2.1: Efficacy data)

In summarizing safety data it is important to examine the safety database thoroughly for any indications of potential toxicity, and to follow up any indications for an associated and supportive pattern of observations. The combination of the safety data from all human exposure to the drug provides an important source of information, because its larger sample size provides the best chance of detecting the rarer adverse events and, perhaps of estimating their approximate incidence. [...] The results from trials which use a common comparator (placebo or specific active comparator) should be combined and presented separately for each comparator providing sufficient data. (Section 7.2.2: Safety data)

In summary, ICH-E9 makes the following proposals and recommendations on the use of meta-analysis:

1. Objectives for meta-analyses might be the gain of more precise overall estimates of the size of the treatment effect, the gain of a complete and concise summary of trial results, and the assessment of consistency of results across trials.
2. Meta-analysis should be prospectively planned with the clinical trials programme in the development of a new treatment. This is extremely important if the meta-analysis is the most appropriate or the only way to provide sufficient evidence of efficacy.
3. Sensitivity analyses are necessary if assumptions or selections are necessary to justify the combination of study results.
4. Safety-results from trials where the same active treatment has been compared with different active substances or placebo should not be combined into an overall estimate of "a treatment effect". In general, separate analyses should be presented for different comparators.

Some other recommendations might be seen too narrow or even contradictory:

1. Studies should address essentially identical efficacy questions (not necessary: The only question is whether the design of the studies and the query for variables, relevant for the meta-analysis, is sufficiently similar)
2. Presentation of results of the single trials should be done by means of estimates and confidence intervals (better: In general, the number of successes and events per treatment group for dichotomous endpoints or sufficient statistics in general should be provided in order to give the reader the option to make his own mind).
3. "Studies should stand on their own" and "studies should address essentially identical key efficacy questions" (little dissent exists on how to interpret the results of meta-analysis in this situation).

And in some instances further clarification is needed:

1. No guidance is given with regard to the exceptional circumstances, for which a proof of efficacy by means of a meta-analysis might be acceptable.
2. Supposed it is accepted that a meta-analysis can increase the precision of an estimate for the treatment effect in a certain situation, then meta-analysis can in fact be confirmatory.
3. The term "prospectively" needs clarification: Should the meta-analysis be planned before the first study that is intended to be included in the meta-analysis is undertaken, or is it sufficient to plan the meta-analysis before its conduct.

7.3 HOW CAN THE CREDIBILITY OF META-ANALYSIS BE INCREASED?

Meta-analyses are non-experimental studies. As a consequence, the classical framework of error probabilities is not applicable for decisions based on p-values or confidence intervals that have been computed in a meta-analysis of results from independent trials. Like in observational studies in general, p-values are primarily a measure for the distance between two success-rates or between two means computed in two different groups in relation to the respective variance. And like in observational studies, too, the author of a meta-analysis must justify his belief that observed differences between two groups, defined by means of the absence or presence of a certain treatment, are in fact due to differences between the two treatments and not a consequence of bias (e.g., due to selection of trials (i.e., publication bias), patients and interventions, or the statistical methodology for the combination of the results from independent trials).

Results and conclusions from meta-analysis are thus more or less credible and this credibility – in contrast to randomized clinical trials – does not only depend on design issues but also on sound argumentation on the absence of bias. The following sections name the factors that influence to our opinion this credibility and name prerequisites for meta-analysis, performed "in an almost confirmatory way". A number of situations are proposed where meta-analysis therefore can contribute valuable information for the decision on licensing of new drugs.

7.3.1 The Aspect of Objectives for Meta-Analyses

For a long time, meta-analyses have had little impact on decisions made by regulatory authorities. This was mainly due to the fact that meta-analyses have been presented almost exclusively in situations where proof of efficacy in two independent clinical trials deemed necessary in the beginning, and this attempt failed in the end (i.e., one significant and one insignificant trial; two only borderline significant results etc.). Whenever meta-analyses are misused to counterbalance for shortcomings in the respective primary research (i.e., the studies to be included in the meta-analysis), credibility is affected: The only aim of this type of meta-analysis is the demonstration of a "significant" treatment effect. Chances are not too bad that at least this aim is reached due to the mere fact that sample size is increased. This unpleasant situation has postponed considerations on good objectives for meta-analyses in the regulatory setting:

1. Substantiation of additional claims on secondary endpoints, especially in a situation where the primary endpoint is based on a surrogate-variable, or on components of multiple endpoints:

   Various situations exist, where in a collection of clinically important variables the choice of the primary endpoint is not only driven by the attempt

to select the most important one, but also by feasibility considerations. This is especially true if the incidence of some of the endpoints is higher and some other endpoints are rare. An example is provided by studies investigating postoperative prophylaxis against thromboembolic complications, where the incidence of deep vein thromboses is usually selected as the primary endpoint. The rare event of a pulmonary embolism is, however, an at least equally important endpoint. The demonstration of equivalence or superiority or equivalence with respect to the rate of the rare event would demand larger clinical trials. A good basis for a claim with respect to the rare endpoint might be a meta-analysis of all pivotal trials.

Similarly, a double endpoint (death or reinfarction) is selected in studies in the treatment of acute myocardial infarction. An additional claim "the experimental treatment reduces the rate of death" may be substantiated by means of a meta-analysis of all pivotal clinical trials.

2. Proof and investigation of efficacy and safety in subgroups of the patient population:

   As soon as the global superiority of an experimental treatment over control has been established by means of evidence from more than one pivotal clinical trial, investigations into subgroups of the patient population might help to understand better for which patients the benefit of the experimental treatment is greatest (e.g., to justify higher costs of the experimental treatment). Similarly, one might be conscious whether a consistent safety profile exists across subpopulations. In both instances a meta-analysis can be helpful to provide evidence based on an acceptable sample size.

3. Proof of efficacy in situations where single studies are contradictory or inconclusive:

   Despite the fact that meta-analysis should not be used to compensate for shortcomings in the pivotal trials, meta-analysis can be helpful to come to a decision if results of clinical trials are not homogeneous. In this situation meta-analysis must be understood as a tool to demonstrate robustness of results and conclusions by means of sensitivity analyses. In general, the same methods should be applied that were proposed for the investigation of heterogeneity in multicenter clinical trials. It should be noted that, due to currently existing limitations with respect to statistical methodology, two studies can not be regarded sufficient to assess an overall "impression" by means of a meta-analysis. This is due to the fact that the likelihood to detect important differences between studies with respect to the treatment effect is small.

4. Evaluation of signals for serious adverse events of treatment:

   In many situations a meta-analysis of "all randomized evidence" will be the only chance to detect early whether a risk for serious adverse events is associated with the experimental treatment at the time a decision on marketing of the new drug has to be made. Whereas in early years simply

a summary of the combined database was provided, during the last years it has been recognized that formal methods for the combination of results from independent studies should also be used in this situation to avoid bias.

7.3.2 The Aspect of Planning

Planning of experimental investigations in humans or animals is mandatory by law. As has been pointed out in the discussion on the need for randomization in clinical trials, the object of trials is both to ensure a high probability of identifying the better treatment (if there is one) and to convince others of the validity of the conclusions (Byar et al., 1976).

In consequence, all scientific investigations and especially observational research should be planned. The crucial point in the context of meta-analysis is that the need for a meta-analysis can become obvious at various points during the conduct of a clinical programme consisting of more than one clinical trial.

It is, however, not only a question of credibility of results whether such a meta-analysis has been planned together with the whole clinical programme or after completion of the most recent study in the program: Credibility is affected if the applicant can not assure that presented conclusions are not driven by observed results and ruling this out is obviously easier in the first case.

One might believe that the credibility of a meta-analysis planned after the completion of the last clinical trial might be increased if the meta-analysis has been performed by a site that is independent or quasi-independent from the sponsor. Even in this situation it might be difficult to assure that results of the meta-analysis have not been known to the sponsor before the "independent" re-analysis has been performed.

When planned in the beginning of the clinical program, the additional opportunity exists to care for consistency in the conduct of the clinical trials that are intended to be included into the meta-analysis. This reduces the need to make assumptions on what can be safely combined (e.g., a study lasting three weeks and another study lasting four weeks, studies where variables have been transformed differently in different studies, or studies with slightly different questionnaires that might affect clinicians behavior with respect to answering), reduces potential sources of heterogeneity, and thus also improves the quality of the meta-analysis.

A meta-analysis planned before the inception of the last study in a clinical program ranges in between the before mentioned extremes: The conformity of the study plans can at this point hardly be influenced. However, which endpoints and which analysis are presented in the end can not be completely derived from the observed results but at least some sort of internal validation is available. This is the reason why one positive meta-analysis and a subsequently initiated clinical study with positive results may constitute a sufficient basis for a licence application (see Example 2 in Section 7.4).

A minimal requirement for meta-analysis in the regulatory setting is that the meta-analysis has been planned in advance to its conduct. In this situation it is, of course, difficult to demonstrate that results have not been available at the point in time, where the plan for the meta-analysis has been presented.

7.3.3 The Aspect of Conduct

Due to unfavorable experiences with publication based meta-analysis recently, meta-analyses based on individual patient data have been recommended and termed the current gold standard in meta-analysis (Clarke & Stewart, 1994). To our present opinion and experiences this might be somewhat too restrictive. Obviously, more questions can be addressed in a meta-analysis based on individual patient data, as the full information on covariables is available. It is also true that more insight into the data at hand is needed to perform this type of meta-analysis. Keeping in mind, however, the enormous workload that is needed to perform a re-analysis of the individual trials, one should also keep in mind that the method which is used for combination of study results should be justified by the question that is to be answered (e.g., in case subgroup analyses are of interest and the respective information is not available from the study report, a re-analysis can not be avoided).

At least in the regulatory framework and with respect to the primary and secondary end-points of pivotal trials, in contrast, it would shade suspicion on the original trial report if results of meta-analyses based on individual patient data and meta-analyses based on published data from the original report would come to different conclusions. Re-definitions of success and treatment failure, in addition, might raise suspicion that again attempts are made to fish for significance (i.e., why should definitions that seemed reasonable at the time when the individual trials had been planned now be obsolete?). Obviously, the plan to perform a meta-analysis based on original patient data can not be the justification for the exclusion of trials from the analysis where original patient data are not available, although it is expected that this problem (like publication bias more generally) is of minor importance in the regulatory setting.

7.3.4 The Aspect of Analysis and Presentation of Results

During the first years the term meta-analysis has been associated in medicine almost completely with the aspect of summarizing the evidence from independent clinical trials. Summary estimates and overall tests of effect or confidence intervals have been presented exclusively. Due to bad experiences the view on meta-analysis is nowadays more differentiated and meta-analysis is more understood as a tool for investigating similarities and dissimilarities between trials that should, at least in principle, be combinable. Unfortunately, in the regulatory setting the mere provision of a summary p-value is still the rule and not the exception. As the evaluation of consistency of results across trials is very important for claims on efficacy, this is not acceptable.

Statistical information on similarities and dissimilarities of study results are important. Critics might say that the currently used tests for homogeneity that are based on weighted squared differences between estimates from the single trials and the meta-analysis estimate have insufficient power to detect departures from the null-hypothesis (Jones, O'Gorman, Lemke, & Woolson, 1989). Critics might further object that they are in addition usually used "to proof their null-hypothesis". As a consequence, it should be not acceptable to conclude that no heterogeneity exists, unless the test for homogeneity rejects the null-hypothesis at a conventional 5%-level. This is true but should, to our opinion, not prevent from making all attempts to use the test as a diagnostic tool (as a well known statistician has pointed out: Statistical methods need not be perfect, it is sufficient if they are better). The following example might support this opinion:

Two randomized double blind placebo controlled studies have been undertaken to investigate the efficacy of omeprazole in functional dyspepsia (Bond study and Opera study) (Talley et al., 1998). Both studies are three armstudies comparing two dosage regimens to placebo. Results of the comparison of the higher dose and placebo are reported here in a slightly simplified way not to discredit a potentially efficacious treatment but to construct an instructive example for decision making. Conclusions of the authors are: Omeprazole is modestly superior to placebo in functional dyspepsia. On an intention to treat analysis ($n = 1248$), complete symptom relief was observed in 38% on omeprazole 20mg compared to 28% on placebo ($p = .002$).

Table 7.1 Complete Relief of Dyspeptic Symptoms

Study	Omeprazole 20mg Relief / Treated	Placebo Relief / Treated
Bond	93 / 219	57 / 219
Opera	68 / 202	62 / 203

Results of the two studies are summarized in Table 7.1. The paper reports results for the combination of the two trials only. The meta-analysis estimate for the difference of the relief rates in the two treatment groups is 10% with a 95% confidence interval ranging from 3.7 to 16.3%, and from this a modest superiority of the experimental treatment over placebo is concluded.

Our confidence in the drawn conclusion might change if it was clearly stated, that first, it is only the Bond-study that came up with a significant treatment effect ($p = .001$), the p-value for the treatment effect in the Opera study was $p = .501$, and that second, despite the fact that tests for heterogeneity are blamed for being insufficient, a clear warning might have been achieved ($p = .039$ for heterogeneity).

This should be general guidance for analysis and presentation of results: Confinement to only meta-analytic results in terms of a summary estimate of the treatment effect and the respective confidence interval or a simple hypoth-

esis test resulting in just one p-value is not appropriate in the setting of observational studies. As has been pointed out before, argumentation is necessary with observational studies. Analysis and presentation of results should always emphasize the need to also clarify the contribution of a single trial to the combined result. Respective recommendations date back to 1993 (Thompson, 1993): For every study the relative weight in a fixed effects model should be presented together with the contribution of the study to the statistics of the heterogeneity test. The first information gives the reader an impression on whether meta-analysis can add useful information to the knowledge from the larger studies (e.g., in a situation with one large trial and two small studies given weights 80%, 10%, 10%, it is very unlikely that the combined analysis will add new information, as the size of the estimate is completely driven by the result of the large trial). The second information can descriptively be assessed with a rule of thumb, comparing each of the contributions to the heterogeneity statistics with the critical value of a χ^2-distribution with one degree of freedom and deciding whether results are homogeneous or whether some extreme results might drive the overall impression.

7.4 SAMPLE SITUATIONS

A series of sample situations of appropriate or inappropriate use of meta-analysis, all motivated by recent applications or collected from the literature, are presented to illustrate the considerations above.

Example 1: Meta-Analysis and Borderline Significant Pivotal Studies

In a situation where proof of superiority has to be based on two separate pivotal trials, both demonstrating that the experimental treatment is superior to control, both studies ended up with only borderline significant results (e.g., a p-value between 5% and 10% was achieved, where the level of significance was initially set to 5%). A combined re-analysis of the two pivotal trials demonstrates "significant" superiority of the experimental treatment over control for the primary endpoint of the pivotal trials.

Even if the meta-analysis has been planned before its conduct, a meta-analysis in this situation is not acceptable. Meta-analysis should not be used as a safety-belt against non-significant results from pivotal trials that were planned to stand on their own.

In a very dialectic discussion on meta-analysis Senn (1997) argued that clinical trials are notoriously too small and that even small true effects might be of enormous public health importance (for example, in the treatment of cancer or myocardial infarction) and that many small drops can make a hole into stone. Nevertheless, he admits that a difference to the drug development setting exists, where the way how experimentation is performed is under the control of

the pharmaceutical company that must demonstrate efficacy beyond reasonable doubt.

If studies that have been planned to stand on their own fail to proof efficacy as expected, this is a strong indication that something substantial went wrong with the original plan.

Example 2: Meta-Analysis and the Need for Replication of Results

Imagine a situation where the sponsor has decided to proof efficacy in a series of three phase III pivotal clinical trials. After a first "significant" trial the sponsor decides to make some minor modifications to the study design (e.g., small changes in the criteria for inclusion or exclusion of patients from the trial or a modification of the primary variable). His intention is to demonstrate even better the superiority of the experimental treatment over control.

Unfortunately, this second and a third attempt with again minor modifications both fail to demonstrate superiority of the experimental treatment over control. However, a meta-analysis including the first trial and "similarly defined subgroups" of the following two trials demonstrates a significant superiority of the experimental treatment over control. Again this is, in our opinion, no appropriate use of meta-analysis. If a need for replication of scientific results exists, this need can not be substituted by a (retrospective) subgroup (or a combination of subgroups) analysis. The sponsor still needs to demonstrate in an independent study that he now can correctly identify those patients that will benefit more from the experimental treatment than from the control treatment.

In consequence, the reversed situation might well be acceptable: Based on two (or more) non-significant trials the sponsor now believes that he can identify the patient population that will benefit from the experimental treatment. A meta-analysis of the respective subgroups of the first two trial populations demonstrates "significant" superiority of the experimental over the control treatment. A new trial is planned according to these restrictions and can verify the result of the meta-analysis. Meta-analysis should thus only be used to replace the first experiment, not the verification step.

The need for independent verification of scientific results has been discussed controversially in the literature (Högel & Gaus, 1999), and it has been even questioned whether the usual procedure of just performing two trials at the same time in different geographical hemispheres reflects a true verification of results. It should be pointed out that in this discussion the question is not whether verification is necessary or not, but that discrepancies between expectation and results need further investigation.

It should be noted that this is also one solution for the problem posed in Example 1: The meta-analysis of the two borderline significant results might, given that no other problems with study design and conduct exist, be accepted as a first pivotal trial. A third study should, however, be planned that can then successfully reproduce this first result.

Example 3: Meta-Analysis and Claims for Secondary Endpoints

Very often in clinical trials the selection of one variable as primary endpoint from a series of others, which then are termed secondary, is not only influenced by clinical importance of the various variables but reflects also considerations on feasibility (i.e., if a more important endpoint (e.g., pulmonary embolism in thrombosis prophylaxis) is a very rare event, chances for demonstrating superiority increase if an endpoint with higher incidence (e.g., deep vein thrombosis) is selected instead). A whole clinical trials program, however, might be designed such that also differences in mortality can be detected. The sponsor might decide to use the more frequent endpoint as primary, however, to plan a meta-analysis of pivotal trials in order to demonstrate superiority with respect to the less frequent event.

Given appropriate results, this is an acceptable prerequisite for an additional claim regarding the secondary, less frequent but potentially clinically more relevant endpoint.

7.5 CONCLUSIONS

Meta-analysis, even if restricted to the combination of only two pivotal trials, might be an extremely helpful tool for decision making in the regulatory setting. This technique could in principle support the task of the medical reviewer, who, at the end of the day, must integrate all the presented knowledge and come to a final decision. Meta-analysis is not playing this role up to now. This is mainly due to the fact that presented meta-analyses fall short with respect to the addressed objective, the conduct and the presentation of results: Still too much emphasis is given to combined estimates of the treatment effect and summary p-values. The potential of meta-analysis to show similarities and dissimilarities between the trials that are to be combined has not been used too often.

Meta-analysis, sometimes routinely presented as part of the clinical expert report, often fall short with respect to the presentation of results: Again, only summary information is presented and the reviewer is referred to the single documentation of clinical trials if he is interested in the consistency with respect to certain information.

Meta-analysis "is here to stay", however, careful consideration of the above mentioned points will help to sharpen the general understanding, where meta-analysis can be helpful and where not, will help to bring up better results and lastly help to find the place for meta-analysis that it should have.

REFERENCES

Bailar III, J. C. (1997). Editorial: The promise and problems of meta-analysis. *New England Journal of Medicine*, *337*, 559–561.

Boissel, J. P., Sacks, H. S., Leizorovicz, A., Blanchard, J., Panak, E., & Peyrieux, J. C. (1988). Meta-analysis of clinical trials: Summary of an international conference. *European Journal of Clinical Pharmacology*, *34*, 535–538.

Byar, D. P., Simon, R. M., Friedewald, W. T., Schlesselmann, J. J., DeMets, D. L., Ellenberg, J. H., Gail, M. H., & Ware, J. H. (1976). Randomized clinical trials. *New England Journal of Medicine*, *295*, 74–80.

Clarke, M. J., & Stewart, L. A. (1994). Obtaining data from randomized controlled trails: How much do we need for reliable and informative meta-analysis? *British Medical Journal*, *309*, 1007–1010.

Eysenck, H. J. (1994). Meta-analysis and its problems. *British Medical Journal*, *309*, 789–792.

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, *48*, 71–79.

Högel, J., & Gaus, W. (1999). The procedure of new drug application and the philosophy of critical rationalism or the limits of quality assurance with good clinical practice. *Controlled Clinical Trials*, *20*, 511–518.

Jones, M. P., O'Gorman, T. W., Lemke, J. H., & Woolson, R. F. (1989). A Monte Carlo investigation of homogeneity tests of the odds ratio under various sample size configurations. *Biometrics*, *45*, 171–181.

Resch, K. L. (1996). Letter: Megatrials for clinical decision making. *Annals of Internal Medicine*, *125*, 621–622.

Senn, S. (1997). *Statistical issues in drug development.* New York: Wiley.

Talley, N. J., Meineche-Schmidt, V., Pare, P., Duckworth, M., Raisanen, P., Pap, A., Kordecki, H., & Schmid, U. (1998). Efficacy of omeprazole in functional dyspepsia: Double-blind, randomized, placebo-controlled trials (the Bond and Opera studies). *Alimentary Pharmacology and Therapy*, *12*, 1055–1065.

Thompson, S. G. (1993). Controversies in meta-analysis: The case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research*, *2*, 173–192.

Victor, N. (1995). "The challenge of meta-analysis": Discussion. Indications and contraindications for meta-analysis. *Journal of Clinical Epidemiology*, *48*, 5–8.

Villar, J., Carroli, G., & Belizan, J. M. (1995). Predictive ability of meta-analysis of randomized controlled trials. *The Lancet*, *345*, 772–776.