The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 31 May 2001
CPMP/EWP/2330/99

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP)

POINTS TO CONSIDER ON APPLICATION WITH 1. META-ANALYSES; 2. ONE PIVOTAL STUDY

DISCUSSION IN THE EFFICACY WORKING PARTY	September 1999/ September 2000
TRANSMISSION TO CPMP	October 2000
RELEASE FOR CONSULTATION	October 2000
DEADLINE FOR COMMENTS	January 2001
SUBMISSION TO THE EWP	February 2001
TRANSMISSION TO CPMP	May 2001
ADOPTION BY CPMP	May 2001

7 Westferry Circus, Canary Wharf, London, E14 4HB, UK
Tel. (44-20) 74 18 84 00 Fax (44-20) 74 18 86 13
E-mail: mail@emea.eudra.org http://www.emea.eu.int/
EMEA 2001 Reproduction and/or distribution of this document is authorised for non commercial purposes only provided the EMEA is acknowledged

POINTS TO CONSIDER ON APPLICATIONS WITH 1. META-ANALYSES; 2. ONE PIVOTAL STUDY

I INTRODUCTION

This PtC provides guidance on two topics. The first of these is meta-analysis. Specifically, the PtC considers the validity and interpretation of this methodology in applications for the marketing authorisation of medicinal products. The second is the use of only a single pivotal study in Phase III clinical development. There are some connections between these two topics that will be indicated in the discussion, but to a certain extent their inclusion in the same PtC is a matter of convenience.

Individual clinical trials should always be designed to satisfy their objectives and the purpose of this document is not to encourage applicants to rely on a submission where stand alone studies have been substituted by a meta-analysis of trials of inadequate size. Nevertheless, there are applications where valuable information has been provided by a meta-analysis of several studies. In the biostatistical guidelines (ICH E9) meta-analytic techniques are recognised as a useful tool to summarise the overall efficacy results of a drug application (e.g. for the clinical expert report) and to analyse less frequent outcomes in the overall safety evaluation. However, when used for other purposes, there is disagreement about the validity and interpretation of meta-analyses. Clarification is needed regarding the selection of studies, the need for pre-specification of the meta-analysis, the outcome of individual studies (lack of significance) in relation to the pooled result, and the clinical relevance and external validity of the pooled result. Finally, it can be argued that a situation where a meta-analysis of all relevant studies is the main proof of efficacy is comparable to an application with only one pivotal study, which is the second topic of this document.

Although in most applications the proof of efficacy consists of several studies standing on their own, there are cases where the application has relied more or less entirely on the results from one study or a meta-analysis of several studies. Indeed, it is recognised in the guidelines (ICH E9) that there are circumstances when such applications can be sufficient for an approval. However, the guidelines are not specific about the circumstances when this applies.

The use of a meta-analysis to provide the pivotal evidence in an application will always be problematic. One really robust trial supported by smaller trials is stronger than either a meta-analysis of studies, none of which is convincing on its own or a meta-analysis of seemingly conflicting results. There are, however, a number of accepted regulatory purposes for meta-analysis. These include:

- To provide a more precise estimate of the overall treatment effects.
- To evaluate whether overall positive results are also seen in pre-specified subgroups of patients.
- To evaluate an additional efficacy outcome that requires more power than the individual trials can provide.
- To evaluate safety in a subgroup of patients, or a rare adverse event in all patients.
- To improve the estimation of the dose-response relationship.
- To evaluate apparently conflicting study results.

EMEA 2001

In accordance with ICH E9 meta-analysis is understood as a formal evaluation of the quantitative evidence from two or more trials bearing on the same question. This most commonly involves the statistical combination of summary statistics from various trials but the term sometimes is also used to refer to the combination of raw data. This PtC document does not deal with a crucial problem for meta-analysis in general, the identification and validation of all relevant studies. It only discusses the problems encountered when meta-analytic techniques are applied to the studies included in a drug application. Nor does it discuss statistical models and methods for carrying out the meta-analyses. It is assumed that adequate methods are used for the estimation of the overall effect as well as for the evaluation of heterogeneity. In this document heterogeneity refers to study-by-treatment interaction.

II META-ANALYSES

II.1 Pre-specification

II.1.1 The meta-analysis protocol

When a meta-analysis is included in an application it should be performed in accordance with a protocol specifying

- The objective of the analysis
- Criteria for inclusion and exclusion of studies, e.g. study populations, study design (the norm is to accept only randomised and preferably blinded studies), dosage, study duration etc.
- In cases when the meta-analysis includes studies not performed by the sponsor, the strategy for the identification of studies should be specified.
- Hypotheses and endpoints.
- Statistical methods (including methods for investigation of heterogeneity). In cases where the meta-analysis provides the pivotal evidence for an indication, a p-value more extreme than the conventional significance level of 0.05 would generally be required, and narrow confidence interval for the treatment effect would be expected. The required degree of significance will be judged on a case by case basis considering factors such as the amount of supportive data, plausibility of the hypothesis tested and whether the analysis is pre-specified or not. Where applicable, the methods for handling studies with different designs, e.g. parallel groups and cross-over designs, should be specified.
- Approaches to maximise the quality of the data in all the studies included and to explore the influence of any variation in quality, including methods to minimise or deal with missing data.
- Plan for evaluation of consistency and robustness. This should aim to demonstrate similar effects in alternative analyses of different endpoints, different sub-populations, different subsets of studies (studies from different regions, of different duration or different design, high-quality studies etc).

II.1.2 Timing of pre-specification

The crucial issue concerns the time at which the protocol for the meta-analysis should be specified in relation to the performance of the individual studies. A retrospective specification when the results from all or some of the studies are known should be avoided. Thus, when the need for a meta-analysis is prospectively identified, the protocol for the meta-analysis should be prospectively

specified during the planning of the clinical development program. Appropriate objectives are to obtain an overall estimate of efficacy and to provide a conclusive analysis of some less frequent secondary endpoint, e.g. in the overall safety evaluation.

There are, however situations where the need for a meta-analysis becomes apparent after the results from some or sometimes all studies are known. This is the case when there is a need to put seemingly conflicting results into perspective, or in the exceptional situation where a meta-analysis seems to be the only way to provide reliable proof of efficacy. To minimise the opportunity for a retrospectively specified meta-analysis protocol in such cases to be data dependent, the primary specifications and definitions set up in the individual studies should be followed. The credibility of the meta-analysis will depend on the degree of adherence to the specifications in the individual study protocols. In this situation, a protocol for the conduct of the meta-analysis should still be written.

II.1.3 Regulatory prerequisites of retrospective meta-analyses

Prerequisites for a retrospective meta-analysis to provide sufficient evidence for a claim include:

- Some studies clearly positive.
- Inconclusive studies showing positive trends in the primary variable.
- No statistically significant heterogeneity.
- Pooled 95% confidence interval well away from zero (or unity for odds ratios, or the pre-defined margin for non-inferiority trials).
- A justification that a biased selection of studies and/or endpoints is unlikely.
- A sensitivity analysis demonstrating robustness of the findings.

For meta-analyses where these requirements are not fulfilled it will prove difficult to get a regulatory acceptance. A retrospective meta-analysis of only two studies originally intended to stand on their own is not expected to add any useful information. In particular, a meta-analysis cannot be used to reconcile the conflicting results of one positive and one inconclusive study.

II.2 Selection of studies

II.2.1 Applications relying on fully reported trials

The applicant has the obligation to submit all studies that are relevant to the claims that are made. In some cases additional studies have been identified by the regulatory agencies, but for new chemical entities it should normally be possible to accept that all relevant studies have been identified in the application.

II.2.2 Bibliographical applications

When meta-analysis is used in applications relying wholly or partly on bibliographical data (e.g. new indications, orphan drugs etc) there is always a substantial risk for selection bias. In such applications the strategy for the identification of relevant studies should be presented together with a thorough discussion of a potential selection bias. In case a regulatory authority identifies additional relevant studies, this may seriously question the validity of the meta-analysis provided by the sponsor.

Even when all potentially relevant studies are identified one can always argue about the inclusion and exclusion of individual studies in the meta-analysis. However, excluded studies should be

listed and their results summarised as well as the reason for excluding them to enable an evaluation of any potential bias due to these exclusions. Furthermore, the robustness of the obtained result should be investigated in alternative analyses based on different subsets of the studies.

II.3 Clinical relevance, heterogeneity and external validity

II.3.1 Clinical relevance

If it is necessary to resort to a meta-analysis in order to reach an acceptable level of significance, this might indicate that there is reason to question the clinical relevance of the findings. It is certainly true that a positive effect established through a methodologically valid meta-analysis does not necessarily mean that a drug application is approvable. The clinical relevance of the findings will be judged on clinical grounds. The magnitude of the effect, and range of effects of different endpoints, might very well be too small or limited to result in a positive overall benefit/risk assessment, and this is more likely to be true when it has required a meta-analysis to establish statistical significance.

II.3.2 Heterogeneity and external validity

The external validity of the average result from a meta-analysis of several studies with more or less inconsistent results has to be judged on a case by case basis considering clinical as well as statistical aspects. From the clinical point of view, are the study populations in the different studies sufficiently representative of a common target population, and to what extent can dosage, treatment duration, sample size and other study conditions differ among studies and still make a common estimate meaningful? Some deviations in these respects are acceptable and might even increase the external validity of the pooled result. From the statistical point of view, are the variable estimates from the individual studies consistent with random fluctuation or is there a reason to suspect a qualitative study-by-treatment interaction? Tests for interaction are usually insensitive and therefore a critical evaluation of heterogeneity should involve graphical and other descriptive methods as well as tests for study-by-treatment interaction. Any suspected interaction, statistically significant or not, should be judged on its clinical relevance and the impact on the overall result should be evaluated in alternative analyses with the aim to demonstrate the robustness of the results (c.f. Plan for evaluation of consistency and robustness in Section II.1.1).

II.3.3 The meta-analysis report

To facilitate a critical evaluation of external validity and heterogeneity, the meta-analysis report should include sufficient data for the assessment of

- the comparability of the individual studies,
- the impact of each study on the overall estimate, and
- the impact of each study on the heterogeneity test,

together with a thorough discussion of the impact of the potential sources of bias outlined above.

II.4 Pooled analyses of safety endpoints

In the evaluation of safety of a new drug it is important to consider adverse events and laboratory data from all sources, dose-finding as well as confirmatory studies, short-term as well as long-term studies, and controlled as well as uncontrolled studies. In order to incorporate also less common and rare outcomes in the estimation of the overall safety profile it is usually necessary to rely on pooled analyses. However, due to the variable data sources and the large number of variables, a

stringent meta-analysis is in many cases not a feasible approach. A meta-analysis of all controlled trials may be valuable for an in-depth follow-up of safety signals from less reliable analyses. To facilitate a meta-analysis as well as a less sophisticated combining of data from different sources, it is recommended that a consistent methodology for the collection and evaluation of safety data is used throughout the clinical program. The possibility of any bias due to differential study factors should be discussed and particular consideration should be given to a potential dilution of negative effects due to the inclusion of studies of inadequately short duration. In comparative safety evaluations it should be a minimum requirement that pooled results are based on intra-study comparisons.

III ONE PIVOTAL STUDY

III.1 The need for replication

There is a general demand for replication of scientific results. This is especially important in strictly experimental studies where a successful result needs to be confirmed in a replicated experiment to achieve general acceptance. However, clinical drug development differs from the situation with strictly experimental studies. The purpose of phase III is to confirm the findings obtained so far in pre-clinical studies, tolerance studies, dose-finding and other phase II studies. The fundamental requirement on the phase III documentation is that it consists of adequate and well-controlled data of good quality from a sufficient number of patients, with a sufficient variety of symptoms and disease conditions, collected by a sufficient number of investigators, demonstrating a positive benefit/risk in the intended population at the intended dose and manner of use. The extent of confirmatory phase III data needed will depend upon what is established for the product in earlier phases, and what is known about related products. The minimum requirement is generally one controlled study with statistically compelling and clinically relevant results. However there are many reasons why it is usually prudent to plan for more than one study in the phase III program. These include:

- Lack of pharmacological rationale (unknown mechanism of action).
- A new pharmacological principle.
- Phase I and phase II data are limited or unconvincing.
- A therapeutic area with a history of failed studies or failures to confirm seemingly convincing results.
- A need to demonstrate efficacy and/or tolerability in different sub-populations, with different co-medication or other interventions, relative to different comparators, etc.
- Any other needs to address additional questions in the phase III program.

III.2 Prerequisites for one pivotal study applications

In cases where the confirmatory evidence is provided by one pivotal study only, this study will have to be exceptionally compelling, and in the regulatory evaluation special attention will be paid to:

- The internal validity. There should be no indications of a potential bias.
- The external validity. The study population should be suitable for extrapolation to the population to be treated.

EMEA 2001

- Clinical relevance. The estimated size of treatment benefit must be large enough to be clinically valuable.
- The degree of statistical significance. Statistical evidence considerably stronger than $p<0.05$ is usually required, accompanied by precise estimates of treatment effects, i.e. narrow confidence intervals. The required degree of significance will depend on factors such as the therapeutic indication, the primary endpoint, the amount of supportive data and whether the alternative analyses demonstrating consistency are pre-specified. When the aim is to demonstrate non-inferiority, one study is more likely to be accepted if the lower 95% confidence bound is well away from the non-inferiority margin.
- Data quality.
- Internal consistency. Similar effects demonstrated in different pre-specified sub-populations. All-important endpoints showing similar findings.
- Centre effects. None of the study centres should dominate the overall result, neither in terms of number of subjects nor in terms of magnitude of effect.
- The plausibility of the hypothesis tested.

To summarise, there is no formal requirement to include two or more pivotal studies in the phase III program. However, in most cases a program with several studies is the most, or perhaps only feasible way to provide the variety of data needed to confirm the usefulness of a product in the intended population. In the exceptional event of a submission with only one pivotal study, this has to be particularly compelling with respect to internal and external validity, clinical relevance, statistical significance, data quality, and internal consistency.