



European Medicines Agency

London, 27 July 2006  
Doc. Ref. CHMP/EWP/83561/2005

**COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE  
(CHMP)**

**GUIDELINE ON CLINICAL TRIALS IN SMALL POPULATIONS**

<b>DRAFT AGREED BY EFFICACY WORKING PARTY / AD HOC GROUP ON CLINICAL TRIALS IN SMALL POPULATIONS</b>	May 2002 – January 2005
<b>ADOPTION BY CHMP FOR RELEASE FOR CONSULTATION</b>	March 2005
<b>END OF CONSULTATION (DEADLINE FOR COMMENTS)</b>	September 2005
<b>AGREED BY EFFICACY WORKING PARTY</b>	July 2006
<b>ADOPTION BY CHMP</b>	27 July 2006
<b>DATE FOR COMING INTO EFFECT</b>	1 February 2007

# GUIDELINE ON CLINICAL TRIALS IN SMALL POPULATIONS

## TABLE OF CONTENTS

<b>1. INTRODUCTION.....</b>	<b>3</b>
<b>2. LEVELS OF EVIDENCE.....</b>	<b>4</b>
<b>3. PHARMACOLOGICAL CONSIDERATIONS.....</b>	<b>5</b>
<b>4. CHOICE OF ENDPOINTS.....</b>	<b>5</b>
<b>5. CHOICE OF CONTROL GROUPS .....</b>	<b>6</b>
<b>6. METHODOLOGICAL AND STATISTICAL CONSIDERATIONS.....</b>	<b>7</b>
<b>7. SUMMARY AND CONCLUSIONS.....</b>	<b>10</b>

## 1. INTRODUCTION

This Guideline considers problems associated with clinical trials when there are limited numbers of patients available to study. Many rare diseases affect only a few thousand or even fewer than one hundred patients in the EU. Under such circumstances, a trial enrolling several hundred patients may not be practical or possible. Accordingly, conduct, analysis, and interpretation of studies in rare conditions at times may be constrained to varying degrees by the prevalence of the disease. Similarly, in paediatric trials it may be difficult to recruit large numbers of patients.

The paper has been prepared by the CHMP Efficacy Working Party (EWP) in joint collaboration with members of the Scientific Advice Working Party (SAWP), the Committee on Orphan Medicinal Products (COMP) and the Paediatric Expert Group (PEG). The expertise within the group includes clinicians, epidemiologists and statisticians from National Regulatory Authorities and from universities.

No methods exist that are relevant to small studies that are not also applicable to large studies. However, it may be that in conditions with small and very small populations, less conventional and/or less commonly seen methodological approaches may be acceptable if they help to improve the interpretability of the study results. In this document, strategies for an approach to trials under such circumstances are briefly outlined. In fact, most orphan drugs and paediatric indications submitted for regulatory approval are based on randomised controlled trials that follow generally accepted rules and guidance. Deviation from such standards is, therefore, uncommon and should only be considered when completely unavoidable and would need to be justified.

Approaches outlined in this document for situations where large studies are not feasible should not be interpreted as a general paradigm change in drug development. The methods described here to increase the efficiency of the design and analysis are also applicable for studies in large populations, but are often not used because of increased complexity.

The general principle can be applied to the following scenarios -: (1) when randomised controlled trials are feasible even though the interpretation will be less clear compared to typically sized phase III trials. This may be improved by special trial designs and/or refined statistical approaches (2) when randomised controlled trials will be severely underpowered. However, controlled studies with low statistical power in case of an important treatment effect may be preferable to no controlled studies. (3) when randomised controlled trials are not feasible and only case series (with external control groups) or even only anecdotal case reports are available. Here alternative approaches are required. Such compromise positions will usually be at the cost of increased uncertainty concerning the reliability of the results and hence the reliability of the effectiveness, safety and risk–benefit of the product. Additional follow-up data, post approval, will be necessary. . We have to consider the trade-off between relatively small amounts of high quality evidence (for example, small randomized trials) and relatively larger amounts of lower quality evidence (for example, large uncontrolled case series). This will always have to be judged on a case-by-case basis.

This document addresses (1) methods where the efficiency of the design or analysis may be increased, and (2) approaches for situations where such methods are not applicable. General principles are presented. They are by no means exhaustive but should encourage further exploration of potentially suitable methods for specific situations. In general, guidelines for trials in large populations are applicable to small populations. Any deviation from such established guidelines should be prospectively considered, and justified in the study protocol and further discussed in the study report. Seeking Scientific Advice, or Protocol Assistance, and expert statistical advice is strongly encouraged. This guideline should be read in conjunction with the following Directives and documents:

Directive 2001/20/EC of the European Parliament and Council of 4 April 2001 on the approximation of the Laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use.

- Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use.
- The Extent of Population Exposure to Assess Clinical Safety for Drugs (ICH E1A).

- Dose Response Information to Support Drug Registration (ICH E4).
- General Considerations for Clinical Trials (ICH E8).
- Statistical Principles for Clinical Trials (ICH E9).
- Choice of Control Group in Clinical Trials (ICH E10).
- Clinical Investigation of Medicinal Products in the Paediatric Population (ICH E11).
- Accelerated Evaluation of Products Indicated for Serious Diseases (Life Threatening or Heavily Disabling Diseases) (CPMP/495/96 *rev.* 1).
- Points to Consider on the Evaluation of the Diagnostic Agents (CPMP/EWP/1119/98).
- Points to Consider on Applications with 1.) Meta-analyses and 2.) One Pivotal Study (CPMP/2330/99).
- Reflection Paper on methodological issues in confirmatory clinical trials with flexible design and analysis plans (CHMP/2459/02).
- Points to Consider on Calculation and Reporting of the Prevalence of a Condition for Orphan Designation (CPMP/436/01).
- Note for Sponsors of Orphan Medicinal Products Regarding Enlargement of the European Union (EMEA/35607/03).

## 2. LEVELS OF EVIDENCE

Applications for marketing authorisations in small populations will be judged against the same standards as for other products, although limitations on patient recruitment will be taken into account.

Hierarchies of evidence have been described which usually place in order:

- Meta-analyses of good quality randomised controlled clinical trials that all show consistent results
- Individual randomised controlled trials
- Meta-analyses of observational studies
- Individual observational studies
- Published case-reports
- Anecdotal case-reports
- Opinion of experts in the field.

All such forms of evidence provide some information (even anecdotal case reports) and none should be ignored. However, high levels of evidence in drug development come from well-planned and well-executed controlled clinical trials, particularly trials that have minimised bias through appropriate blinding and randomisation. At their conclusion, the treatment effect should ideally be clinically relevant, confidence intervals for that effect should be narrow, and the effect size statistically significant. Well-planned and well-conducted meta-analyses of such trials will provide even stronger evidence. It must be recognised, that poor meta-analyses will not give reliable conclusions.

In very rare diseases, the combined evaluation of single case studies may be the only way to provide evidence. In such situations, treatment conditions and data collection should be standardised and data should be of high quality and adhere to GCP standards. Such studies should be prospectively planned and described in study protocols. A systematic review of all data (including data from other sources) will add weight to the evidence. Also combined analysis of individual case reports or observational studies should be considered.

Generally, for a given size of treatment effect, a larger sample size and/or a smaller variance will result in narrower confidence intervals and more extreme levels of statistical significance. Although  $p < 0.05$  is a common but arbitrary threshold for 'statistical significance', no such value is adequate to

confirm that a treatment effect truly does exist. In almost all cases, confidence intervals of estimates of the treatment effect are much more informative than *P*-values.

### 3. PHARMACOLOGICAL CONSIDERATIONS

Detailed knowledge of the pathophysiology of the disease and the pharmacology of the drug will facilitate the design of efficient clinical studies and will help determine the amount of clinical data required.

For rare diseases pre-clinical pharmacodynamic studies can be of importance if there exist adequate animal models and may be informative for the design of clinical studies. Such studies may also give important information for dosing and/or route of administration and the investigation of these features in man can be focused.

In deficiency diseases for ‘substitution studies’ (typically enzyme or hormone replacement), well-characterised short- and long-term consequences of the deficiency, and a clear understanding of the pharmacokinetics and pharmacodynamics of the compound, provide guidance for designing studies. Regulatory requirements for licensing ‘substitution products’ (notably recombinant products) may deviate from those for other compounds provided that symptoms related to the deficiency are clearly understood and that the pharmacokinetics and pharmacodynamics of the product are well documented in clinical studies. See for example the note for guidance on the clinical investigation of recombinant factor VIII or IX products. (CPMP BPWG 1561/99). Within-patient comparisons in a relentlessly and predictably progressive disorder might provide sufficient data to support a benefit–risk assessment. However, in other situations comparative trials may be needed/expected.

Variability (whether in terms of disease phenotype, underlying pathophysiology, pharmacodynamics or pharmacokinetics) is a threat to successful drug development. Efficient study design and analysis requires as clear an understanding as possible of all of these potential sources of variability.

The credibility of study results may be enhanced if a dose-response relationship is seen or in cases where a chain of events can be identified (for example, drug exposure to target occupancy, to pharmacodynamic measures, to clinical outcome). Cases where no such clear chain of events exists are much less convincing and will increase the data requirements regarding robustness and persuasiveness of study results.

In very rare disorders, it is important that every patient participating in a study contributes as much information as possible to make a benefit–risk assessment possible. Therefore, the well-planned use of the best available techniques to obtain and analyse information is crucial. This applies throughout the study process from pharmacokinetic and pharmacodynamic modelling to handling and analyses of biopsy material.

### 4. CHOICE OF ENDPOINTS

Ideally a ‘hard’ and clinically relevant endpoint is used. At one extreme, the endpoint may be complete ‘cure’ of disease, overall survival.

*Time to disease progression* is an endpoint of intermediate level and it requires a measure of disease severity or of disease progression. Ideally, this should be validated as a tool for use in clinical trials, but it is recognised that there might be too few patients to use some for validating endpoints and others for testing treatments. In studies whose endpoint is time to progression or time to remission, adequate length of follow up of the patients is important; this can be done in ‘open-label extensions’ or randomised studies. It is preferable, to be able to identify a causal relationship between treatment and a particular (beneficial) outcome.

A *clinical endpoint* like renal failure (e.g. in Fabry’s disease) is a good example of a clinically relevant endpoint because it severely impairs a patient’s survival and well-being. *Relief of symptoms* and the resulting patient preference may be a valuable study endpoint even in the absence of demonstration of a benefit on disease progression or survival. However, since it does not truly reflect the impact of treatment on disease progression or death, this must always be accompanied by a disease and treatment-dependent justification.

Health related quality of life is important when patients remain severely disabled (such as poor neurological status following resuscitation or after an intracranial bleed). If quality of life is measured, it should always be assessed using scales validated for the particular indication being treated. It is recognised that sometimes there are too few patients for validation exercises as well as separate treatment evaluation. However, improvements in quality of life alone (that is, in the absence of any other clinical benefit) is unlikely to be sufficient to grant a Marketing Authorisation. Quality of life data should ideally be considered as supportive evidence. It may be important to assess any influence of the observed study effect on activities of daily life and social functioning.

The choice of the primary endpoint may pose considerable problems. In some cases, the ‘most appropriate’ clinical endpoint may not be known or widely agreed or a validated clinical endpoint may not exist. In other cases, the mode of action of the test treatment may not be well enough known to predict which of several possible outcomes will be affected. In such circumstances, the usual approach of pre-specifying the primary endpoint may be too conservative and more knowledge may be gained from collecting all sensible/possible endpoints and then presenting all the data in the final study report. Still, every effort should be made to identify an appropriate hierarchy in the endpoints. If, collectively, the data look compelling, then a Marketing Authorisation may be grantable.

In the context of rare disorders for a given clinical endpoint or validated surrogate endpoint recruitment of a sufficient number of patients would be difficult or demonstration of this endpoint would take an unreasonable length of time. Then use of other surrogate markers as substitutes for a clinical endpoint may be considered. The term ‘surrogate endpoint’ should only be used for biomarkers, which have been validated. However, selection of a surrogate marker as study endpoint requires it to be reasonably likely – based on epidemiologic, pathophysiologic, or other evidence – to predict benefit. Prediction in itself may not be sufficient to establish efficacy. Considerations should include:

- How closely changes in the surrogate endpoint are causally linked to changes in a clinical endpoint or symptom.
- How much risk is associated with the therapy
- What other therapies (if any) are available for the same condition.

Demonstrating that a surrogate endpoint adequately reflects the true clinical endpoint is difficult. Epidemiological data and data from patient registers may provide some help. These data may be limited when there are very few patients.

A risk–benefit assessment may become very difficult, since the size of benefit may be impossible to determine based on a surrogate endpoint.

Also it has to be pointed out that surrogate markers cannot serve as final proof of clinical efficacy or long-term benefit. If they are intended to be the basis for regulatory review and approval then, unless they are properly validated, there should be a predetermined plan to supplement such studies with further evidence to support clinical benefit, safety and risk/benefit assessment.

## **5. CHOICE OF CONTROL GROUPS**

Ideally, we wish to obtain an unbiased estimate of the effect of the treatment being investigated compared to placebo or to another active compound and, for this reason, every effort should be made to randomise patients from the beginning of the therapeutic testing phase. The goal of obtaining an unbiased estimate of the size of effect is true in studies in small populations as well as large trials for common diseases. Thus, in developing any treatment, a comparative trial will usually be *preferable* but may not always be *possible*. In cases where there is no existing treatment, even in life-threatening diseases, the use of placebo as a comparator should be considered. Where a placebo control may not be possible, an appropriate control group may be ‘best standard of care’. When other treatments are available they an active comparator could be used as control group. However, if the active comparator does not have its own good evidence base, then superiority to that comparator will usually be necessary.

In general, there are two approaches to selecting control patients: internal controls or external controls, who may be historical or concurrent. The ideal is a comparative trial using an internal control group, as there are several well-known problems inherent with historical (or other external) controls.

Although internal controls are the preferred option for comparative trials, under exceptional circumstances external controls may be acceptable. Historical controls (using patients treated with 'current' therapies, or not treated at all) might, in some circumstances (even if not routinely), be acceptable to demonstrate efficacy, safety, ease of administration and so on, of a new treatment. In general, the absence of any control data is only likely to be acceptable if the natural course of the disease is very well known.

Patient registers may supply important information on the natural course of disease and may help in the assessment of effectiveness and safety. Further, such registers might be used as a source for historical controls. In cases where biomarkers are used in short-term (placebo-) controlled studies and the need for long-term follow-up is foreseen (for example, in enzyme replacement therapy), it is recognised that placebo control might not be feasible. The situation and the possible need for a historical control should be recognised early in the development of the product and a suitable database set up. In general, the setting up of such registers should be timely and pre-planned with a view to using them to help establish possible benefits of new therapies. They should capture the data necessary and measures taken to avoid bias should be carefully considered.

If only active controlled studies are possible, then showing equivalence or non-inferiority may be difficult because assay sensitivity of the study cannot be assured and so obtaining a licence in these circumstances becomes extremely difficult. Arguments concerning the natural course of a disease may help to support assay sensitivity of studies.

## **6. METHODOLOGICAL AND STATISTICAL CONSIDERATIONS**

The following text discusses a range of approaches, which may be helpful in particular situations. As already mentioned at the beginning of the document, any given list of possible approaches cannot be exhaustive and this list is not intended to be so.

Each approach – not only those listed here – has to be weighed according to its merits and drawbacks on a case-by-case basis and will always have to be carefully and fully justified. Sponsors are encouraged to seek scientific advice or protocol assistance when considering use of alternative designs and methodologies.

### **6.1 Design stage**

In conventional phase III trials sponsors often enrol several hundred or even thousands of participants. The design and conduct of any trial should be such that bio-noise is minimised. Bio-noise is the sum of avoidable and unavoidable non-systematic errors in the design and conduct of a trial. It usually (although not always) leads to a bias towards failing to show a difference between treatments. As an example, a typical error, which has avoidable and unavoidable elements, is loss-to-follow up. Researchers cannot force patients to stay in a study but there is empirical evidence that some measures may help to reduce the loss-to-follow-up rate in longitudinal studies. Examples include ensuring visits are scheduled at reasonable intervals and at times convenient to patients, providing transport for patients where necessary, etc. Further to this point, it is imperative that the data obtained are of the highest quality (particularly with respect to accuracy, lack of bias, precision, etc.). This further helps to reduce 'bio-noise'.

While in a large trial the impact of this noise-to-effect ratio can usually be reduced simply by increasing the sample size, it can become a severe problem in small studies. Therefore it is of utmost importance that sponsors pay great attention to the minimisation of avoidable errors. Important considerations are given in the ICH guideline on the statistical principles for clinical trials (E9) as well as to other standard texts on the design and analysis of clinical trials. Many of the following methods are likely to reduce the amount of bio-noise and thus increase the efficiency of a study – but in nearly all cases at a cost of increased complexity and possibly also bias.

Continuous variables usually allow for higher precision/smaller sample size than those that have been categorised or even split into 'responder' vs. 'non-responder'. This is particularly true if the baseline

value is accounted for in an appropriately pre-specified analysis of covariance (ANCOVA) model. Even when baseline measurements of the eventual outcome variable may not be available, other important prognostic variables are likely to increase the efficiency of an ANOVA or ANCOVA. Unreliability of one particular outcome can also be avoided by choosing another outcome (as long as this outcome is clinically meaningful and chosen prior to the start of the study), training of outcome assessors, and by using multiple ratings. All of these aspects should be considered before the study starts. The size of the clinical effect is always important when considering the balance of risks and benefits.

#### *Randomisation procedures*

Matching or stratification also improves power, particularly if matching or stratification is based on important prognostic variables. Such procedures, accompanied by pre-specified stratified analyses and sensitivity analyses may, therefore, be useful.

#### *Response-adaptive methods*

Response-adaptive designs change the allocation ratio based on which treatment appears to be 'best'. As patients complete a trial, if one treatment is beginning to emerge as better, then new patients entering the study are more likely to be allocated to that treatment. These designs are sometimes called 'play-the-winner' designs. The allocation probabilities can be continuously changing and do not rely on 'good evidence' of one treatment being superior (when, possibly, a study might be terminated anyway). As soon as one treatment appears better, the allocation of new patients is biased in favour of that treatment. As the study continues, the apparently 'best' treatment may change and allocation bias can change with it. Such methods rely on outcome data being available quickly (relative to patient recruitment) and also rely on continuously unblinding individual patients as they complete the study. The analysis may be very complex, as it is not based on standard assumptions of equal and constant probability of being assigned to either treatment.

A variation of response-adaptive designs is those used for dose finding – they are typically referred to as 'continual re-assessment' methods. They are sometimes, but rarely, used. The properties of such methods far outstrip those of conventional 'up and down' dose finding designs. They tend to find the optimum (however defined) dose quicker, they treat more patients at the optimum dose, and they estimate the optimum dose more accurately. Such methods are encouraged.

#### *Sequential designs*

Sequential designs – with a goal to demonstrate 'statistical significance' if a treatment is genuinely superior to control - generally reduce the required sample size. There can be several different types of sequential design – all providing valid statistical conclusions but each tailored to specific balances of expected outcomes and patient availability. Some designs are 'open-ended' and (in theory) continue to recruit patients until a reliable positive or negative conclusion about the treatments can be made. Other designs are 'closed' and so have a fixed upper limit to recruitment (but may stop before this). Stopping boundaries for benefit and harm need not be symmetrical; boundaries for showing benefit of treatment relative to an active control also need not be symmetrical. Stopping boundaries for futility may also be introduced. In all cases, the choices of stopping boundaries should be discussed and justified. Sequential designs, as with response-adaptive designs, require treatment outcomes to be available quickly (relative to the patient recruitment rate). This will almost never be the case if we are looking for long-term survival data, for example, but may be the case if we are looking at shorter term clinical or surrogate/bio-markers. A common problem with trials in rare diseases is that recruitment *is* slow because patients are so rare; hence such methods may have more of a place in these situations than in more common diseases. Ultimately, however, the extent to which the sample size can be reduced depends on the size of the effect. Variations on sequential methods are the, so-called, group-sequential methods and adaptive designs. See also draft guidance on adaptive designs (CHMP/2459/02).

#### *n-of-1 trials*

In this design the unit of randomisation is the intervention, rather than the patient. They are rather like crossover studies, but carried out in single patients. The patient's first treatment is determined at random and at the end of a treatment period, the patient is randomised again; a switch to the alternative instead of randomisation is also possible. Multiple switches may occur. The outcome of

such a study is a conclusion about the best treatment *for this particular patient*. Series of *n*-of-1 trials may begin to show trends for repeated preference of one treatment over another. The advantage of such a design is that each patient is assured of ‘eventually’ ending up on the treatment that is best for him or her. Different investigators can use the same treatment comparisons in ‘their own’ *n*-of-1 trials and do not have to conform to a standard protocol, which may be too restrictive in some individual cases. So rather than a patient being excluded from a trial because they do not meet the inclusion criteria, or because they would not be able to follow all the necessary trial procedures, each trial can be tailored to each patient. *n*-of-1 trials have many of the same limitations as crossover trials. They are most useful for fast-acting symptomatic treatments and in diseases that quickly return to stable baseline values after treatment. Results of many *n*-of-1 trials may then be combined in a manner similar to both a crossover study and a meta-analysis. Pre-planned *n*-of-1 trials, and pre-planned sequences of them will always be preferred and will usually be more convincing than unplanned cases.

## 6.2 Data analysis

### *Assumptions*

Studies with few patients are often perceived as presenting a rather simple situation: there is not much information (data) and so simple (often descriptive) analyses are all that are warranted. It seems quite counterintuitive, therefore, that for ‘simple’ situations more complex approaches should be applied but this is exactly what is necessary. Crude (simple) methods may often be adequate when we have huge amounts of data – but when there are very few data, it is imperative that the most efficient and informative analytical methods should be used. Many of these methods involve ‘statistical modelling’. Such models usually make assumptions about the data or the form of the treatment effect. With few data, these assumptions may not be testable or verifiable. However, assumptions *add* to the data so that more complex statistical models give us more information than simple descriptive statistics. Hence, sensitivity analyses consisting of various analyses/models should be presented, which may make different assumptions about the data. Such sensitivity analyses should be pre-planned. Then it can be seen if the conclusions are heavily reliant on the model assumptions or if, in fact, they are robust to a variety of plausible assumptions.

### *Non-parametric methods*

Contrary to the above, non-parametric, or ‘distribution-free’ methods may often be used when we cannot determine if data are from a Normal (or other specified) distribution. There are a wide variety of methods that make ‘few’ (although not usually ‘no’) assumptions about the data or about the form of any treatment effect.

### *Prognostic variables*

Adjustment for baseline variables may greatly improve the efficiency of an analysis. Factors used to stratify the randomisation in a study should be used to stratify the analysis – at least as ‘main effects’ but interaction terms between them should generally not be included in the primary analysis. Including prognostic variables in a model can greatly enhance the precision of a treatment effect but caution should be used not to include too many factors.

### *Repeated Measurements*

Repeated measurements over time – or in different body locations – may also improve the efficiency of an analysis. A commonly encountered problem in the analysis of such data is the non-independence between observations. Non-independence occurs when data fall into groups or clusters, e.g. in different body locations or in longitudinal studies. Standard statistical methods, such as generalized linear models (GLM) cannot be applied when analyzing dependent data, since the assumption of independence between observations is violated. Neglecting dependencies in these situations can lead to false conclusions. In general, the precision of the results and thereby their significance is usually overestimated. There are different methods available to analyze clustered dependent data e.g. the Generalized Estimated Equations (GEE) method, Hierarchical Linear Models or Mixed-effects models. These modern statistical approaches take the correlation within subjects into account and can also allow an unequal number of observations per subject, (e.g. caused by missing values) so that valid inferences can be assured. Judging the appropriateness of models and checking model assumptions is very important – but also very difficult with limited data.

### *Bayesian methods*

Bayesian methods are a further source of ‘adding assumptions’ to data. They are a way to formally combine knowledge from previous data or prior ‘beliefs’ with data from a study. Such methods may be advantageous when faced with small datasets, although introducing prior beliefs is often a concern in drug regulation. As with sensitivity analyses mentioned above, a variety of reasonable prior distributions should be used to combine with data from studies to ensure that conclusions are not too heavily weighted on the prior beliefs.

### **6.3 Reporting**

In the study report, (as well as in the study protocol), a justification for any deviations from the principles laid down in regulatory guidance is expected as well as justification for any alternative study design chosen (e.g. choice of surrogate end point, lack of randomisation, lack of control group) and a justification for the statistical considerations should be given. An orphan drug status as justification will not be sufficient.

Sensitivity analyses in order to show robustness of the data is expected.

The need for, and plans for long term follow-up for efficacy and safety should be discussed.

## **7. SUMMARY AND CONCLUSIONS**

- There are no special methods for designing, carrying out or analysing clinical trials in small populations. There are, however approaches to increase the efficiency of clinical trials. The need for statistical efficiency should be weighed against the need for clinically relevant/interpretable results; the latter being the most important.
- Guidelines (ICH, CHMP and others) relating to common diseases are also applicable to rare diseases.
- In situations where obtaining controlled evidence on the efficacy and safety of a new treatment is not possible, the regulatory assessment may accept different approaches if they ensure that the patients’ interests are protected.
- Detailed knowledge of the pharmacology of a compound may help when designing studies. Pharmacology studies may help identify sources of heterogeneity in patients. Non-clinical pharmacology may sometimes be helpful, especially in conditions where very few patients are available.
- Surrogate endpoints may be acceptable but need to be fully justified. Their relation to clinical efficacy must be clear so that the balance of risks and benefits can be evaluated.
- Controls and comparator groups are very important. Their absence compromises the reliability of studies.
- Patient registers may supply important information on the natural course of disease and may help in the assessment of effectiveness and safety. Furthermore, such registers can be used as a source for historical controls. Registers used in this way should contain high quality data; GCP inspection might be anticipated.
- It is strongly recommended that scientific advice/protocol assistance be sought during all phases of development to guide sponsors as to the acceptability of their planned approaches for later marketing.