The European Agency for the Evaluation of Medicinal Products
Veterinary Medicines and Inspections

COMMITTEE FOR VETERINARY MEDICINAL PRODUCTS

GUIDELINE ON STATISTICAL PRINCIPLES FOR VETERINARY CLINICAL TRIALS

AGREED BY EFFICACY WORKING PARTY	9 - 10 October 2000
ADOPTION BY CVMP FOR RELEASE FOR CONSULTATION	8 November 2000
START OF CONSULTATION	9 November 2000
END OF CONSULTATION	9 May 2001
REVIEWED BY EFFICACY WORKING PARTY	23-24 October 2001
FINAL ADOPTION BY CVMP	5 December 2001
DATE OF COMING INTO EFFECT	5 June 2002

STATISTICAL PRINCIPLES FOR VETERINARY CLINICAL TRIALS

1. INTRODUCTION

1.1 Background and Purpose

The efficacy and in-use safety of veterinary medicinal products should be demonstrated by clinical studies that follow the guidance in the current VICH guideline for Good Clinical Practice (GL9). In that guideline the role of statistics in clinical trial design and analysis is acknowledged as essential. The guideline which follows is written primarily to harmonise the principles of statistical methodology applied to clinical trials to support the applications for a marketing authorisation for veterinary medicinal products within Europe and complements and supplements the current guideline on Good Clinical Practice.

As a starting point, this guideline utilised the ICH (International Conference for Harmonisation) and the CPMP (Committee for Proprietary Medicinal Products) Notes for Guidance entitled "Statistical principles for clinical trials" and "Biostatistical methodology in clinical trials in applications for marketing authorisations for medicinal products" (December, 1994), respectively.

This guideline is intended to give direction to sponsors in the design, conduct, analysis, and evaluation of clinical trials of an investigational veterinary product in the context of its overall clinical development. The guidance will also assist scientific experts charged with preparing application summaries or assessing evidence of efficacy and in-use safety in the target species.

This guideline should be read in conjunction and integrated with other guidelines adopted within the European Union that deal with clinical development.

1.2 Scope and Direction

The focus of this guideline is on statistical principles. It does not address the use of specific statistical procedures or methods. Specific procedural steps to assure that principles are properly implemented are the responsibility of the sponsor. Integration of data across clinical trials is discussed, but is not a primary focus of this guideline. Selected principles and procedures related to data management or clinical trial monitoring activities are covered in the current guideline on Good Clinical Practice. This guideline should be of interest to all scientific personnel involved in the conduct of a clinical trial.

It is assumed that the responsibility for all statistical work associated with clinical trials will lie with an appropriately qualified and experienced statistician. The involvement of the statistician is to assure, in collaboration with other clinical trial professionals, that statistical principles are applied appropriately in clinical trials from the protocol development phase through to the final trial report. All important details of the design, conduct and proposed analysis of each clinical trial contributing to a marketing application should be clearly specified in a protocol written before the trial begins. The extent to which the procedures in the protocol are followed and the primary analysis is planned *a priori* will contribute to the degree of confidence in the final results and conclusions of the trial.
The protocol and subsequent amendments should be approved by responsible personnel including the statistician. The statistician should ensure that the protocol and any amendments cover all the relevant statistical issues clearly and accurately, using appropriate terminology.

Many of the principles delineated in this guideline deal with minimising bias and confounding, and maximising precision. It is important to identify potential sources of bias so that attempts may be made to limit such bias.

2.   CONSIDERATIONS FOR OVERALL CLINICAL DEVELOPMENT

2.1   Study Context

2.1.1 Type of Clinical Trial

The broad aim of the process of clinical development of a veterinary medicinal product is to determine whether there are a dose or dose range and a dosing schedule at which the product can be shown to be simultaneously safe and effective and thus strike a balance between the risks and benefits associated with the use of the product. The target population that may benefit from the product, and the specific indications for its use, also need to be defined.
Satisfying these broad aims usually requires an ordered programme of clinical trials, each with its own specific objectives. This should be specified in an ordered development plan, with appropriate decision points and flexibility to allow modification as knowledge accumulates.

Depending on the aim of the trial, it can be classed in one of the following three categories: confirmatory, exploratory, or composite.

2.1.2 Confirmatory Trial

Confirmatory trials can concern dose determination trials, dose confirmation trials as well as controlled field trials. For some specific product studies the design may be subject to other guidance such as that provided by the CVMP and the European Pharmacopoeia.

- Confirmatory trials are carried out in conformity with the current guideline for Good Clinical Practice which embodies the following points : Confirmatory trials

    - are controlled.
    - have an agreed protocol written and signed before the study begins.
    - test a hypothesis that is stated in advance.
    - only address a limited number of questions.
    - are necessary to provide firm evidence of efficacy and/or in-use safety.
    - estimate with due precision the size of the effects attributable to the treatment under evaluation and relates these effects to their clinical significance.
    - are justified in terms of their design and other statistical aspects such as planned analysis.
    - clearly and definitively answer each question relevant to support the stated hypothesis.
    - explain the generalisation from the chosen study animal population to the intended target animal population.
    - produce robust results.

In some circumstances, the weight of evidence from a single confirmatory trial may be sufficient.

2.1.3 Exploratory Trial

The rationale and design of confirmatory trials often rests on earlier clinical work carried out in a series of exploratory studies. Exploratory trials

    - are precursors to confirmatory trials.
    - also have clear and precise objectives.
    - such objectives may not always lead to simple tests of predefined hypotheses.
    - sometimes require data exploration during analysis.
    - the choice of the hypothesis may be data dependant.
    - contribute to the total body relevant evidence, but cannot be the sole basis of formal proof of efficacy or in-use safety.

2.1.4 Composite Trials

Any individual trial may have both confirmatory and exploratory aspects.

In confirmatory trials the opportunity may exist to subject the data to further exploratory analyses, which may serve to explain and support the trial findings and to suggest further hypotheses for research. The protocol should make a clear distinction between those aspects of the trial which are confirmatory and those which are exploratory.

2.2   Study Scope

2.2.1   Population

In the earlier phases of new product development the choice of subjects for a clinical trial may be heavily influenced by the wish to maximise the chance of observing specific clinical effects of interest, and hence, they may come from a very narrow sub-group of the total animal patient population for which the product may eventually be indicated. By the time confirmatory field trials are undertaken the study animals should more closely mirror the intended population. Hence, in these trials it is generally helpful to relax the inclusion and exclusion criteria as much as possible within the target indication, whilst maintaining sufficient homogeneity to permit the successful conduct of the trial. No individual clinical trial can be expected to be totally representative of the future target population because of potential influences of, for example, geographical location, timing, animal husbandry, and local veterinary clinical practices. Wherever possible the influence of such confounding factors should be taken into account and subsequently discussed during the interpretation of the trial results.

2.2.2   Primary and Secondary Variables

The primary variable also known as the primary endpoint variable, should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial. Reference to CVMP/VICH guidelines may provide guidance in selection of such variables for some specific product studies. Generally there should only be one primary variable. This will usually be an efficacy variable, which reflects the accepted norms and standards in the relevant field of research or clinical studies. The variable should be reliable and validated and derived from experience in previous studies or in published scientific literature. There should be sufficient evidence that the chosen primary variable can provide a valid and reliable measure of some clinically relevant and important clinical benefit in the study animal population as defined by the inclusion and exclusion criteria. This is generally the variable used to estimate the sample size.

In many cases, and especially when treatment is directed at a chronic rather than an acute process, the approach to assessing subject outcome may not be straightforward and needs to be carefully defined. The primary variable should be specified in the protocol, along with the rationale for its selection. Redefinition of the primary variable after unblinding will almost always be unacceptable, since the biases this introduces are difficult to assess.

Secondary variables are either supportive measurements related to the primary objective or measurements of effects related to the secondary objectives. Their pre-definition in the protocol is also important, as well as an explanation of their relative importance and roles in interpretation of trial results.

In some situations it may be useful to combine the multiple measurements into a single or "composite" variable, using a pre-defined algorithm. The method of combining the multiple measurements should be specified in the protocol, and an interpretation of the resulting scale should be provided in terms of the size of a clinically relevant benefit. Combining multiple measurements addresses the multiplicity problem without requiring adjustment for multiple comparisons. When composite variables are used as primary variables, the individual components of these variables are often also analysed separately. When a rating scale is used as a primary variable, it is especially important to address such factors as

relevance, inter- and intra- assessor variability and sensitivity for discriminating different clinical conditions.

In some cases, « global assessment » variables may be developed to measure the overall in-use safety, overall efficacy, and/or overall usefulness of a treatment. Global assessment variables generally have a subjective component. The use of a global assessment variable as a primary or secondary variable, requires fuller details in the protocol (see glossary).

In exploratory trials, in order to cover the range of effects of the therapies, it may be desirable to use more than one primary variable, each or some of which, in the context of confirmatory trials, could provide sufficient basis for an efficacy and/or safety claim.

The primary hypothesis or hypotheses should be clearly stated with respect to the primary variables identified, and the approach to testing these hypotheses described. When direct assessment of the clinical benefit to the study animal through observing actual clinical efficacy is not practical, indirect criteria (surrogate variables) may be considered and justified depending on their biologic plausibility and prognostic value.

Criteria of "success" and "response" are common examples of dichotomies which require precise specification in terms of, for example, a minimum percentage improvement (relative to baseline) in a continuous variable or a ranking categorised as at or above some threshold level (e.g., "good") on an ordinal rating scale. Categorisations are most useful when they have clear clinical relevance. The criteria for categorisation should be pre-defined in the protocol, as knowledge of trial results could easily bias the choice of such criteria. Because categorisation normally implies a loss of information, a consequence will be a loss of power in the analysis: this should be accounted for in the sample size calculation.

2.3   Design Techniques to Avoid Bias

Some sources of bias arise from the design of the trial; one example would be the systematic assignment of treatments such that study animals with a poor prognosis would all be assigned to the same treatment group. Other sources of bias arise during the conduct and analysis of a clinical trial. For example, protocol deviations and exclusion of subjects from analysis based upon knowledge of treatment outcome for that animal.

The two most important design techniques for avoiding bias in clinical trials are blinding and randomisation. These techniques should always be considered when designing clinical trials to support an application for marketing authorisation.

2.3.1 Blinding

Blinding is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial which may arise from the knowledge of specific treatment a study animal may be receiving or is about to receive. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed. Therefore, it is important to try achieving optimum blinding (i.e. full-blinded study).

A full-blinded trial is one in which the investigator and sponsor staff involved in the treatment or clinical evaluation, the owner or carers of the study animals, or any other persons associated with administrating the treatment, are unaware of the treatment received by the study animals. This includes anyone determining subject eligibility, evaluating endpoints, or assessing compliance with the protocol. Where possible, this may also include the statistician. This level of blinding is maintained until all the study data are cleaned and only then are appropriate personnel unblinded. The sponsor should have adequate standard operating procedures (SOPs) or recommendations in the protocol to guard against inappropriate dissemination of treatment codes to blinded personnel by staff, who by the nature of their work and responsibilities have to remain unblinded.

Difficulties in achieving full blinding may arise particularly where the treatments are of a different nature. Nevertheless, every effort should be made to overcome these difficulties and any suitable method to keep the trial blinded should be considered. One way of achieving full blinding conditions under these circumstances is to use a "double dummy" technique.

If a full blinded trial is not feasible, it should be justified and then the partially blinded trial should be considered. If a study is to be conducted with partial blinding, it should be clearly specified which members of the sponsor or investigator's staff are to be blinded, and whether the owner or study animal carer are to be blinded and at what stage of the study blinding was achieved.

In an open-label trial the identity of treatment is known to all. An open label study can be avoided and partial blinding achieved by denying personnel involved with clinical assessments access to treatment information.

In partially blinded or open-label trials, every effort should be made to minimise known sources of bias and make the primary variable as objective as possible. The reasons for the degree of blinding to be achieved and the measures to be taken to minimise bias should be explained in the protocol. Breaking the blind (for a single study animal) should be considered only when knowledge of the treatment assignment is deemed essential to the veterinary care and welfare of the study animal. Any intentional or unintentional breaking of the blind should be reported and explained at the end of the trial, irrespective of the reason for its occurrence.

2.3.2 Randomisation

Randomisation introduces a deliberate element of chance into the assignment of treatments to subjects in a clinical trial. During subsequent analysis of the trial data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. It also tends to produce treatment groups in which the distributions of prognostic factors (known and unknown) are similar. The randomisation schedule of a clinical trial documents the random allocation of treatments to study animals. In the simplest form it could be a sequential list of treatments (or treatment sequences in a crossover trial) or corresponding codes by subject number. Different study designs will require different procedures for generating randomisation schedules.

Although unrestricted randomisation is an acceptable approach, some advantages can generally be gained by randomising subjects into blocks. These include: an increase in comparability of the treatment groups particularly when the study animal characteristics change over time; provision of a better guarantee that the treatment groups will be of nearly equal size; provision of finding a way of obtaining balanced designs in crossover studies with greater efficiency and easier interpretation. Care must be taken to choose block lengths which are sufficiently short to limit possible imbalance, but suitably long enough to avoid predictability.

In multicentre trials the randomisation procedures should be organised centrally. There may be advantages in having stratification by centre or allocating several whole blocks to each centre. In a properly randomised multicentre trial, the next study animal to be randomised into a study should always receive the treatment corresponding to the next free number in the appropriate randomisation schedule or in the respective stratum as appropriate. It is preferable for the subsequent animal to be processed only after this procedure has occurred.

Details of the randomisation which facilitate predictability, such as block length, should not be included in the protocol. The randomisation schedule itself should be filed securely by the sponsor or an independent party to ensure blindness is maintained.

The procedure to be followed, the documentation required, and the subsequent treatment and assessment of the study animal for which the blinding has been broken as a result of an emergency should be described in the protocol.

3. STUDY DESIGN CONSIDERATIONS

3.1 Study Configuration

3.1.1 Parallel Group Design

The most common clinical trial design for confirmatory trials is the parallel group design in which study animals are randomised to one of two or more arms, each arm being allocated a different treatment. These treatments will include the investigational product at one or more doses, and generally one or more control treatments, such as placebo and/or an active comparator. The assumptions underlying this design are less complex than for most other designs. However, there may be additional features of the design which complicate the analysis and interpretation (e.g. covariates, repeated measurements over time, interactions between design factors, protocol deviations, dropouts and withdrawals).

3.1.2 Cross-over Design

In the cross-over design, each study animal is randomised to a sequence of two or more treatments, and hence acts as its own control for treatment comparisons. This simple manoeuvre is attractive primarily because it reduces the number of animals and usually the number of assessments required to achieve a specific power, sometimes to a marked extent. In the simplest 2x2 cross-over design each animal receives each of two treatments in randomised order in two successive treatment periods, often separated by a wash-out period. The condition of the animal under study, either diseased or normal, should be stable. The relevant effects of the medication must develop fully within the treatment period. The wash-out periods should be sufficiently long for complete reversibility of drug effect. The fact that these conditions are likely to be met should be established in advance of the trial by means of prior information and data.

The most common use of the 2x2 cross-over design is to demonstrate the bioequivalence of two formulations of the same medication (see guidelines for the conduct of bioequivalence studies for Veterinary medicinal products (EMEA/CVMP/016/00)).

3.1.3 Factorial Designs

In a factorial design two or more treatments are evaluated simultaneously in the same set of subjects through the use of varying combinations of the treatments. The simplest example is the 2x2 factorial design in which study animals are randomly allocated to one of the four possible combinations of two treatments, A and B say. These are: A alone; B alone; both A and B; neither A nor B. In many cases this design is used for the specific purpose of examining the interaction of A and B. The statistical test of interaction is model dependent and may lack power to detect an interaction if the sample size was calculated based on the test for main effects. This consideration is important when this design is used for examining the joint effects of A and B, in particular, if the treatments are likely to be used together.

Another important use of the factorial design is to establish the dose-response characteristics of a combination product e.g. one combining treatments C and D especially when the efficacy of each monotherapy has been established at some dose in prior studies. A number, m, of doses of C is selected, usually including a zero dose (placebo), and a similar number, n, of doses of D. The full design then consists of mn treatment groups, each receiving a different combination of doses of C and D. The resulting estimate of the response surface may then be used to help to identify an appropriate combination of doses of C and D for clinical use.

3.2. Multicentre Trials

In general, only laboratory studies or exploratory field studies can be carried out at a single site. .

Multicentre trials are carried out for two main reasons. Firstly, it is an accepted way of evaluating a new medication more efficiently; under some circumstances, it may present the only practical means

of accruing sufficient study animals to satisfy the trial objective within a reasonable time-frame. Multicentre trials of this nature may, in principle, be carried out at any stage of clinical development. They may have several centres with a large number of animals per centre or, in the case of a rare disease, they may have a large number of centres with very few subjects per centre.

Secondly, a study may be designed as a multicentre (and multi-investigator) study primarily to provide a better basis for the subsequent generalisation of its findings. This arises from the possibility of recruiting the subjects from a wider population and administering the medication in a broader range of clinical settings, thus presenting an experimental situation which is more typical of future use. In this case the involvement of a number of investigators also gives the potential for a wider range of clinical judgement concerning the value of the medication. The multicentre study might sometimes be conducted in a number of different countries in order to facilitate generalisability even further.

If multicentre studies are to be meaningfully interpreted and extrapolated, then the manner in which the protocol is implemented should be clear and similar at all centres. Furthermore, the usual sample size and power calculations depend upon the assumption that the differences between the compared treatments in the centres are unbiased estimates of the same quantity. It is important to design the common protocol and to conduct the trial with this background in mind. Procedures should be standardised as completely as possible. Variation of evaluation criteria and schemes can be reduced by investigator meetings, by the training of personnel in advance of the study and by careful monitoring during the study. Good design should generally aim to achieve the same distribution of subjects to treatments within each centre and good management should maintain this design objective.
The statistical model to be adopted for the comparison of treatments should be described in the protocol. In particular it needs to differentiate between those variables which are classified as fixed effects and those which are classed as random effects. Any rules for combining centres in a multicentre analysis should be justified and specified prospectively in the protocol where possible.

If appropriate, i.e. when centres are a fixed effect, a treatment-by-centre interaction should be explored, as this may affect the generalisation of the conclusions. Marked treatment-by-centre interaction may be identified by graphical display of the results of individual centres of by analytical methods, such as a significance test of the interaction.
In the absence of a true centre-by-treatment interaction, the routine inclusion of interaction terms in the model reduces the efficiency of the test for the main effects.
In the presence of a true centre-by-treatment interaction the interpretation of the main treatment effect is controversial.

In some studies, for example some large mortality studies with very few subjects per centre, there may be no reason to expect the centres to have any influence on the primary or secondary variables because they are unlikely to represent influences of clinical importance. In other studies it may be recognised from the start that the limited numbers of subjects per centre will make it impracticable to include the centre effects in the statistical model. In these cases it may not be appropriate to include a term for centre in the model, because in this situation randomisation is rarely stratified by centre.

3.3 Type of Comparison

All the studies should be designed to control the risk of concluding erroneously what they are supposed to demonstrate.

3.3.1 Trials to Show Superiority

Superiority studies are designed to detect a difference. Scientifically, efficacy is most convincingly established by demonstrating superiority to placebo in a placebo-controlled trial, by showing superiority to an active control treatment or by demonstrating a dose-response relationship. This type of trial is referred to as a "superiority" trial (see Section 5.2.3).

For serious illnesses, when an appropriate positive control exists, a placebo-controlled trial may be considered unethical. In that case the scientifically sound use of the active control should be considered. The appropriateness of placebo-control vs. active control must be considered on a study-by-study basis.

3.3.2 Trials to Show Equivalence or Non-inferiority

An investigational product can be compared to a reference treatment without the objective of showing superiority. This type of trial is divided into two major categories according to its objective; one is an "equivalence" trial and the other is a "non-inferiority" trial.

Bioequivalence trials fall into the former category (more details are given in a specific guideline: Conduct of Bioequivalence Studies for Veterinary Medicinal Products (EMEA/CVMP/016/00)). In some situations, clinical equivalence trials are also undertaken for other regulatory reasons such as demonstrating the clinical equivalence of a generic product to the marketed product when the compound is not absorbed and therefore not present in the blood stream.

An equivalence margin should be specified in the protocol: This margin is the largest difference which can be judged as being clinically acceptable. The result is a theoretical interval. For the active control equivalence trial, both the upper and the lower equivalence margins of this interval are needed. The choice of equivalence margins requires clinical justification. Equivalence is inferred when, for a given variable, the entire confidence interval of the investigational product falls within the equivalence margins of the theoretical interval. This is the same as the method of using two simultaneous one sided tests to test the (composite) null hypothesis that the treatment difference is outside of the equivalence margins versus the (composite) alternative that the treatment difference is within the limits. With this method, the overall Type I error can be controlled at the required level of significance.

Many active control trials are designed to show that the efficacy of an investigational product is no worse than that of the active comparator, and hence fall into the latter category of « non-inferiority studies ». An equivalence margin should be specified in the protocol: this margin is the largest difference which can be judged as being clinically acceptable. For non-inferiority trials, the lower or upper equivalence margin, depending on the criteria chosen, is the only one needed. The confidence interval approach has a one-sided hypothesis test counterpart testing the null hypothesis that the treatment difference (investigational product minus control) is equal to the lower equivalence margin versus the alternative that the treatment difference is greater than the lower equivalence margin.

Active control equivalence or non-inferiority trials may also incorporate a placebo, thus pursuing multiple goals in one trial, for example, establishing superiority to placebo and hence validating the study design and evaluating the degree of similarity of efficacy and safety to the active comparator. There are well known limitations associated with the use of the active control equivalence (or non-inferiority) trials that do not incorporate a placebo. These relate to the implicit lack of any measure of internal validity (in contrast to superiority trials), thus making external validation necessary. The equivalence (or non-inferiority) trial is not conservative in nature, so that many flaws in the design or conduct of the trial will tend to bias the results towards a conclusion of equivalence. For these reasons the design features of such trials need special attention.

Active comparators should be chosen with care. An example of a suitable active comparator would be a widely used therapy whose efficacy in the relevant indication has been clearly established and quantified in well designed and well documented superiority trial(s) and which can be reliably expected to exhibit similar efficacy in the contemplated active control study.

It is vital that the protocol of a trial designed to demonstrate equivalence or non-inferiority contains a clear statement that this is its explicit intention.

Concluding equivalence or non-inferiority based on observing a non-significant test result of the null hypothesis that there is no difference between the investigational product and the active comparator is inappropriate.

There are also special issues in the choice of analysis sets. Subjects that are withdrawn from the treatment group or the comparator group will tend to have a lack of response, and hence the analysis of all randomised subjects may be biased toward demonstrating equivalence (see Section 5.2.3).

3.3.3 Dose-response Designs

Dose response studies may serve a number of objectives, amongst which the following are of particular importance: The confirmation of efficacy; the investigation of the shape and location of the dose-response curve; the estimation of an appropriate starting dose; the identification of optimal strategies for individual dose adjustments; and/or the determination of a maximal dose beyond which additional benefit would be unlikely to occur.

These objectives need to be addressed using the data collected at a number of doses under investigation, including a placebo (zero dose). For this purpose the application of estimation procedures, including the construction of confidence intervals, and of graphical methods is often as important as the use of statistical tests. The hypothesis tests which are used may need to be tailored to the natural ordering of doses or to particular questions regarding the shape of the dose-response curve (e.g. monotonicity). The details of the planned statistical procedures should be given in the protocol.

3.4   Group Sequential Designs

Group Sequential designs are used to facilitate the conduct of interim analysis (see section 4.5). While group sequential designs are not the only acceptable types of designs permitting interim analysis, they are the most commonly applied because it is more practicable to assess grouped subject outcomes at certain intervals during the trial than on a continuous basis as data from each subject become available. The statistical methods should be fully specified in advance.

3.5 Experimental unit

In veterinary clinical studies there are a variety of situations where the experimental unit is not the animal but a pen, room, pasture or litter, as well as an udder quarter for milking cows. For example, dogs and cats tend to be presented in a veterinary surgery singly or may be group housed in a kennel or cattery. Chickens are usually housed in groups of hundreds (layers) or many thousands (broilers). Pigs, on the other hand, may be seen singly (sow or boar), as a litter (sow plus 10-12 piglets), a weaner pool (25-50) or a fattening group (pens of 10-40). A fish tank or cage can also constitute an experimental unit. It still is possible for the individual animal to be the experimental unit even when the animals are group housed. This occurs when individual animals within the group are able to receive different treatments.

However, the follow up of the clinical condition should be done at the individual animal level. The experimental unit should be clearly specified in the protocol, since it is essential to the sample size calculation.

3.6   Sample size

The number of subjects in a clinical trial should always be large enough to provide reliable answers to the questions addressed. This number is usually determined by the primary objective of the trial. If the sample size is determined on some other basis, then this should be made clear and justified. For example, a trial sized on the basis of safety questions or requirements may need a larger number of subjects than one sized on the basis of efficacy questions.

The usual method for determining the appropriate sample size requires that the following items should be specified. A primary variable, the test statistic, the null hypothesis, the alternative ("working") hypothesis at the chosen dose(s) embodying consideration of the treatment difference to be detected or rejected at the dose and in the subject target population selected), the probability of erroneously rejecting the null hypothesis (the type I error) and the probability of erroneously failing to reject the null hypothesis (the type II error), as well as the approach to dealing with treatment withdrawals and protocol deviations. In some instances, the event rate is of primary interest for evaluating power, and assumptions should be made to extrapolate from the required number of events to the eventual sample size for the study.

The method by which the sample size is calculated should be given in the protocol, together with the estimates of any quantities used in the calculations (such as variances, mean values, response rates, event rates, difference to be detected). The basis of these estimates should also be given. In the case of more than one primary variable, the most unfavourable (i.e. the largest) sample size obtained from each variable should be retained. Moreover, in this case, the sample size calculation should take into account the multiplicity of the planned tests.

In confirmatory studies, assumptions should normally be based on published data or on the results of earlier studies. The treatment difference to be detected may be based on a judgement concerning the minimal effect that has clinical relevance in the management of animal patients or on a judgement of the anticipated effect of the new treatment, where this is larger.

Conventionally the probability of type I error is set at 5% or less or as dictated by any adjustments made necessary for multiplicity considerations; the precise choice is influenced by the prior plausibility of the hypothesis under test and the desired impact of the results. The probability of type II error is conventionally set at 20% or less; it is in the sponsor's interest to keep this figure as low as feasible especially in the case of studies which are difficult or impossible to repeat. When the hypotheses to be tested are well written (i.e. in a way that the null hypothesis is the one to be rejected), it is not useful for guidelines to impose any specific value for the type II error.

Sample size calculations should refer to the number of experimental units required for the primary analysis.

The sample size of an equivalence trial (see Section 3.3.2) should normally be based on the objective of obtaining a confidence interval for the treatment difference that shows that the treatments differ at most by a clinically acceptable difference. The power is usually assessed at a true difference of zero but can be estimated inappropriately if the true difference is not zero. In equivalence testing the relevant null hypothesis is that a difference of at least x exists between response with the test product and the control product, and the trial is targeted at disproving this in favour of the alternative hypothesis that no difference exists. In non-inferiority studies, the relevant null hypothesis is that response where the test product is less than "x" below the control, and the alternative hypothesis is that the response is equal to or greater than with the control. The chosen « x » should generally be smaller than in a superiority trial. In comparative trials against placebo, « x » is often set equal to a difference of undisputed clinical importance, and hence may be above the minimum difference of clinical interest. However, when comparing a new agent with standard comparator it is necessary to show that the new agent is sufficiently similar to the standard to be clinically not distinguishable. This entails using smaller values of « x » than were used to detect the effect of the standard relative to placebo.

For non-inferiority studies, the power is usually assessed at an expected (non-zero) difference, but can be underestimated if the true difference is less than expected. The choice of a « clinically acceptable » difference requires justification, and again may be smaller than the « clinically relevant » difference referred to above in the context of superiority trials designed to establish that a difference exists.

The sample size in a group sequential study cannot be fixed in advance because it depends upon the play of chance in combination with the chosen stopping rule and the true treatment difference. The design of the stopping rule should take into account the consequent distribution of the sample size usually embodied in the expected and maximum sample sizes.

When event rates are lower than anticipated or variability is larger than expected, methods for sample size re-estimation are available without unblinding data or making treatment comparisons (see Section 4.3.1).

3.7   Data capture and processing

The data capture and processing should be performed in accordance with the VICH guideline for « Good Clinical Practice ».

4.   STUDY CONDUCT

4.1   Changes in Inclusion and Exclusion Criteria

Inclusion and exclusion criteria should remain constant, as specified in the protocol, throughout the period of subject recruitment. However, occasionally changes may be appropriate, for example, as a result from the discovery by monitoring staff that regular violations of the entry criteria are occurring, or that seriously low recruitment rates are due to over-restrictive criteria. Changes should be made without breaking the blind and should always be described by a protocol amendment which should cover any statistical consequences, such as sample size adjustments arising from different event rates, or stratification of the analysis according to modified inclusion/exclusion criteria.

4.2   Recruitment Rates

In studies with a long period for the recruitment of study animals, it is necessary to monitor the rate of recruitment in order to take remedial measures if it falls below the projected rate in order to protect the power of the trial. In a multicentre trial, this applies to the individual centres.

4.3   Interim analysis

4.3.1 Sample Size Adjustment

In long term trials there may be an opportunity to check the assumptions that underlie the original design and sample size calculations. This may be important in trials where assumptions have been made on preliminary or uncertain information. An interim check can be made and, where necessary, a revised sample size recalculated based on newly revised assumptions. The potential need for sample size re-estimation should be planned for in the protocol or, in exceptional circumstances, be recorded and justified in a protocol amendment. The steps to preserve blindness and the consequences of sample size re-estimation on the Type I error and the widths of the confidence intervals should also be explained.

4.3.2 Interim Analysis and Early Stopping

The schedule of interim analyses, or at least the considerations which will govern its generation, should be stated in the protocol or in a protocol amendment before the time of the first interim analysis; as flexible statistical methods are available to conduct interim analyses according to a variety of needs (early or late in a trial), the stopping guidelines and their properties should be clearly stated in the protocol or amendments.

It is recognised that drug development plans involve the need for sponsor access to comparative treatment data for a variety of reasons, such as planning other studies, and that only a subset of trials will involve the study of serious life-threatening outcomes or mortality which may need sequential monitoring of accruing comparative treatment effects for ethical reasons. Interim analysis may be particularly useful if the trial specifications have been made on preliminary and/or uncertain information. In long-term studies, for example, growing knowledge either from outside the trial or from interim analyses may suggest a change of entry criteria. An interim check may reveal that overall response variances, event rates or survival experience are not as anticipated.

When an interim analysis is planned with the intention of deciding whether or not to terminate a trial, this is usually accomplished by the use of a group sequential design which employs statistical monitoring schemes as guidelines (see Section 3.4). The goal of such an interim analysis is to stop the trial early if the superiority of the treatment under study is clearly established, if the demonstration of a relevant treatment difference has become unlikely or if unacceptable adverse effects are apparent. Generally, boundaries for monitoring efficacy require more evidence to terminate a trial early (i.e., more conservative) than do boundaries to terminate a trial for safety reasons. When the trial design and monitoring objective involve multiple endpoints, then this aspect of multiplicity may also need to be taken into account.

Deviations from the planned procedure always bear the potential of invalidating the study results. If it becomes necessary to make changes to the trial, any consequent changes to the statistical procedures should be specified in an amendment to the protocol at the earliest opportunity, especially discussing the impact on any analysis and inferences that such changes may cause. The procedures selected should always ensure that the overall probability of Type-I error is controlled.

4.3.3 Confidentiality of Interim Analysis

The execution of an interim analysis must be a confidential process because unblinded data and results are potentially involved. All investigators involved in the conduct of the trial should remain blind to the results of such analyses, because of the possibility that their attitudes to the trial will be modified and cause changes in recruitment patterns or biases in treatment comparisons. This principle applies to the staff of the investigators and to staff employed by the sponsor that come into contact with clinic staff or subjects. Investigators should only be informed about the decision to continue or to discontinue the trial, or to implement modifications to trial procedures.

Any interim analysis that is not planned in the protocol or in an amendment to the protocol prior to unblinding the data (with or without the consequences of stopping the trial early) may flaw the results of a trial and possibly weaken confidence in the conclusions drawn. Therefore, such analyses should be avoided. If unplanned interim analysis is conducted, the study report should explain why it was necessary, the degree to which blindness had to be broken, provide an assessment of the potential magnitude of bias introduced, and the impact on the interpretation of the results.

5. DATA ANALYSIS

5.1 Pre-specified statistical analysis

When designing a clinical trial, the principal features of the statistical analysis should be described in the statistical section of the protocol.

5.1.1 Statistical Section of the Protocol

This statistical section of the protocol should include the principal features of the statistical analysis. These include, where relevant:

- Definition of the experimental unit
- Hypothesis to be tested
- Treatment effect(s) to be estimated
- Statistical model, test(s) and construction of confidence intervals
- Justification of the use of one- or two-sided tests
- Use of covariate(s)
- Significance threshold
- Power $(1-\beta)$ and other assumptions used in sample size estimation
- Set of experimental units to be included in the analysis
- Planned data transformations
- Bayesian estimates
- Reporting of summary data
- Comparison of groups at baseline
- Alternative methods to be used in case of expected problems (heteroscedasticity, non-normality…).

For exploratory trials, this section could describe more general principles and directions.

5.1.2 Standard Operating Procedures

The statistical analysis section of the protocol may refer to SOPs for the most common statistical methods, for the handling of the most common problems (protocol deviations, missing data, outliers…), and for the transformation of data generally used to stabilise the variances and/or to increase the symmetry of the residuals distribution. If any reference to SOPs is made, the SOPs in questions should be attached to the study report.

5.1.3 Statistical Section of the Report

In the statistical section of the clinical study report, the statistical methodology should be clearly described. It should also describe when methodology decisions were made in the clinical trial process.

5.2 Analysis sets

If all study animals randomised into a clinical trial satisfied all entry criteria, followed all trial procedures and provided complete data records, then all the animals would be protocol compliant and would be used in the analysis. While the design and conduct of a trial should aim to approach this ideal, the protocol may prospectively address how to handle data from clinical studies where biological and environmental realities deviate from the ideal. To limit deviations, the protocol can also define acceptable ranges for compliance for visit times, treatment doses, etc. The protocol should also specify procedures aimed at minimising any anticipated irregularities in study conduct that might impair a satisfactory analysis, including various types of protocol violations, withdrawals and missing values. The protocol should consider ways both to reduce the frequency of such problems and handle problems that occur in the analysis of data. The blind review of data to identify possible amendments to the analysis plan due to the protocol violations should be carried out before unblinding. It is desirable to identify any important protocol violation with respect to the time when it occurred, its

cause and its influence on the trial result. The frequency and type of protocol violations, missing values and other problems should be documented in the study report and their potential influence on the trial results should be described.

5.2.1 All Randomised Study Animals Dataset

Some qualifying animals may not be randomised to treatments due to an excess of animals, and some qualifying animals may be excluded from randomisation due to non-study attributes that may affect the ability of the animal to complete the study, e.g., those with concurrent illnesses or owners who the Investigator doubts will return to study visits.

All animals randomised into a treatment group and receiving at least one dose of medication will comprise the All Randomised Study Animals Dataset.

5.2.2 Per Protocol Set of Study Animals

All study animals that received the required level of study medication and reasonably complied with the protocol comprise the Per Protocol Dataset. Minor deviations from the ideal may still have occurred with these animals; however, the deviations are not expected to have any bearing on the evaluation of the primary or secondary outcomes.

This Per-protocol set of study animals excludes animals that do not meet entry criteria and whose removal from the analysis does not introduce bias. Animals that have severe protocol deviations during the conduct of the study are also removed, and the analysis should discuss if the exclusions tended to be from any single treatment that could potentially be due to bias. To prevent bias, decisions to include or exclude an animal with a protocol deviation should be performed before the study is unblinded, whenever possible. All animals that received even one dose of study medication should be maintained in the tabulation and analysis of safety variables.

5.2.3 Roles of the Different Analysis Sets

The "All Randomised Study Animals analysis" set is more likely to mirror the treatment effect(s) observed in practice, whereas the "Per protocol" analysis maximises the opportunity for a new treatment to show additional efficacy in the analysis, and most closely reflects the scientific model underlying the protocol.

The "All Randomised Study Animals" set and the "Per Protocol" set may play different roles in superiority trials and in equivalence or non-inferiority trials. In superiority trials the " All Randomised Study Animals" set tends to avoid over-optimistic estimates of efficacy (non-compliers will generally diminish the estimated treatment effect). In non-inferiority or equivalence trials, similar conclusions from both the " All Randomised Study Animals" and the "Per Protocol" analysis sets may allow for a robust interpretation in the analysis of the primary outcome(s). However, it should be recognised that for clinical efficacy parameters the "Per protocol" dataset may be the only analysable dataset.

5.2.4 Comparison of Baseline Values

The most relevant summary data should also be provided for the «Per Protocol» analysis set, when this sub-sample is used for the analysis of key variable(s). If treatment groups are to be compared for demographic data, baseline values, prognostic variables, etc., the «All Randomised Study Animals» analysis set may be considered.

5.3 Missing Values and Outliers

Missing values and the presence and/or exclusion of outliers represent a potential source of bias in a clinical trial. Hence, every effort should be undertaken to fulfil all the requirements of the protocol concerning the collection and management of data.

The handling of missing data and outliers should be described as part of the statistical section of the protocol or in the study report. The decision on whether to keep or to exclude extreme values should be discussed. For the main endpoint, two separate analyses may be provided, with and without outlier(s), and the differences between their results discussed.

5.4 Data Transformation/Modification

Transformation of data is often necessary for confirming basic statistical assumptions. However, transformations should only be applied where necessary.

The decision to transform key variables prior to analysis is best made during the design of the trial on the basis of *a priori* knowledge (from previous studies, publications, guidelines…). Transformations (e.g. square root, logarithm…) should be specified in the protocol and a rationale provided wherever possible.

5.4.1 Assumptions Underlying the Statistical Models

The general principles guiding the use of transformations to ensure that the assumptions underlying the statistical methods are met are to be found in standard texts; conventions for particular variables have been developed in a number of specific clinical areas. This can sometimes lead to the use of unplanned transformations. In this case, a justification should be given in the report.

5.4.2 Clinical Interpretation of Analysis Made on Transformed Data

Transforming endpoints back to the original scale after statistical analysis facilitates clinical interpretation.

5.4.3 Data Modifications (or Derived Variables)

Data modifications are sometimes used to create a new variable for analysis, for example, "change from baseline," « Area under the curve », or ratio of two different variables. Such derivations should be detailed in the protocol and/or statistical report. For complex derivations examples should be supplied.

5.5 Estimation, Confidence Intervals and Hypothesis Testing

5.5.1 Estimates of Treatment Effects

It is important to estimate the size of the difference between treatments, in order to assess whether the effect is clinically relevant. This point estimate could be the mean of the observed difference for normally distributed variables, the odds ratio for proportions, or other appropriate summary statistics.

5.5.2 Confidence Intervals

Estimates of treatment effects should be accompanied by confidence intervals, whenever possible, and the way in which these will be calculated should be described.

5.5.3 Significance Tests

The reporting of precise p-values (e.g.'P=0.034') should be preferred, rather than exclusive reference to critical values (e.g. 'P<0.05'). It is also important to clarify whether one- or two-sided tests are used.

5.5.4 Statistical methods

The particular statistical model chosen should reflect the current state of veterinary knowledge and statistical science about the variables to be analysed. All effects to be fitted in the analysis (for example in ANOVA models) should be fully specified, and the manner, if any, in which this set of effects might be modified in response to preliminary results should be explained. The same considerations apply to the set of covariates fitted in an analysis of covariance (See also Section 5.7.).

5.5.5 Bayesian methods

This guideline largely refers to frequentist methods when discussing hypothesis testing and/or confidence intervals. However, the use of Bayesian approaches may be considered when the reason for their use is clear and when the resulting conclusions are sufficiently robust compared to alternative assumptions.

5.6   Adjustment of Type I Error and Confidence Levels

When multiplicity is present, the usual frequentist approach to the analysis of clinical trial data may require an adjustment to the type I error. Multiplicity may arise, for example, from multiple primary variables (see Section 2.2.2), multiple comparisons of treatments, repeated evaluation over time and/or interim analyses (see Section 4.6). Methods to avoid or reduce multiplicity are sometimes preferable when available, such as the identification of the key primary variable (multiple variables), the choice of a critical treatment contrast (multiple comparisons), the use of a summary measure such as 'area under the curve' (repeated measures). In confirmatory analyses, any aspects of multiplicity which remain after steps of this kind have been taken should be identified in the protocol; adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan.

5.7   Subgroups, Interactions and Covariates

5.7.1 Influence of Covariates

The primary variable(s) is/are often systematically related to other influences apart from treatment. For example, there may be relationships to covariates such as gender, breeding conditions, or prognostic factors. Or there may be differences between specific subgroups such as those treated in different countries. In some instances an adjustment for the influence of covariates or for subgroup effects is an integral part of the statistical section of the protocol. Pre-study deliberations should identify those covariates and factors expected to have an important influence on the primary variable(s), and should consider how to account for these in the analysis in order to improve precision and to compensate for any lack of balance between treatment groups. Special attention should be paid to the role of baseline measurements of the primary variable(s).

5.7.2 Post-randomisation Adjustment

It is not advisable to adjust the main analyses for covariates measured after randomisation where they may be affected by the treatments. This does not include protocol-defined covariates that are measured daily, such as ambient temperature.

5.7.3 Primary Analysis

When the potential value of an adjustment is in doubt, it is advisable to nominate the unadjusted analysis as the one for primary attention, the adjusted analysis being supportive.

5.7.4 Interactions

The treatment effect itself may also vary with subgroup or covariate. For example, the effect may decrease with age, or may be larger in a particular diagnostic category. In some cases such interactions are anticipated, and hence a subgroup analysis, or a statistical model including interactions, is part of the statistical confirmatory analysis, rather than the exploratory analysis. In most cases, however, subgroup and interaction analyses are exploratory and should be clearly identified as such. They should explore the uniformity of any treatment effects found overall. In general, such analyses should proceed first through the addition of interaction terms to the statistical model in question, complemented by additional exploratory analysis within relevant subgroups of subjects, or within strata defined by the covariates. When exploratory, these analyses should be interpreted cautiously; any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted.

5.8   Integrity of Data and Computer Software

The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software used both for data management (data entry, storage, verification, correction and retrieval) and also for processing the data statistically. Data management activities should therefore be documented; it may be helpful to describe basic data management procedures in specific SOPs. The computer software(s) used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

6.   EVALUATION OF SAFETY AND TOLERANCE

Safety variables (both for pharmaceuticals and biologicals) are evaluated, where appropriate, according to the same statistical principles as clinical efficacy endpoints. One additional requirement is the need to refer to normal ranges for safety variables when interpreting the results of any statistical analysis. In general, the incidence of adverse events within a clinical trial is too low to allow a meaningful statistical analysis. The use of descriptive summary statistics and graphs should also be considered.

7.   REPORTING

Primary data should normally be provided as part of the reporting process and sufficient information, summary tables, and reports on analyses be included in the statistical output of the report so that the reviewer can easily review the study report from raw data to the final inferential claims. In particular, a reviewer should be able to check a statistical procedure by taking the raw data, applying the statistical method and software to arrive at the same conclusions presented in the report.

The data analysis should proceed according to the statistical section of the protocol. Particular attention should be paid to any differences between the planned statistical analysis and the actual analysis. An explanation should be provided for deviations from the planned analysis.

All experimental units entering the trial should be accounted for in the report, whether or not they are included in the analysis. All reasons for exclusion of any experimental unit from the analysis should be documented. The measurements of all important variables should be accounted for at all relevant time-points.

The effect of all losses of experimental units or data, withdrawals from treatment and major protocol deviations on the main analyses of the primary variable(s) should be considered. Experimental units lost to follow-up, withdrawn from treatment, or with a severe protocol deviation should be identified, and a descriptive analysis of them provided, including the reasons for their loss and its relationship to treatment and outcome.

Descriptive statistics form an indispensable part of reports. Suitable tables and/or graphical presentations should illustrate clearly the important features of the primary and secondary variables. The results of the main analyses relating to the objectives of the trial should be the subject of a descriptive presentation.

Although the primary goal of the analysis of a clinical trial should be to answer the questions posed by its main objectives, new questions based on the observed data may emerge. Additional and perhaps complex statistical analysis may be the consequence. This exploratory work should be distinguished in the report from work that was planned in the protocol.

Chance may lead to unforeseen imbalances between the treatment groups in terms of baseline measurements not pre-defined as covariates, but having some prognostic importance nevertheless. This may be dealt with by showing that a subsidiary analysis that accounts for these imbalances reaches essentially the same conclusions as the planned analysis. If this is not the case, the effect of the imbalances on the conclusions should be discussed.

Subsidiary analyses are sometimes carried out when it is thought that the treatment effect may vary according to some other factor or factors. An attempt may then be made to identify subgroups of experimental units for whom the effect is of particular importance. Such exploratory analysis must be properly assessed and should therefore be reported critically.

Statistical judgement should be brought to bear on the analysis, interpretation and presentation of the results of a clinical trial. To this end the trial statistician should be a member of the team responsible for the study report and should approve the final report.

8.   GLOSSARY

All Randomised Cases Dataset

Dataset that includes all cases actually enrolled in the study, randomised to a treatment group and receiving at least one dose of study medication.
Bayesian Approaches

Approaches to data analysis that provide a posterior probability distribution for some parameter (e.g. treatment effect), derived from the observed data and a prior probability distribution for the parameter. The posterior distribution is then used as the basis for statistical inference.

Bias (Statistical & Operational)

The systematic tendency of any factors associated with the design, conduct, analysis and evaluation of the results of a clinical trial to make the estimate of a treatment effect deviate from its true value. Bias introduced through deviations in conduct is referred to as 'operational' bias. The other sources of bias listed above are referred to as 'statistical'.

Covariate/Covariant

Secondary explanatory variable to the measured clinical variable, that likely influences the observed result. Example: Baseline (pre-treatment) levels of a clinical variable, ambient temperature. Analyses that account for the influence of these variables typically yield a more accurate representation of the true treatment effect by partitioning the raw variability.

Dichotomous variables

A special type of categorical (qualitative or discrete) variable which has only two categories, e.g. Yes/No, Present/Absent. Sometimes multiple categories for a variable, e.g. clinical scores, are dichotomised into two categories to simplify the statistical analysis. When continuous variables are dichotomised to produce, for example, the values present/absent or success/ failure this will reduce the power of any statistical comparisons.

Double-Dummy

This is a technique in which the investigator and animal owners are blinded by the systematic use of two treatments. For example, an injectable is compared to an intramammary product. Group A will be treated with both, the active injectable and a placebo intramammary whereas Group B will be treated with a placebo injectable and the active intramammary. This technique is used where blinding cannot be assured because the formulations of the two products to be compared are too dissimilar.

Dropout

An animal in a clinical trial which for any reason fails to complete the study as defined in the study protocol.

Equivalence Trial

A trial with the primary objective of showing that the response to two or more treatments differs by an amount which is clinically unimportant. This is usually demonstrated by showing that the true treatment difference is likely to lie between a lower and an upper equivalence margin of clinically acceptable differences.

Experimental Unit

The smallest unit to which the treatment is applied and forms the basic unit for the statistical analysis. For an injectable product, the experimental unit would be the individual animal. For an in-feed product, the experimental unit may be the pen of animals. (See also 'Observation unit').

Fixed effect

Explanatory variables, such as treatment group or gender, in which all levels of the factor about which inferences are to be drawn from the results of the measured clinical variable, are included in the experimental design and analysis. (See also 'Random effect')

Frequentist Methods

Statistical methods, such as significance test and confidence intervals, which can be interpreted in terms of the frequency of certain outcomes occurring in hypothetical repeated realisations of the same experimental situation.

Generalisability, Generalisation

The extent to which the findings of a clinical trial can be reliably extrapolated from the animals that participated in the trial to a broader animal population and a broader range of clinical settings.

Global Assessment Variable

A single variable, usually a scale of ordered categorical ratings, which integrates objective variables and the investigator's overall impression about the state or change in state of an animal. It has to be relevant to the primary objective of the trial.

Group Sequential Designs

These trials have one or more planned interim analyses and require stopping rules based on repeated significance testing.

Interaction (Qualitative & Quantitative)

The situation in which a treatment contrast (e.g. difference between investigational product and control) is dependent on another factor (e.g. day of study). A quantitative interaction refers to the case where the magnitude of the contrast differs at the different levels of the factor, whereas for a qualitative interaction the direction of the contrast differs for at least one level of the factor.

Inter-Assessor Reliability

The property of yielding equivalent results when used by different assessors.

Intra-Assessor Reliability

The property of yielding equivalent results when used by the same assessor on different occasions.

Interim Analysis

Any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to the formal completion of the trial.

Meta-Analysis

The formal evaluation of the quantitative evidence from two or more trials of similar, but not necessarily identical experimental structure, designed to answer similar question(s).

Mixed Model

Experimental design that includes both fixed and random effects.

Multicentre Trial

A clinical trial conducted according to a single protocol, but at more than one site, and therefore, carried out by more than one investigator.

Multiplicity/Multiple Comparisons

The consequence of performing more than one hypothesis test on a data set or parameter. When multiplicity is present, the usual frequentist approach to the analysis of clinical trial data requires the use of an appropriate Multiple Comparison procedure to preserve the type I error (See **Statistical Significance**). Multiplicity can occur because of: multiple treatments, or multiple endpoints, or repeated measurements, or subgroup analyses or interim analyses.

Non-homogeneity of variance (Heteroscedasticity)

Many common statistical procedures assume the variances are homogeneous for the different treatment groups (ANOVA) or for different time points (repeated measures ANOVA) or for different values of the independent variable (regression analysis). Where the variances are non-homogeneous, transforming the data is one common way of achieving homogeneity. Also, modern statistical procedures e.g. PROC MIXED in SAS allows non-homogeneity to be modelled in the statistical analysis.

Non-Inferiority Trial

A trial with the primary objective of showing that the response to the investigational product is not clinically inferior to a comparative agent (active or placebo control). This is usually demonstrated by showing that the true treatment difference is likely to lie above a lower limit of clinically relevant differences.

Observation Unit

The smallest unit that is independently observed for a clinical sign. This is typically the individual animal, even if the treatment and analysis is based on the larger experimental unit. (See also 'Experimental unit')

Per Protocol Set (Valid Cases, Efficacy Sample, Evaluable Subjects Sample)

The set of data generated by the subset of animals which complied with the protocol sufficiently to ensure that these data would be likely to exhibit the effects of treatment, according to the underlying scientific model. Compliance covers such considerations as exposure to treatment, availability of measurements and absence of major protocol violations.

Power

The power of a statistical test $(1-\beta)$ is the probability that it correctly rejects the null hypothesis when it is false. The probability of erroneously not rejecting a false null hypothesis, is referred to as the type II error (β). Power estimation relies on assumptions of the distributions of the variables tested, on the size of the effect to be detected, on the design, and on the sample sizes.

Random Effect

Explanatory variables, such as site in a large multicentre study, in which only a subset of the possible levels of the factor are included in the experiment. (See also 'Fixed effect').

Randomisation

The process of assigning study animals (or groups of study animals) to treatment or control groups using an element of chance to determine the assignments, in order to reduce bias.

Robustness

Robustness in the results of a statistical analysis implies that the results are insensitive to small deviations in the assumptions on which the analysis is based.

Statistical Analysis Plan

A statistical analysis plan is a document that contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

Statistical Significance

The difference between two treatment group mean values (*e.g.* investigational product versus placebo) is statistically significant if the probability of such a difference occurring by chance alone is less than an agreed value (usually 0.05, i.e. the 5% level of significance, and also referred to as the type I error, probability of erroneously rejecting the null hypothesis). Thus, it is a measure of whether a difference is likely to be real, but it does not indicate whether the difference is small or large, important or trivial.

Study/Trial definition

For the purpose of this guideline, study and trial are synonymous. A study may be conducted on individual animals at a single site or at multiple sites (See **Multicentre Trials**).

Superiority Trial

A trial with the primary objective of showing that the response to the investigational product is superior to a comparative agent (active or placebo control). This may be demonstrated by using confidence limits and/or hypothesis tests to show that the true treatment difference is likely to be greater than zero.

Surrogate Variable

A variable that provides an indirect measurement of effect in situations where direct measurement of clinical effect is not feasible or practical.

Treatment Effect

An effect attributed to a treatment in a clinical trial. In most clinical trials the treatment effect of interest is measured by comparing (or contrasting) two or more treatments.

Trial Statistician

A statistician who has a combination of education/training and experience sufficient to implement the principles in this guidance and who is responsible for the statistical aspects of the trial.

Type I Error

See **Statistical Significance**

Type II Error

See **Power**