EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

23 July 2021
EMA/CVMP/EWP/81976/2010-Rev.1*
Committee for Medicinal Products for Veterinary use (CVMP)

Guideline on statistical principles for clinical trials for veterinary medicinal products (pharmaceuticals)

Adoption of guideline by Committee for Medicinal Products for Veterinary Use (CVMP)	12 January 2012
Revised guideline (revision 1) agreed by Efficacy Working Party (EWP-V)	June 2021
Adopted by CVMP	15 July 2021
Date for coming into effect	28 January 2022

This guideline replaces the guideline on statistical principles for clinical trials for veterinary medicinal products (pharmaceuticals) (EMA/CVMP/EWP/81976/2010).

Keywords	***Veterinary medicinal products, statistical principles, pre-clinical studies, clinical trials, study design, study conduct, data analysis, reporting***

*The current revision consists of administrative changes made in order to align the guideline to the new definitions and terminology provided by Article 4 of Regulation (EU) 2019/6. The references to the legislation applicable and other scientific guidelines have also been updated. As no changes were made to the scientific content, no concept paper and no public consultation were deemed necessary.

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

Guideline on statistical principles for clinical trials for veterinary medicinal products (pharmaceuticals)

Table of contents

Executive summary

This revised note is intended to provide guidance on the statistical principles to be considered in the design, conduct, analysis and evaluation of clinical trials to demonstrate efficacy and/or safety of an investigational veterinary pharmaceutical product in animals. The guideline is basically similar to its counterpart in human medicine (Note for Guidance on Statistical Principles for Clinical Trials, CPMP/ICH/363/96) and addresses, in addition, specific veterinary issues. A number of issues relating to hypothesis testing (superiority, non-inferiority), confidence intervals for response variables, power calculations and other statistical methods have been identified by regulators in the recent years that would need clearer guidance. Therefore, the guideline has been updated accordingly.

1. Introduction (background)

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

This guideline is intended to give direction to sponsors in the design, conduct, analysis, and evaluation of clinical trials of an investigational veterinary pharmaceutical product in the context of its overall pre-clinical and clinical development. The guidance will also assist scientific experts 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. Veterinary pharmaceutical products are often covered by specific guidelines (e.g. anthelmintics, ectoparasiticides) where specific efficacy thresholds are set, as well as numbers of animals – in that case, those guidelines have to be respected in first line.

2. Scope

The focus of this guideline is on statistical principles. It assumes the use of validated and clinically meaningful parameters as well as relevant inclusion and exclusion criteria. It does not address the use of specific statistical procedures or methods. Specific procedural steps to assure that principles are properly implemented are within 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 VICH GL9.

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.

If it is anticipated that it is not possible to follow the required study design to allow meaningful statistics (e.g. in case of minor use/minor species or for particular animal welfare issues), applicants

should consider alternative methods of efficacy demonstration, or seek scientific advice prior to the start/planning of the trial.

The protocol and subsequent amendments should be signed 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.

3. Legal basis

This note for guidance has to be read in conjunction with Regulation (EU) 2019/6. Applicants should also refer to other relevant European and VICH guidelines, including those listed under "References".

4. Overall considerations on design

4.1. Type of clinical trial

For the purpose of this guideline, the terms "study" and "trial" are used synonymously for a single scientific experiment conducted in a target species.

The broad aim of the process of clinical development of a veterinary medicinal product is to determine 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 a series of clinical trials, each with its own specific objectives. This should be specified in a 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 two categories: exploratory (pilot) or confirmatory (pivotal) trial. In confirmatory trials, the option may exist to use the data for 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.

4.1.1. 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 pilot studies, i.e. precursors to confirmatory trials.
- also have clear and precise objectives; however, hypotheses may not be predefined.
- allow data exploration during analysis, which may lead to hypotheses to be tested in future studies.
- contribute to the proof of concept, but cannot be the sole basis of formal proof of efficacy or in-use safety.

Exploratory trials are often non-GCP and, therefore, are not in the main focus of this guideline. Nevertheless, their conduct should have quality, be ethical and pre-planned.

4.1.2. Confirmatory trial

Confirmatory trials can concern dose determination and dose confirmation studies as well as controlled clinical trials.

Confirmatory trials are carried out in conformity with the VICH GL9 (Good Clinical Practice) which embodies the following points:

Confirmatory trials

- are controlled.
- have suitable randomisation and blinding procedures.
- 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 relate 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.
- use validated or well-established, recognised parameters that are clinically relevant.
- are planned to allow for robust conclusions.

Usually, the weight of evidence from a single confirmatory trial is sufficient, and there is no need for replication of the results. But if there are weaknesses with respect to internal or external validity, clinical relevance, statistical significance, data quality, or internal consistency, a second confirmatory trial should be performed.

4.2. Study scope

4.2.1. 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 clinical trials are undertaken, the study animals (selected from the study population) should more closely mirror the intended population (target 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.

4.2.2. Primary and secondary variables

The primary variable, also known as the primary endpoint variable, should be the variable capable of providing the clinically most 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. The variable should be a reliable and validated or well-established measure derived from experience in previous studies or in published scientific literature.

The primary variable 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 treatment response 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 in knowledge of the data 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.

Composite variables

In some situations it may be useful to combine 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 (including a description how to deal with missing values), and an interpretation of the resulting scale should be provided in terms of the size of a clinically relevant benefit. This interpretation is of particular importance, since there might be different outcomes with the same composite value but of different relevance. Combining multiple measurements addresses the multiplicity problem (see Section 7.7) 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 as secondary variables.

Rating scale variables

Rating scales (such as "none/mild/moderate/severe" classifications or visual analogue scales) – in particular as primary variable – should be used with particular care. Content validity, intra-assessor reliability, as well as inter-assessor reliability should be addressed. Furthermore, responsiveness for appropriately detecting relevant differences is of importance, in particular, if the study aims at proving similarity. When possible, assessors should be trained to enhance the reliability of the rating scales. These should be sufficiently validated, e.g. the number of categories comprising treatment success and treatment failure should be balanced. The degree of validity of a rating scale determines the weight of the results in an overall assessment of risk-benefit balance. It should be noted that performing arithmetics with ratings (e.g. calculating sums, differences, averages or percentages of rates or rate changes) is generally considered at least problematic. Treating ordered categorical data as if it were continuous would imply linearity of observations and equal spaces which is rarely true. Statistical tests performed using rating scales should be appropriate for this type of data.

Global assessment variables

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 detailed and precise description in the protocol (see Definitions).

Multiple primary variables

In some 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 could provide sufficient basis for an efficacy and/or safety claim. Rules for the interpretation of the results, e.g. whether one or all variables have to be significant for the study, have to be pre-specified in the protocol. The influence of the use of multiple primary variables on type I and type II errors and, hence, also on the sample size should be discussed.

The primary hypothesis or hypotheses should be clearly stated with respect to the primary variables identified, and the approach to testing these hypotheses described.

Surrogate variables

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. Such surrogate variables have to be validated in order to allow for confirmatory conclusions.

Categorised variables

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.

Time-to-event variables

There are particular situations where time-to-event data are of interest. For example, time-to-event data could be the time span from randomization to death, or to tumour progression in oncology, but also to a "positive" event like recovery. In studies dealing with such data often a substantial proportion of losses to follow-up (e.g. by treatment withdrawal or death unrelated to treatment) has to be expected. To statistically analyse such studies, methods of 'Survival Analysis' should be used.

To obtain reliable results, the event in question as well as the rules for censoring data should be uniquely defined in advance.

There might be unequal proportions of censored data in different treatment groups – patients getting worse could be more likely to drop out of the study; such systematic censoring might bias the results of comparisons. The potential influence of this bias on the trial results should be described.

4.3. Design techniques to avoid bias

Random errors lead to low precision – they can be kept small by increasing sample sizes or at least their size can be estimated by confidence intervals. On the contrary, biases, i.e. systematic errors, may lead to low validity of results – they distort measures of association in any (possibly unknown) direction. While random errors influence the width of confidence intervals, biases could shift them to an incorrect location.

Biases could arise from the design, during the conduct or during the analysis of a clinical trial. There are different types of bias – the main ones are:

- Selection bias: arises from the way the animals are selected for enrolment into the study resulting in systematic differences between comparator groups in prognosis or responsiveness to treatment.
- Performance bias: arises when there is a systematic difference in care or concomitant treatment between groups.
- Detection/assessment bias: arises from errors in measuring caused by, e.g., differential misclassification or accuracy of information between different groups (in particular, the risk of this type of bias is present if there is no blinding).
- Attrition bias: is caused by different amounts of protocol deviations or by different numbers of withdrawals between different groups.
- Reporting bias: Selective reporting of the endpoints according to their results.
- Confounding bias: might occur when third factors are linked to the outcome of interest and unevenly distributed between the study groups.

Efforts should be undertaken to avoid or at least minimize all kinds of bias. If there are known confounders, stratification and/or adjustment techniques should be used. The study population should be as close as possible to the randomised population, and exclusions should be justified. The two most important design techniques to reduce bias in clinical trials are blinding and randomisation including allocation concealment (see Section 4.3.1. and 4.3.2). These techniques should always be considered when designing clinical trials to support an application for marketing authorisation.

4.3.1. 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.

Randomisation and allocation concealment (i.e. keeping investigators and animal owners unaware of upcoming treatment assignments – without allocation concealment randomisation might become corrupted; note that allocation concealment is also possible in non-blinded trials) help to avoid possible bias in the selection and allocation of subjects arising from the predictability of treatment assignment.

The randomisation method 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 methods. The most common randomisation methods are simple randomisation, stratified randomisation and block randomisation.

The randomisation process should be reproducible, but allocation concealment (see above) should be ensured.

Although simple 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. It is possible to randomly vary block lengths to increase the level of randomness and reduce predictability.

Stratification by prognostic factors is a useful randomisation method: It ensures that the numbers of animals in each comparator group are closely balanced within each stratum and can minimize confounding biases. Stratified randomisation is of particular usefulness if there are subgroups anticipated to be clinically relevant.

In multicentre trials the randomisation procedures should ideally 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.

To ensure allocation concealment, the allocation sequence and 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.

4.3.2. Blinding

Blinding is a procedure to reduce potential study bias in which designated study personnel are kept uninformed on the treatment assignment(s).

The optimum is a double-blinded trial in which the investigator and trained personnel involved in the treatment or clinical evaluation, the owner of the study animals, or any other persons associated with administering 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, who by the nature of their work and responsibilities have to remain unblinded.

Difficulties in achieving double blinding may arise particularly where the treatments are of a different nature. One way of achieving double blinding conditions under these circumstances is to use a "double dummy" technique.

If a double blinded trial is not feasible, it should be justified and a single blinded trial should be considered. If a study is to be conducted with single 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 single blinding achieved by denying personnel involved with clinical assessments access to treatment information.

In single 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 or for a pen of grouped animals) 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.

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.

5. Types of Study Design

5.1. Study configuration

5.1.1. Control groups

The choice of control group is always a critical decision in designing a clinical trial; their major purpose is to allow discrimination of patient outcomes caused by the test treatment from outcomes caused by other factors. Therefore, test and control groups should be similar with regard to all baseline and on-treatment variables that could influence outcome, except for the study treatment.

There are different types of control, each of which is appropriate in some circumstances, but none is usable or adequate in every situation: There are different types of concurrent controls (i.e. chosen from the same population as the test group and treated in a defined way as part of the same trial that studies the test treatment, and over the same time period): placebo, no treatment, different dose(s) of the test product, and active (positive) control. In addition to these internal control groups, there are external controls, e.g. historical controls, baseline controls, or patients treated at the same time but in another setting. External controls are considered less appropriate as they generally differ from the test group in more factors than just the treatment (for more details see ICH Topic E 10). The use of external controls should be avoided.

5.1.2. Type of design

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).

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 2x2 cross-over design is commonly used in veterinary clinical trials to demonstrate the bioequivalence of two formulations of the same medication (see Guideline on the Conduct of Bioequivalence Studies for Veterinary Medicinal Products (EMA/CVMP/016/2000)).

Main advantages of a cross-over design are the reduced need in animals to participate in the study and the reduction of the influence of confounding covariates as each animal serves as its own control. However, there are also some problems specific for this design, e.g. possible sequence and carry-over effects, discontinuations and the difficulty of assigning adverse events to the correct treatment.

Factorial designs

In a factorial design, two or more factors are evaluated simultaneously in the same set of subjects by the use of varying combinations of the treatments. The simplest design is the 2x2 factorial design in which study animals are randomly allocated to one of the four possible combinations of factors (for example factor "age" with levels "young" and "old", and factor "gender" with levels "male" and "female" these possible combinations are "young male", "young female", "old male", and "old female"). Factors may also be treatments in a dose finding study for a combination product, with application of distinct doses as levels. In many cases this design is used for the specific purpose of examining the interaction of two (or more) factors. 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 two treatments, in particular, if the treatments are likely to be used together.

In addition, factorial designs might be used to make efficient use of trial subjects by evaluating two treatments in the same subjects in the absence of interaction.

5.1.3. Multicentre trials

In general, only pre-clinical studies or exploratory clinical trials can be carried out at a single site.

Multicentre clinical 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 timeframe. 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) clinical trial to provide a better basis for the subsequent generalisation of its findings. This arises from the possibility of recruiting the animals 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 clinical trial might sometimes be conducted in a number of different countries in order to facilitate generalisation even further.

If multicentre clinical trials 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 treatment effects and variances do not differ between centres. 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 study design should generally aim to achieve the same distribution of animals to treatments within each centre and good trial management should maintain this design objective.

Centre effects and treatment-by-centre interactions should be explored as random effects, since such effects may affect the generalisation of the conclusions. Marked treatment-by-centre interaction may be identified by graphical display of the results of individual centres or by analytical methods, such as a significance test of the interaction. In the absence of a true treatment-by-centre 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 treatment-by-centre interaction the interpretation of the main treatment effect is controversial. In any case, the strategy relating to inclusion of interaction terms has to be pre-defined in the statistical protocol.

Problems may arise when there are a few large-sized centres dominating the trial compared to the included small-sized centres. The possible impact of unbalanced centre sizes has to be assessed appropriately.

5.2. Type of comparison

All studies should be designed to control the risk of drawing wrong conclusions. A p-value estimates the probability of the type I error, i.e. the probability of erroneously rejecting the null hypothesis and accepting the alternative hypothesis. Therefore, the alternative hypothesis should state what the test is aiming at to demonstrate, while the null hypothesis should state the opposite. In general, keeping the probability of the type I error low increases the probability of the type II error, i.e. the probability of erroneously not rejecting the null hypothesis (see also List of Definitions).

The type I error rate is generally set to 5% two-sided (or 2.5% one sided) for all types of comparisons, i.e. α=0.05 and 95% confidence intervals are used for statistical inference.

5.2.1. Trials to show superiority

Superiority trials are designed to detect a significant difference between two or more treatments, hence the null hypothesis should state equality and the alternative hypothesis should state difference between the treatments. When the test is significant (two-sided p-value < 0.05 or one-sided p-value <0.025) one can reject the null-hypothesis and conclude that there is evidence of a difference between the treatments tested. If the test does result in non-significance we can only conclude that there is no evidence of any difference between the treatments. 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. Superiority should be demonstrated for the full analysis set (see Section 7.2.4).

A successful superiority trial shows a statistically significant difference between the test and the control group or a statistically significant dose-response relationship. Treatment differences should be provided in the study report, and the clinical relevance of the observed effects, and the additional benefit in relation to possible adverse effects should be discussed.

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

5.2.2. 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. A non-inferiority or equivalence test is aiming at demonstrating non-inferiority or equivalence between two treatments. The null hypothesis should state inferiority or non-equivalence, respectively, and the alternative hypothesis should state non-inferiority or equivalence, respectively.

The use of a superiority test for proving non-inferiority or equivalence, i.e. the conclusion of non-inferiority or equivalence from the non-rejection of the null hypothesis of no difference, is never acceptable, as this would mean the interchange of type I and II errors and in general one does not control the type II error (refer to the definition list). Thus, the lack of demonstrating a significant difference is not the same as concluding that the two treatments are equally good.

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.

Equivalence trials

Bioequivalence trials fall into the former category (more details are given in a specific guideline: Conduct of Bioequivalence Studies for Veterinary Medicinal Products (EMA/CVMP/016/2000), as amended). In some situations, clinical equivalence trials may also be undertaken for other regulatory reasons such as demonstrating the clinical equivalence of a generic product to an authorised product when the compound is not absorbed and therefore not present in the blood stream.

In an equivalence trial, the relevant null hypothesis is "The response to test treatment is at least δ_1 lower or at least δ_2 higher than the response to control treatment", and the trial is targeted at rejecting this in favour of the alternative hypothesis "The responses to test and control treatment differ at most by δ_1 or δ_2, respectively" (the margins δ_1 and δ_2 might be equal, but they need not).

For the positive 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 the entire confidence interval for the difference of treatment responses falls within the equivalence margins of the theoretical interval [-δ1; δ2]. 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. Generally, the confidence interval should be the two-sided 95% confidence interval; alternatively, the two simultaneous one-sided tests should be at the 2.5% level. One generally accepted exception from this rule is the use of 90% confidence intervals in bioequivalence studies.

Non-inferiority trials

For non-inferiority trials, positive control trials are designed to show within a predefined non-inferiority margin that the efficacy of an investigational product is not worse than that of the comparator; such trials are a one-sided version of equivalence trials. A non-inferiority 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 non-inferiority margin, depending on the criteria chosen, is the only one needed – non-inferiority is inferred when the entire confidence interval for the difference of treatment responses lies above (or below, respectively) the margin. 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 (or upper, respectively) non-inferiority margin versus the alternative that the treatment difference is greater than the lower (or upper, respectively) non-inferiority margin. Generally, the confidence interval should be the one-sided 97.5% confidence interval; alternatively, the one-sided test should be at the 2.5% level.

For both, non-inferiority and equivalence trials, the point estimates for the treatment difference should be provided in the study report. However, the decision on non-inferiority or equivalence will be based solely on the confidence intervals (as described above), as these account for the possible imprecision of the point estimates due to small sample sizes and/or high variability (see also Section 7.6.2).

Non-inferiority and equivalence trials should be designed and conducted in a way to demonstrate a recognised level of efficacy of the comparator product. Ideally, 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 comparator. There are well known limitations associated with the use of the positive 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 leading to the necessity of substantiating the results from outside the study. Therefore, comparators should be chosen with care. A suitable comparator would be a veterinary medicinal product authorised in the EU, for which 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 positive control study.

An important issue is assay sensitivity (see Definitions). If a trial is intended to demonstrate efficacy by showing a test treatment to be non-inferior or equivalent to an comparator, but the trial is lacking assay sensitivity, an ineffective treatment might be interpreted to be non-inferior and could lead to an erroneous conclusion of efficacy.

Equivalence (or non-inferiority) trials are not conservative in nature, so that flaws in the design or conduct of the trials will tend to bias the results towards a conclusion of equivalence. Thus, the use of the full analysis set may bias the results because protocol deviations might blur treatment differences. Generally, no primary analysis set can be defined a priori, thus the full and the per-protocol analysis sets should be considered co-primary (see Section 7.2.4).

Equivalence / non-inferiority margin

The equivalence / non-inferiority margins should be specified in the protocol: These margins are the largest differences of the particular endpoint which can be judged as being clinically acceptable. The selection of the equivalence or non-inferiority margin(s) should provide assurance that the investigational product has a clinically relevant effect greater than zero.

It is not appropriate to define the margin(s) as a fixed proportion of the parameter difference between comparator and placebo – if the reference product has a large advantage over placebo, this does not mean that large differences are not relevant; it just means that the reference product is very efficacious.

It is also not appropriate to define the margin(s) as a fixed proportion of the inter-individual parameter variability of the reference product – if there are large differences between the effects of the reference product to different individuals, this does not mean that for a single individual a large difference is not relevant.

Furthermore, the choice of the margin(s) should be independent of considerations of statistical power – as the size of a clinically relevant difference is not altered by the dimension of the study, a small sample size is not a justification for a wider confidence interval.

If the endpoint in an equivalence or non-inferiority test is a dichotomous variable, one should keep in mind that a margin for a success rate always corresponds with a margin for a failure rate (and vice versa). Furthermore, for a proportion near 50%, a difference by a certain number of percent points has not the same meaning as the same difference for a proportion near 0% or near 100%, e.g. a difference in success rates between 65% and 55% might not be of the same importance as the same absolute difference between 95% and 85% as the latter one implies a three-fold non-success rate. Since confidence intervals for proportions become shorter for extreme proportions, margins should become smaller, too. For equivalence or non-inferiority tests of rates, the use of margins in terms of odds ratios could be considered.

Switch between superiority and non-inferiority trials

There is the possibility to switch between non-inferiority and superiority trials: If in a trial planned as non-inferiority the confidence interval for the treatment effect not only lies entirely above (or below, depending on the criteria chosen) the non-inferiority margin but also above (or below, respectively) zero then there is evidence of superiority in terms of statistical significance; in this case it is acceptable to calculate the p-value for rejecting the null hypothesis of no difference and to interpret the trial as a superiority trial. When superiority to the comparator is claimed, the clinical relevance of the determined difference and the additional benefit in relation to possible adverse effects should be discussed.

If a superiority trial fails to detect a significant difference between treatments, there may be interest in the lesser objective of establishing non-inferiority. Such a downgrading of the objective is only acceptable if its eventuality is stated in the study protocol together with an acceptable, prospectively defined margin for non-inferiority. Therefore, in any superiority trial where non-inferiority may be an acceptable outcome for licensing purposes, it is prudent to specify a non-inferiority margin in the protocol in order to avoid the serious difficulties that can arise from later selection.

Note that the different study populations have different emphasis in non-inferiority and superiority trials (see Section 7.2.4); this has to be accounted for when switching from non-inferiority to superiority or vice versa.

5.2.3. 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 dose-response relationship; 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 dose levels 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.

5.3. Group sequential designs

Group sequential designs are used to facilitate the conduct of interim analysis (see section 6.3). While group sequential designs are not the only acceptable types of designs permitting interim analysis, they are the most commonly applied ones 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. (See also the section on interim analyses (6.3)).

5.4. Experimental unit

In veterinary clinical trials there are a variety of situations where the experimental unit is not a single animal but a group of animals (e.g. a pen, room, pasture or litter), or subunits within the same animal such as udder quarters in dairy cows. For example, dogs and cats tend to be presented in a veterinary surgery individually 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 individually (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 animals are housed in a group. This occurs when individual animals within the group are able to receive individual treatments.

However, whenever possible the follow up of the clinical condition should be done at the individual animal level. If not single animals but groups are the experimental units, the use of nested analyses should be considered.

The experimental unit should be clearly specified in the protocol, since its choice has an essential impact on many statistical issues, e.g. the statistical analysis model, sample size estimations, considerations regarding variability, thus significances and confidence intervals, etc.

5.5. Sample size

The number of animals 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 animals 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. The type of the primary variable (e.g. binary or continuous), the test statistic, the type of comparison (e.g. superiority, non-inferiority or equivalence, type of statistical test, one- or two-sided test), 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 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 (see Section 7.7) of the planned tests.

In confirmatory trials, 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 for 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. If null/alternative hypotheses are stated appropriately (see Sections 5.2.1, 5.2.2), 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), since the type II error is the risk of not rejecting a false null hypothesis.

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

The sample size of an equivalence or non-inferiority trial (see Section 5.2.2) should normally be based on the objective of obtaining a confidence interval for the treatment difference that shows that the outcomes of treatments differ at most by a clinically acceptable difference. The power is usually assessed assuming that there is no difference, as it can be estimated inappropriately if the true difference is not zero. Consequently, the sample size might be estimated inappropriately (often too low).

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 decision rule when to terminate the trial, and the true treatment difference. The design of this decision 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 6.3).

5.6. Meta-analyses

Meta-analysis is the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. As such, it constitutes a relevant issue in evidence-based medicine. However, this methodology is controversially discussed.

On the one hand, there is a series of benefits:

- The probability of detecting treatment differences (i.e. the statistical power) could be increased by the inclusion of a large number of subjects.
- Results could become more precise since confidence intervals become narrower.
- Results could become more reliable since biases of single studies are reduced in their overall effect.
- Seeming discrepancies between the results of different studies could be clarified by investigating sources of inter-study variability.
- Using already available data rather than conducting another study might be preferable regarding animal welfare.

On the other hand, these benefits come along with a number of problems. The main ones are:

- "Garbage in – garbage out": A well-performed meta-analysis of badly designed or performed studies will not lead to reliable results. All studies included in the meta-analysis should be assessed for their quality or "risk of bias". Only studies of adequate quality may be included into a meta-analysis, or the single studies should be weighted according to their quality, or sensitivity analyses may be conducted to determine the effect of including studies of different qualities.
- Heterogeneity: There could be differences between the study characteristics (e.g. different study conditions, inclusion criteria, doses, definition and measurement methods of endpoints, etc.) that would make a meta-analysis impossible or, at least, should be addressed by use of appropriate statistical methods (e.g. fixed effects or random effects models).
- Selection bias: The subjective choice of which studies to include and exclude might bias the results. Therefore, the approach for selection and exclusion of studies should be as objective as possible and must be clearly defined in advance. In addition, publication bias might occur if a meta-analysis is relying on published studies alone, as studies with significant results are more likely to be published than those with negative results, and this may bias the findings. The possible consequences of publication and selection bias on the overall results should be assessed.

Furthermore, in contrast to a standard clinical trial for which at first a question is formulated and then data are collected and analysed, in a meta-analysis generally the question is formulated in awareness of at least some of the data and results; this might lead to a bias in the direction of false positive results.

As each study, a meta-analysis has to be planned carefully. Clear and objective criteria to select studies for inclusion and their weighting have to be defined in advance and described in the meta-analysis study plan. The study report of the meta-analysis should not only list the single studies that were included but also those ones that were checked for eligibility but not included, and the particularly applied criteria should be specified.

Statistical methods have to be appropriately chosen and described in advance.

Meta-analyses including the raw data of each single study are generally considered more reliable, since no information is getting lost, than meta-analyses including summary data only.

6. Study Conduct

6.1. 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 that regular violations of the entry criteria are occurring, or if 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.

6.2. 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.

6.3. Interim analysis

There are many recognized reasons for performing interim analyses, that is, comparing treatment arms at any time prior to formal completion of a trial, generally aiming at adapting the study design:

- The assumptions that underlie the original sample size estimation might be made on preliminary information – in this case, an interim analysis may be needed to revise the assumptions and re-estimate the sample sizes.
- In the course of the study, clear evidence of efficacy or futility may emerge, or serious adverse events may occur to an unacceptable extent – in this case an interim analysis might support a decision on early stopping the trial.
- Ineffective or unsafe study arms – in particular, ineffective or unsafe dose groups in dose finding studies – could be dropped as the consequence of an interim analysis; or new study arms or dose groups could be added.

In addition, interim response data might be used for starting follow-up studies sooner, for planning accompanying studies or for portfolio planning.

Because the number, methods and consequences of comparisons affect the interpretation of the trial's results – type I and II errors may be increased and biases may be introduced –, all interim analyses should be clearly justified, carefully planned in advance and described in the protocol. This includes a description of methods to be used for adjustment for type I error inflation, a clear system of rules on which interim results would lead to which modifications (e.g. early stopping rules, see also Section 5.3) and a discussion of the possible influence of biases on the results.

The execution of an interim analysis must be a confidential process to avoid a bias caused by the use of unblinded data and results: 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. Blinding can best be warranted if these analyses are performed by person(s) that is/are not directly involved in the study; the independence of this/these person(s) has to be made plausible. Investigators should only be informed about the decision to continue or to

discontinue the trial, or about modifications to the trial procedures that are necessary to be implemented.

As deviations from the planned procedure always bear the potential of invalidating the study results, unplanned interim analyses should be avoided. Any interim analysis that is not planned appropriately may flaw the results of a trial and possibly weaken confidence in the conclusions drawn. But special circumstances may dictate the need for an interim analysis that was not defined at the start of a trial. To keep a certain level of interpretability of the data in these cases, a protocol amendment describing the interim analysis and rules for possible consequences should be completed prior to unblinded access to treatment comparison data, and the clinical trial report should explain why the interim analysis was necessary, describe the degree to which blindness had to be broken, and provide an assessment of the potential magnitude of bias introduced as well as of the impact on the interpretation of the results.

7. Data Analysis

7.1. 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.

7.1.1. Statistical section of the study protocol

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

- Definition of the experimental unit and populations
- Methods and details of randomisation and allocation concealment
- Methods of blinding
- Definition of variables (incl. aggregated variables) and handling of missing data
- Hypothesis to be tested, and specification of the primary one(s)
- Justification of the use of one-sided tests
- Treatment effect(s) to be estimated
- Methods, assumptions on the data variability and the size of clinically relevant differences, choices of statistical power (1-ß) and significance levels as well as possibly other assumptions used in sample size estimation
- Assumptions for using the statistical analysis (e.g. test for normal distribution when using an analysis of variance)
- Planned data transformations
- Statistical model, test(s) and construction of confidence intervals
- Alternative methods to be used in case of expected problems (heteroscedasticity, non-normality, etc)
- Use of covariate(s), adjusted analyses, sensitivity analyses and planned subgroup analyses
- Significance thresholds

- Equivalence and non-inferiority margins
- Planned interim analyses; stopping rules
- Reporting of summary data
- Comparison of groups at baseline

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

7.1.2. Data capture and processing

The data capture and processing should be performed in accordance with the VICH guideline for "Good Clinical Practice". (See also Section 7.8)

7.1.3. Statistical section of the study report

In the statistical section of the clinical trial report, the statistical methodology should be clearly described. It should also describe when methodology decisions were made in the clinical trial process. See also Section 9.

7.2. 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 trials 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.

7.2.1. Full analysis set of study animals

The Intention-To-Treat (ITT) principle asserts that the effect of a treatment can be best assessed by evaluating on the basis of the intention to treat a subject (i.e. the planned treatment regimen) rather than the actual treatment given. It has the consequence that animals allocated to a treatment group should be followed up, assessed and analysed as members of that group irrespective of their compliance to the planned course of treatment. Thus, it may provide estimates of treatment effects which are more likely to mirror those observed in subsequent practice.

This principle implies that the primary analysis should include all randomised subjects. Compliance with this principle would necessitate complete follow-up of all randomised subjects for study outcomes. In practice this ideal may be difficult to achieve, for reasons to be described. In this document the term 'full analysis set' is used to describe the analysis set which is as complete as possible and as close as possible to the ITT ideal of including all randomised subjects.

There are circumstances that might lead to excluding randomised subjects from the full analysis set including the failure to satisfy major entry criteria (eligibility violations), the failure to take at least one dose of trial medication and the lack of any data post randomisation. Such exclusions should always be justified.

In some situations, it may be reasonable to eliminate from the set of all randomised subjects any subject that took no trial medication. The ITT principle would be preserved despite the exclusion of these patients provided, for example, that the decision of whether or not to begin treatment could not be influenced by knowledge of the assigned treatment group. In other situations it may be necessary to eliminate from the set of all randomised subjects any animal without data post randomisation. No analysis is complete unless the potential biases arising from these specific exclusions, or any others, are addressed.

When the full analysis set of subjects is used, violations of the protocol that occur after randomisation may have an impact on the data and conclusions, particularly if their occurrence is related to treatment assignment. In most respects it is appropriate to include the data from such subjects in the analysis, consistent with the ITT principle.

7.2.2. 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.

Note that the need to exclude a substantial proportion of subjects from the per-protocol analysis might throw doubt on the overall validity of the trial.

7.2.3. Safety Dataset

All animals that received even one dose of study medication should comprise the safety dataset. They should be included into the analysis of target animal safety variables according to the treatment actually received.

7.2.4. Roles of the different analysis sets

The full 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. In general, it is advantageous to conduct analyses of both sets and to demonstrate a lack of sensitivity of the principal trial results to alternative choices of the set of subjects analysed. When the full analysis set and the per-protocol set lead to essentially the same conclusions, confidence in the trial results is increased, else differences should be discussed, and the final conclusion should be justified.

The full analysis set and the per-protocol set may play different roles in superiority trials and in equivalence or non-inferiority trials. In superiority trials, the full analysis set should be used in the primary analysis as it tends to avoid over-optimistic estimates of efficacy (non-compliers will generally diminish the estimated treatment effect). To the contrary, in non-inferiority or equivalence trials the use of the full analysis set is generally not conservative, as protocol deviations might blur treatment differences and thus might bias the results towards equivalence. Therefore, its role should be considered very carefully. In non-inferiority and equivalence trials, the full and the per-protocol analysis sets have equal importance and their use should lead to similar conclusions for a robust interpretation in the analysis of the primary outcome(s).

7.3. Comparison of baseline values

Baseline data should be collected from participants prior to randomisation. Demographic data, data on prognostic indicators and baseline values for relevant endpoint variables should be collected. These data are used to provide information on the study population, to assess the success of randomisation in producing comparable groups and to provide baseline values for key response variables, where needed. Baseline data should be summarised for each group separately. As any imbalance between groups in a randomised study is due to chance, it is not appropriate to conduct hypothesis tests to compare groups.

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).

7.4. 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 avoid missing 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.

Imputation techniques, ranging from the carrying forward of the last observation to the use of complex mathematical models, may be used in an attempt to compensate for missing data. The use of any of these strategies should be described and justified in the statistical section of the protocol, and the assumptions underlying any mathematical models employed should be clearly explained. It is also important to demonstrate the robustness of the corresponding results.

Special problems arise in connection with subjects withdrawn from treatment after receiving one or more doses and providing no data after this, and subjects otherwise lost to follow-up. Measurements of primary variables made at the time of the loss to follow-up of a subject for any reason, or subsequently collected in accordance with the intended schedule of assessments in the protocol, might be valuable for the assessment of study outcome; subsequent collection is especially important in clinical trials where the primary variable is mortality or serious morbidity. The intention to collect data in this way should be described in the protocol whenever possible.

The decision on whether to keep or to exclude outlier values should be discussed; one should be aware that exclusion of outlier values always is a possible source of bias. For the main endpoint, two separate analyses may be provided, with and without outlier(s), and the differences between their results discussed.

7.5. Data transformation/modification

Transformation of data is often necessary to confirm 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, etc.). Transformations (e.g. square root, logarithm, etc.) should be specified in the protocol and a rationale provided wherever possible.

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 (e.g. plasma concentrations have to be log-transformed prior to ANOVA in bioequivalence studies). This can sometimes lead to the use of unplanned transformations. In this case, a justification should be given in the report.

Transforming endpoints back to the original scale after statistical analysis facilitates clinical interpretation. Therefore, when possible this should be done.

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.

7.6. Estimation, confidence intervals and hypothesis testing

7.6.1. 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.

7.6.2. Confidence intervals

Point estimates may lack precision due to small sample sizes and/or high individual animal data variability. Therefore, to assess the precision of point estimates of treatment effects, these should be accompanied by confidence intervals, whenever possible. For example, a 2-sided 95% confidence interval for a treatment difference is interpreted in such a way that with a probability of 2.5% the difference is larger than the upper confidence limit, and with the same probability the difference is smaller than the lower confidence limit. In particular, a positive (or negative) point estimate with a confidence interval containing zero is considered not significant. A point estimate near zero is not considered sufficient to demonstrate the absence of a clinically relevant difference, if one or both confidence limits are of unacceptably large absolute values.

7.6.3. Significance tests

To allow for an appropriate assessment of the risk-benefit balance, the reporting of precise p-values (e.g. "P=0.034") is preferred, rather than exclusive reference to critical values (e.g. "P<0.05"). Generally, two-sided tests should be performed unless the use of one-sided tests is clearly justified.

Statistical significance and clinical relevance should not be confused. P-values should always be accompanied by estimations of confidence intervals for the effect sizes to allow the discussion of the results' clinical relevance.

7.6.4. 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 7.8).

7.6.5. Bayesian methods

Because the predominant approaches to the design and analysis of clinical trials have been based on frequentist statistical methods (see definitions), this guideline largely refers to the use of these methods when discussing hypothesis testing and/or confidence intervals. This should not be taken to imply that other approaches are not appropriate: the use of Bayesian (see definitions) and other approaches may be considered when the reasons for their use are clear, and when the resulting conclusions are sufficiently robust.

7.7. Adjustment of type I error and confidence levels

When multiplicity is present, the usual frequentist approach to the analysis of clinical trial data requires an adjustment to the type I error. Multiplicity may arise, for example, from multiple primary variables (see Section 4.2.2), multiple comparisons of treatments, repeated evaluation over time, subgroup analyses (see Section 7.8.2) and/or interim analyses (see Section 6.3). Methods to avoid or reduce multiplicity are always preferable when available, such as the identification of the key primary variable (multiple variables), the choice of a critical treatment contrast (multiple comparisons), or 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.

7.8. Covariates, interactions and subgroups

7.8.1. 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 centres of a multicentre trial. 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: it could provide the most appropriate p-value for a treatment difference, or an unbiased estimate and confidence interval for a treatment effect. 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).

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 treatment independent covariates that are measured daily, such as ambient temperature.

When adjusted analyses are performed, the results of both, the adjusted and the unadjusted analyses, should be reported. The protocol should clearly state which of both analyses the main one is. 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.

7.8.2. Interactions and subgroup analyses

The treatment effect itself may also vary with subgroups, based on demographic, genomic or disease characteristics. In some cases such interactions are anticipated, and hence a subgroup analysis, or a statistical model including interactions, is a critical part of the statistical confirmatory analysis. In other cases, however, subgroup and interaction analyses are exploratory (and should be clearly identified as such): they should demonstrate the homogeneity of any treatment effects, or identify possible subgroups where benefit-risk is different to the full analysis set.

All subgroup and interaction analyses should be clearly justified and carefully planned in advance; their rationale should be described in the protocol. Subgroups have to be precisely defined in advance, and these definitions should not be based on factors measured after randomisation where they may be affected by treatment. A stratified randomisation to balance the treatment groups with regard to the subgroups (see 4.3.1) should be considered. Issues relating to multiplicity (see 7.7) and statistical power should be addressed – in case of extensive subgroup analyses, false positive (due to type I error inflation) as well as false negative findings (due to small subgroup sizes) are to be expected.

The statistical methods have to be chosen appropriately. In general, such analyses should proceed first through the addition of interaction terms to the statistical model in question, and only in case of a significant interaction they should be complemented by additional analyses within relevant subgroups.

Unplanned subgroup and interaction analyses should be avoided as they may flaw the results of the trial and possibly weaken confidence in the conclusions drawn. There might be circumstances dictating the need for such unplanned analyses; in this case, a protocol amendment describing the analyses as well as the statistical methods should be completed prior to unblinding, and the clinical trial report should explain the reason for the subgroup or interaction analyses and provide an assessment of the potential magnitude of bias introduced as well as of the impact on the interpretation of the results. However, these analyses should be interpreted cautiously. Any conclusion of treatment efficacy or safety (or lack thereof) based solely on exploratory subgroup analyses (planned or unplanned) is unlikely to be accepted without further proof of efficacy.

7.9. 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. The computer software(s) used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

8. Evaluation of Target Animal Safety

Safety variables collected during clinical trials may be evaluated, where appropriate, according to the same statistical principles as clinical efficacy endpoints. One additional requirement is the need to refer to appropriate reference ranges for measurable safety variables for a relevant interpretation of 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. It should be noted that performing difference tests

here seldom is meaningful since the absence of statistical significance does not imply that there is no difference.

The use of descriptive summary statistics and graphs should be considered.

9. Reporting

Primary data should be provided as part of the reporting process and sufficient information, summary tables, and reports on analyses should 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 conclusions. For each analysis, it should be reproducible which subjects in each treatment group were included, and the respective numbers should be specified. Visual presentations are considered useful – they are, therefore, strongly recommended.

The data analysis should proceed according to the statistical section of the protocol. A brief description of the assumptions made for sample size calculation as well as the statistical methods used for analysis should be included in the report. 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 animals entering the trial should be accounted for in the report, whether they are included in the analysis or not. The use of participant flow diagrams is encouraged. 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, withdrawal 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 provided as suitable tables and, whenever possible, graphical presentations which 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 graphical 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 an additional 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.

Ancillary analyses are sometimes carried out when it is thought that the treatment effect may vary according to some other factor(s). 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.

Definitions

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.

Assay sensitivity

The ability of distinguishing an effective treatment from a less effective or ineffective one; the ability of recognizing relevant differences between single measurements of a parameter. In non-inferiority or equivalence trials with two active arms only, assay sensitivity implies that if a placebo could have been included then both active treatments would have been superior to placebo. Generally, assay sensitivity is inferred from previous studies.

Bayesian approaches

In contrast to "Frequentist methods" (see there) Bayesian approaches focus on the determination of the probability, that a hypothesis is true, given the observed data.

Bias (statistical or 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 parameter 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".

Blind review

The checking and assessment of data during the period of time between trial completion (the last observation on the last subject) and the breaking of the blind, for the purpose of finalising the planned analysis.

Clinical trial

Means a study which aims to examine under field conditions the safety or efficacy of a veterinary medicinal product under normal conditions of animal husbandry or as part of normal veterinary practice for the purpose of obtaining a marketing authorisation or a change thereof.

Content validity

The extent to which a variable (e.g. rating scale) measures what it is supposed to measure.

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/Withdrawal

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 which 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").

External validity

Validity of generalisability of the conclusions from the study population to the target population and from study settings to other environments.

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. It is focused on the probability of results of a trial assuming that a particular hypothesis is true (alternatively see "Bayesian approaches").

Full analysis set

The set of subjects that is as close as possible to the ideal implied by the intention-to-treat principle. It is derived from the set of all randomised subjects by minimal and justified elimination of subjects.

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 or 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 on different occasions.

Internal validity

Internal validity refers to the extent to which one can accurately state that the test product led to the observed effect.

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.

ITT principle = intention-to-treat principle

The principle that asserts that the effect of a treatment policy can be best assessed by evaluating on the basis of the intention to treat a subject (i.e. the planned treatment regimen) rather than the actual treatment given. It has the consequence that subjects allocated to a treatment group should be followed up, assessed and analysed as members of that group irrespective of their compliance to the planned course of treatment.

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 allow 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. This is usually demonstrated by showing that the true treatment difference is likely to lie above a lower limit of clinically relevant differences.

Null hypothesis

The null hypothesis is a statement on a parameter implying that a claimed relationship does *not* hold. Generally, a study is aiming at rejecting the null hypothesis. (See also "Type I error" and "Type II error").

P-value

The probability to reject the null hypothesis although it is true, that is, the probability to erroneously conclude on the study claim stated in the alternative hypothesis (see also "Null hypothesis", "Statistical significance" and "Type I error"). Note that in a difference test, a small p-value only means that the presence of a difference is likely, but not, that this difference is important from a clinical perspective.

PP analysis set is the per-protocol analysis 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.

Pre-clinical study

Means a study not covered by the definition of clinical trial which aims to investigate the safety or efficacy of a veterinary medicinal product for the purpose of obtaining a marketing authorisation or a change thereof.

Protocol amendment

A written change or modification of the study protocol effected prior to the implementation of the protocol or execution of the changed or modified task. Study protocol amendments should be signed and dated by the investigator and sponsor and incorporated into the study protocol.

Protocol deviation

A departure from the procedures stated in the study protocol. Study protocol deviations should be recorded as a statement signed and dated by the investigator describing the deviation and the reason for its occurrence (if identifiable).

Protocol violation

A serious protocol deviation that may affect the completeness, accuracy and reliability of the study data.

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

A result is considered statistically significant if the corresponding p-value is less than the pre-defined significance level (usually 0.05). Note that a non-significant result does not imply that the null hypothesis holds but just that data are not sufficient to reject it.

Study population

The group of individuals in a study.

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.

Target population

The entire group of individuals the treatment is aiming at.

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 (α)

The error made when the null hypothesis is rejected although true. (See also "P-value" and "Statistical significance").

Type II error (β)

The error made when the null hypothesis is not rejected although not true. (See also "Power").

References

Regulation (EU) 2019/6 of the European Parliament and of the Council of 11 December 2018 on veterinary medicinal products and repealing Directive 2001/82/EC

VICH GL9: Guideline on Good Clinical Practices (GCP) (CVMP/VICH/595/1998)

ICH Topic E 9, Statistical Principles for Clinical Trials (CPMP/ICH/363/96)

ICH Topic E 10, Choice of Control Group in Clinical Trials (CPMP/ICH/364/96)

Biostatistical Methodology in Clinical Trials (Directive 75/318/EEC)

CVMP Guideline on the Conduct of Bioequivalence Studies for Veterinary Medicinal Products (EMA/CVMP/016/2000).

Guideline on the Choice of the Non-Inferiority Margin (EMEA/CPMP/EWP/2158/99)

Points to Consider on Switching Between Superiority and Non-inferiority (CPMP/EWP/482/99)

Points to Consider on Adjustment for Baseline Covariates (CPMP/EWP/2863/99)

Points to Consider on Multiplicity Issues in Clinical Trials (CPMP/EWP/908/99)

Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ, 27 March 2010, Vol 340, 698-702.