BIOSTATISTICAL METHODOLOGY IN CLINICAL TRIALS

Guideline Title	**Biostatistical Methodology in Clinical Trials**
Legislative basis	**Directive 75/318/EEC as amended**
Date of first adoption	**May 1993**
Date of entry into force	**October 1993**
Status	**Last revised May 1993**
Previous titles/other references	***Biostatistical methodology in clinical trials in applications for Marketing Authorisations for Medicinal Products*/ III/3630/92,**
Additional Notes	**This note for guidance concerns the application of Part 4, sections C and F of the Annex to Directive 75/318/EEC as amended with a view to the granting of a marketing authorisation for a medicinal product. It is intended to give support to applicants when designing, conducting, documenting, evaluating and reporting clinical trials of new medicinal products in the context of their overall clinical development. It is written primarily to harmonise the biostatistical methodology applied to clinical trials within member states, during both clinical development and subsequent review.**

CONTENTS

8. PRIMARY AND SECONDARY VARIABLES

9. THE CASE REPORT FORM

10. PRE SPECIFIED DATA ANALYSIS

11. THE CONDUCT AND MONITORING PHASE

12. INTEGRITY OF DATA AND COMPUTER SOFTWARE

13. EVALUATION AND REPORTING

14. OUTLINE OF SAFETY ISSUES

15. SUMMARISING THE CLINICAL DATABASE

BIOSTATISTICAL METHODOLOGY IN CLINICAL TRIALS

1. INTRODUCTION

This note for guidance should be read in the light of Directive 75/318/EEC as amended which sets up the relevant regulatory framework. It should also be read in conjunction with note for guidance on *Good Clinical Practice* (GCP) and which emphasises the importance of statistical principles and expertise in the design, conduct and analysis of clinical trials.

This Note is intended to give support to applicants when designing, conducting, documenting, evaluating and reporting clinical trials of new medicinal products in the context of their overall clinical development. It will also assist scientific experts charged with preparing Expert Reports or assessing evidence of efficacy and safety, principally from Phase III clinical trials. It is written primarily to harmonise the biostatistical methodology applied to clinical trials within member states, during both clinical development and subsequent review. It is largely concerned with planning and methodology, and does not attempt to cover systematically the structure and content of clinical reports which is covered in the note for guidance on *Structure and Content of Clinical Reports*.

The efficacy and clinical safety of medicinal products must be demonstrated by clinical trials which follow the guidance given in GCP. Within this document biostatistical issues are clearly identified as an integral part of the design and conduct of clinical trials. Biostatistical principles underlie the multi-disciplinary approach necessary to ensure that clinical trials provide adequate evidence to support the decisions which rest upon them and, therefore, an appropriately qualified and experienced statistician should be responsible for each clinical trial at all stages from design through to reporting.

In the early phases of product development many clinical trials are of an exploratory nature. Some of the content of this note is strictly relevant only to confirmatory trials such as the pivotal trials carried out in the later phase of development (late Phase II / Phase III) in order to give definitive proof of efficacy, and these trials are certainly the main target of the statistical material covered here. Nevertheless, biostatistical considerations are relevant to all clinical trials; hence the substance of this note should be applied as far as possible to all phases of product development and to all types of study. In section 14 of this note for guidance a brief outline of some of the statistical issues relevant to safety evaluation is given. However, a separate guideline will be necessary to cover this topic fully.

This note for guidance is intended:

- to outline the relevant principles of design and analysis,
- to describe generally appropriate approaches to these tasks,
- to describe some approaches which should not be adopted,
- to set conventions where choice would otherwise be entirely arbitrary. It is not intended to give detailed (textbook or “cookbook”) specification of methodology, thus leaving applicants and other users responsible for making reasoned choices between

alternative statistical approaches and procedures to solve specific problems. This may sometimes require innovation.

All important details of the design, conduct and proposed analysis of each clinical trial contributing to a marketing authorisation application should be clearly specified in a protocol written before the trial begins. Any subsequent changes to the protocol, and the reasons behind them, should be carefully documented. The clinical trial protocol and all subsequent amendments should be signed and dated by the responsible personnel (including the trial statistician). The trial statistician should ensure that the protocol covers all relevant statistical issues clearly and accurately, using technical terminology as appropriate.

Although this note for guidance is written largely from the classical (frequentist) viewpoint, the use of Bayesian or other well-argued approaches is quite acceptable.

2. OVERALL DEVELOPMENT CONSIDERATIONS

2.1 Clinical development plan

The broad aim of the process of clinical development of a new active substance is to find out whether there is a dose range and schedule at which the substance can be shown to be simultaneously effective and safe, to the extent that the risk/benefit relationship is acceptable. The particular patients who benefit from the medicinal product, and the specific indications for its use, also need to be defined.

Satisfying these broad aims usually requires an ordered programme of clinical trials, each with its own specific objectives. This should be specified in a clinical plan, or a series of plans, with appropriate decision points and flexibility to allow modification as knowledge accumulates. An application for marketing authorisation should clearly describe the main content of such plans, and the contribution made by each trial. Interpretation and assessment of the evidence from the total programme of trials involves synthesis of the evidence from the individual trials (see Section 15). This is facilitated by ensuring that common standards are adopted for a number of features of the trials such as dictionaries of medical terms, definition and timing of the main measurements, handling of protocol deviations and so on. A formal statistical overview or meta-analysis may be informative when medical questions are addressed in more than one trial. Where possible this should be envisaged in the plan so that the relevant trials are clearly identified and any required common features of their designs are specified in advance.

Other major statistical issues (if any) which are expected to affect a number of trials in a common plan should be addressed in that plan.

2.2 Confirmatory studies

As a rule, confirmatory trials are necessary for the definitive proof of efficacy. In such trials the key hypothesis of interest is always pre-defined, and corresponds to the hypothesis which is subsequently tested when the trial is complete. However, in a confirmatory trial it is equally important to estimate with due precision the size of the effects attributable to the treatment of interest and to relate these effects to their clinical significance.

Confirmatory trials are the cornerstones of decision making and hence adherence to their planned design and procedures is particularly important; unavoidable changes must be explained and documented, and their effect examined. A detailed and justified account of

the design of each such trial, and of all other statistical aspects such as the planned analysis, should be set out in the protocol. Each trial should address only a limited number of questions.

Firm evidence of efficacy requires that the results of the pivotal confirmatory trials demonstrate that the medical product under test has unequivocal clinical benefits. The number of pivotal trials should therefore be sufficient to answer each key clinical question relevant to the efficacy claim clearly and definitively. In addition it is important that the basis for generalisation to the intended patient population is understood and explained; this may also influence the number of pivotal trials required. Replication of important studies is of great value during the interpretation of results, especially when unforeseen problems have arisen during conduct or analysis or when earlier studies provide an insecure basis for the main hypothesis to be tested.

2.3 Exploratory studies

The rationale and design of confirmatory studies nearly always rests on earlier clinical work carried out in a series of exploratory studies. Like all clinical trials, these exploratory studies should have clear and precise objectives. However, in contrast to confirmatory studies, their objectives may not always lead to simple tests of pre-defined hypotheses. In addition, exploratory studies may sometimes require a more flexible approach to design so that changes can be made in response to accumulating results. Their analysis may entail data exploration; tests of hypothesis may be carried out, but the choice of hypothesis may be data dependent. Such studies cannot be the basis of the formal proof of efficacy, although they may contribute to the total body of relevant evidence.

Any individual study may have both confirmatory and exploratory aspects. For example, in most confirmatory studies the data are also subjected to exploratory analyses which serve as a basis for explaining or supporting their findings and for suggesting further hypotheses for later research. The protocol should make a clear distinction between the aspects of a study which will be used for confirmatory proof and the aspects which will provide data for exploratory analysis.

3. CLINICAL TRIAL DESIGNS

3.1 The parallel group design

The most common design of clinical trial in Phase III is the parallel group design in which patients are randomised to one of two or more arms, each arm being allocated a different treatment. These treatments will include the new medication under investigation, at one or more doses, and one or more control treatments, such as placebo and/or an active comparator. In general terms, the analysis of such trials and the interpretation of their results is straight-forward and is described in standard statistical textbooks. However, even in this relatively simple situation biostatistical advice is still necessary (in accordance with GCP), because there may be additional features of the design which it is important or essential to take into account, but which complicate the analysis and interpretation (e.g. covariates, repeated measurements over time, interactions between design factors).

3.2 The cross-over design

In the cross-over design of clinical trial, each patient is randomised to a sequence of two or more treatments, and hence acts as his/her own control for treatment comparisons. This simple manoeuvre is attractive primarily because it reduces the number of patients required to achieve a specific power, sometimes to a marked extent. In the simplest 2x2 cross-over design, each patient receives each of two treatments in randomised order in two successive treatment periods, often separated by a washout period. The most common extension of this entails comparing n(>2) treatments in n periods, each patient receiving all n treatments. Numerous variations exist such as designs in which each patient receives a subset of n(>2) treatments, or ones in which treatments are repeated within a patient.

Cross-over designs have a number of problems which can invalidate their results. The chief difficulty concerns carryover, that is the residual influence of treatments in subsequent treatment periods. In an additive model the effect of unequal carryover will be to bias direct treatment comparisons. In the 2x2 design the relevant contrast is totally aliased with the interaction between treatment and period and, what is more, the test for either of these lacks power because it is a "between patient" contrast. This problem is less acute in higher order designs, but cannot be entirely dismissed.

When the cross-over design is used it is therefore important to avoid carryover. This is best done by selective and careful use of the design on the basis of adequate knowledge of both the disease area and the new medication. The disease under study should be chronic and stable. The relevant effects of the medication should develop rapidly and be completely reversible. The washout periods should be sufficiently long for residual effects to have disappeared. 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.

A common, and generally satisfactory, use of the 2x2 cross-over design is to demonstrate the bioequivalence of two formulations of the same medication in healthy volunteers. Here, the CPMP note for guidance on *Investigation of Bioavailability and Bioequivalence* gives detailed advice on how to plan, conduct and evaluate such trials. In this particular application in healthy volunteers, carryover effects are rather unlikely to occur if the wash-out time between the two periods is sufficiently long. However it is still important to check this assumption during analysis on the basis of the data obtained, for example by demonstrating that no substance is detectable at the start of each period.

There are additional problems which need careful attention in cross-over trials. The most notable of these are the complications of analysis and interpretation arising from the loss of patients, and the difficulties of assigning those adverse events which occur in later treatment periods to the appropriate treatment. These, and other issues, are described in the guidelines on *Dose Response Information to Support Product Authorisation* prepared within the International Conference on Harmonisation process. The cross-over design should generally be restricted to situations where losses of patients from the trial are expected to be small.

3.3 Factorial designs

In a factorial design two or more treatments are evaluated simultaneously in the same patient population through the use of varying combinations of the treatments. The simplest example is the 2x2 factorial design in which patients are randomly allocated to one of the four possible combinations of two treatments, A and B say. These are: A alone; B alone; both

A and B; neither A nor B. The usual intention of using this 2x2 design is to make efficient use of clinical trial patients by evaluating the efficacy of the two treatments with the same number of patients as would be required to evaluate the efficacy of either one alone. The success of this approach depends upon the absence of any relevant interaction between treatments A and B so that the effects of A and B on the primary efficacy variables follow an additive model, and hence the effect of A is virtually identical whether or not it is additional to the effect of B. As for the cross-over trial, evidence that this condition is likely to be met should be established in advance of the trial by means of prior information and data. In principle this design may also be used for the specific purpose of examining the interaction of A and B to look for non-additivity, but then the advantages of efficiency no longer apply because much larger trials are necessary to detect clinically relevant interaction. The 2x2 design has proved to be particularly valuable for very large mortality studies.

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

In the clinical development of medicinal products there are a number of other situations where these and other factorial designs are occasionally appropriate. The detailed design and analysis of these studies is a matter for professional statistical attention.

4. OTHER DESIGN ISSUES

4.1 Multicentre trials

A multicentre trial is an accepted way of establishing the efficacy of a new medication more efficiently. Sometimes it provides the only means of accruing sufficient patients within a reasonable time-frame. A further advantage commonly claimed for the multicentre trial is that it provides a better basis for subsequent generalisation of findings, because the patients are recruited from a wider population, and the medication is administered in a broader range of clinical settings, thus presenting an experimental situation which is more typical of future use. In addition the involvement of a larger number of investigators gives the potential for more clinical judgement concerning the value of the medication.

The additional efficiency of the multicentre trial depends upon the assumption that the differences between the compared treatments are constant from centre to centre; during analysis, the robustness of the conclusions to this assumption generally needs to be explored. It is important to design and conduct the trial with this background in mind. Procedures should be standardised as completely as possible: variation of evaluation criteria and schemes can be reduced by investigator meetings, by the training of personnel in advance of the study and by careful monitoring during the study. Good design should ensure a balanced distribution of patients to treatments within each centre and good management should ensure that centres do not diverge excessively in the number of patients entered. Failure to take these precautions, combined with doubts about the homogeneity of the results, may, in severe cases, reduce the value of a multicentre trial to such a degree that it cannot be regarded as giving convincing evidence for the applicant's claims.

The analysis plan (see Section 10) for a multicentre trial should clearly define the centres (e.g. by investigator, location or region) in the context of the particular trial whenever there is room for doubt. The plan should also specify prospectively any intention to combine small centres in the analysis, and the rules for so doing. The statistical model to be adopted for the comparison of treatments should be described in the protocol. The model should generally allow for centre differences – exceptions relate, for example, to large mortality studies with very few patients per centre – and similarly the possibility of qualitative or quantitative centre by treatment interaction should generally be explored, as this may affect the generalisability of the conclusions.

4.2 Trials to show superiority

Scientifically, efficacy is most convincingly established by demonstrating superiority to placebo in a placebo-controlled trial. Similarly, efficacy would normally be supported by showing superiority to an active control treatment. This type of trial is later referred to as a "comparative" trial (see Section 10.1.3).

4.3 Trials to show equivalence

When placebo-controlled trials are considered unethical, however, the demonstration of comparable efficacy to that of a standard therapy (rather than superiority) may also be acceptable provided a suitable standard therapy exists. [An example of a suitable standard therapy would be a widely used therapy whose efficacy in the relevant indication has been clearly established and quantified in well designed and well documented placebo-controlled trial(s).] This type of trial is often called an "equivalence" trial. In some indications equivalence trials are also undertaken for other regulatory reasons (see, for example, note for guidance on *Clinical Requirements for Locally Applied, Locally Acting Products, containing Known Constituents*). Equivalence trials may also incorporate a placebo thus pursuing two goals in one trial i.e. establishing superiority to placebo and evaluating the degree of similarity to a reference treatment.

There are well known dangers associated with the use of the active control equivalence trial. These relate to the implicit lack of any yardstick of internal validity (in contrast to placebo-controlled trials), thus making external validation necessary. The equivalence trial is not conservative in nature, so that most flaws in the design or conduct of the trial will tend to bias the results towards a conclusion of equivalence. For these reasons a number of design features need special attention.

First and foremost, it is vital that the protocol of a trial designed to demonstrate equivalence contains a clear statement that this is its explicit intention. Secondly, the primary measurement should be one which has been shown to be valid and reliable, and to reflect the major aspect of the disease which the new treatment is intended to improve. Thirdly, the dose(s) selected for study, the patient population and other aspects of the design should be such as to ensure adequate sensitivity of the trial to detect relevant differences if present. Reassurance on this (and other issues) can often be gained by modelling the study on earlier placebo-controlled studies of the active comparator in which the comparator clearly demonstrated clinically valuable efficacy. Finally, classical hypothesis testing (with the null hypothesis that there is no difference between the treatments) is inappropriate, so that sample size calculations need an alternative basis (see Section 5).

There are also special issues in analysis. Patients who withdraw or drop out from treatment can bias the results towards showing equivalence, and hence the "intention to treat" strategy is insecure (see Section 10.1). Analysis is generally based around the use of confidence intervals (see Section 10.4) and is directed at establishing the largest possible treatment difference which might reasonably be consistent with the observed results.

4.4 Dose Response designs

How response is related to the dose of a new substance is a question to which answers may be obtained in all phases of development, and by a variety of approaches. Dose response studies may serve a number of objectives, amongst which the following are of particular importance: the confirmation of efficacy; the investigation of the shape and location of the dose response curve; the estimation of an appropriate starting dose; the identification of optimal strategies for individual dose adjustments; the determination of a maximal dose beyond which additional benefit would be unlikely to occur. These objectives need to be addressed using the data collected at a number of doses under investigation, perhaps including a placebo (zero dose). For this purpose the application of estimation procedures, including the construction of confidence intervals, and of graphical methods is as important as the use of statistical tests. The hypothesis tests which are used may need to be tailored to the special ordering of study arms or to the particular questions of interest in these studies: does response increase monotonically with dose?; does response plateau or decline at higher doses? The details of the planned statistical procedures should be laid down in the protocol.

4.5 Group sequential designs

The group sequential design is the most commonly applied form of sequential design. The purpose of a group sequential design is to allow the trial to be monitored at specific time intervals so that a treatment arm, or the entire trial, may be stopped early if there is clear evidence of efficacy or of unacceptable adverse effects. Trials may also be stopped early because the chance of ever showing a clinically valuable benefit is assessed as too small. However, the logistics of a group sequential trial are complicated and require careful planning: an independent monitoring committee must generally be employed: the statistical procedures must be fully specified in advance (see Section 11.4) by the monitoring committee before they conduct their first interim look at the data. This design is best reserved for large long-term trials of mortality or major non-fatal end-points.

5. THE SAMPLE SIZE

The number of patients in a clinical trial should always be large enough to provide a reliable answer to the questions addressed, but should also be the minimum necessary to achieve this aim. This number is usually determined by the primary efficacy 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 will usually need larger numbers of patients. (See, for example, note for guidance on *The Extent of Population Exposure to assess Clinical Safety for Medicines intended for Long-term Treatment of Non-life-threatening Conditions*.)

The usual method for determining the appropriate sample size requires specification of a primary variable, the test statistic, the null hypothesis, the alternative ('working') hypothesis

(embodying consideration of the treatment difference to be detected or rejected), the probability of erroneously rejecting the null hypothesis (the type I error), and the probability of erroneously accepting the null hypothesis (the type II error), as well as the approach to dealing with dropouts and other protocol deviations. 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, difference to be detected). The basis of these estimates should also be given; in confirmatory studies, they 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 which has clinical relevance in the management of patients, or on a judgement concerning the anticipated effect of the new treatment, where this is larger. Conventionally the type I error is set at 5% or less; the precise choice of value is influenced by the prior plausibility of the hypothesis under test and the desired impact of the results. The type II error is conventionally set at 20% or less; it is in the investigator's interest to keep this figure as low as possible especially in the case of studies which are difficult or impossible to repeat.

When the variance of the primary variable, or the control response rate, is particularly uncertain, it may be worthwhile to carry out a suitably blinded interim estimation of the quantity in question in order to update the sample size appropriately. Any such plans should be carefully described in the protocol – the steps taken to preserve blindness should be explained, together with the consequences (if any) for the type I error and the width of confidence intervals. (See also Section 11.3).

Sample size calculations should make appropriate allowance for dropouts and any other anticipated protocol deviations which may affect power. Not only should the sample size be increased to offset the loss of patients from the study or from their intended treatments, but it also should allow for the fact that the size of effect observed may be reduced by the need to include any available data from treatment failures or other early terminators in an intention-to-treat analysis (see Section 10.1.1).

The sample size of an equivalence trial (see Section 4.3) should normally be based on the objective of obtaining a confidence interval for the treatment difference which is sufficiently narrow to exclude a non-negligible difference, given that the treatments are truly equivalent. The choice of “non-negligible” difference requires justification, and may be smaller than the minimal clinically relevant difference referred to above in the context of trials designed to establish that a difference exists. In bioequivalence trials the use of a 90% confidence interval is conventional; therapeutic clinical trials may require the use of confidence intervals with greater coverage probability (e.g. 95%) in order to achieve sufficient impact.

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

6. THE POPULATION

In the earlier phases of product development the choice of patients 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 patient population for which the substance may eventually be indicated. However by the time the pivotal Phase III trials are undertaken, the patient populations should more closely mirror the intended users. Hence in these trials it is generally helpful to relax the inclusion and exclusion as much as possible within the target indication, whilst maintaining sufficient homogeneity to permit a successful trial to be carried out. No individual clinical trial can be expected to be totally representative of future users, because of the possible influences of the geographical location, the time when it is conducted, the medical practices of the particular investigator(s) and clinics, and so on. However the influence of such factors should be reduced wherever possible, and subsequently discussed during the interpretation of the trial results.

It is unusual for all eligible patients considered for entry to a trial to be included. In pivotal trials, it is recommended that a log should be kept of all screened patients so as to allow some assessment of the numbers and characteristics of the excluded patients, and the reasons for their exclusion.

7. DESIGN TECHNIQUES TO AVOID BIAS

The two most important design techniques for avoiding bias in clinical trials are blinding and randomisation, and these should be a normal feature of most controlled clinical trials intended to be included in the regulatory application for a new product. Most such trials follow a double-blind approach in which treatments are pre-packed in accordance with a suitable randomisation schedule, and supplied to the trial centre(s) labelled only with the patient number and the treatment period so that no-one involved in the conduct of the trial is aware of the specific treatment allocated to any particular patient, not even as a code letter. This approach will be assumed in Sections 7.1 and most of Section 7.2, exceptions being considered at the end.

7.1 Blinding

Blinding is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence which the knowledge of treatment may have on the recruitment and allocation of patients, their subsequent care, the attitudes of patients to the treatments, the assessment of end points, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias are passed.

In a double-blind trial neither the patient nor any of the staff involved in the treatment or clinical evaluation of the patient is aware of the treatment received. In a single blind trial the investigator and/or his staff are aware of the treatment but not the patient. In an open-label trial the identity of treatment is known to all. The double-blind trial is the optimal approach. This requires that the treatments to be applied during the trial cannot be distinguished in any way (appearance, taste, etc.) either before or during administration.

Difficulties in achieving the double-blind ideal can arise: the treatments may be of a completely different nature – for example, surgery and medicinal product therapy; two substances may have different galenic forms and, although they could be made indistinguishable by the use of capsules, changing the galenic form might also change the pharmacokinetic and/or pharmacodynamic properties and hence require that bioequivalence

of the forms be established; the daily pattern of administration of two treatments may differ. One way of achieving double-blind conditions under these circumstances is to use a double dummy technique. This technique may sometimes force an administration scheme which is sufficiently unusual to influence adversely the motivation and compliance of the patients. Ethical difficulties may also interfere with its use when, for example, it entails dummy operative procedures. Nevertheless extensive efforts must be made to overcome these difficulties.

If a double-blind trial is not feasible, then the single blind option should be considered. In some cases only an open-label trial is practically or ethically possible. Under either of these circumstances it is preferable, where possible, for clinical assessments to be made by medical staff who are not involved in treating the patients and who remain blind to treatment. In single blind or open-label trials every effort should be made to minimise the various known sources of bias. The reasons for the degree of blinding adopted should be explained in the protocol, together with steps taken to minimise bias by other means.

Any intentional or unintentional breaking of the blind must be reported at the end of the trial, irrespective of the reason for its occurrence.

7.2 Randomisation

Randomisation introduces a deliberate element of chance into the assignment of treatments to patients 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. In combination with blinding, randomisation helps to avoid possible bias in the selection and allocation of patients arising from the predictability of treatment assignments.

The randomisation schedule of a clinical trial documents the random allocation of treatments to patients. In the simplest situation it is a sequential list of treatments (or treatment sequences in a cross-over study) or corresponding codes by patient number. The logistics of some studies, such as those with a screening phase, may make matters more complicated, but the unique pre-planned assignment of treatment, or treatment sequence, to patient should be clear. Different designs of study will require different procedures for generating randomisation schedules. This procedure should be capable of being reproduced (if the need arises) through the use of the same random number table, or the same computer routine and seed for its random number generator.

Although unrestricted randomisation is an acceptable approach, some advantages can generally be gained by randomising patients in blocks. This helps to increase the comparability of the treatment groups particularly when patient characteristics may change over time, as a result, for example, of changes in recruitment policy. It also provides a better guarantee that the treatment groups will be of nearly equal size. In cross-over trials it provides the means of obtaining balanced designs with their greater efficiency and easier interpretation. Care must be taken to choose block lengths which are sufficiently short to limit possible imbalance, but which are long enough to avoid predictability towards the end of the sequence in a block. Investigators should generally be blind to the block length; the use of two or more block lengths, randomly selected for each block, can achieve the same purpose. (Strictly, in a double-blind trial predictability does not matter, but the pharmacological effects of medicinal products often provide the opportunity for intelligent guesswork.)

In multicentre studies the randomisation procedures should always be organised centrally. It is advisable to have a separate random scheme for each centre i.e. to stratify by centre. More generally, stratification by important prognostic factors measured at baseline (e.g. severity of disease, age, sex, etc.) may sometimes be valuable in order to promote balanced allocation within strata; this has greater potential benefit in small trials. The use of more than two or three stratification factors is rarely necessary, is less successful at achieving balance and is logistically troublesome. Where it *is* necessary the use of a dynamic allocation procedure (see below) may help to achieve balance across all factors simultaneously provided the rest of the trial procedures can be adjusted to accommodate an approach of this type.

The next patient to be randomised into a study should always receive the treatment corresponding to the next free number in the appropriate randomisation schedule (in the respective stratum, if randomisation is stratified). The appropriate number and associated treatment pack for the next patient should only be allocated when entry of that patient to the randomised part of the trial has been confirmed. These tasks will normally be carried out by staff at the investigator's centre, who will then dispense the relevant blinded trial supplies. Details of the randomisation which facilitate predictability (e.g. block length) should not be contained in the main study protocol, but should be set out in an addendum which can be withheld from the study sites. The randomisation schedule itself should be filed securely by the applicant in a manner which ensures that blindness is properly maintained throughout the trial. Access to the randomisation schedule during the trial must take into account the possibility that, in an emergency, the blind may have to be broken for any patient, either partially or completely. The procedure to be followed, the necessary documentation, and the subsequent treatment and assessment of the patient should all be described in the protocol.

Alternative randomisation procedures generally require some relaxation of the degree of blindness. An important case is when the treatment *code* of each patient is known although the assignment of actual treatment to treatment codes is not; this is still technically double-blind but the groups of patients on a common treatment are identifiable. This relaxation is necessary to permit the normal methods of dynamic allocation to be used. In these, the allocation of treatment to a patient may be influenced by the current balance of allocated treatments and, in a stratified trial, by the stratum to which the patient belongs and the balance within each prognostic factor. Knowledge of the treatment code may be restricted to a central trial office from where the dynamic allocation is controlled, generally through telephone contact. This in turn permits additional checks of eligibility criteria and establishes entry into the trial – these features can be valuable in certain types of multicentre trial. The usual system of pre-packing and labelling product supplies for double-blind trials can then be followed, but the order of their use is no longer sequential. (It is even possible through the use of appropriate computer algorithms to keep the personnel at the central trial office blind to the treatment code.)

Single blind and open-label trials provide additional flexibility, but in these it is particularly important to take care that the investigator's knowledge of the next treatment does not influence his decision to enter the patient as randomised: his decision should always precede knowledge of the randomised treatment. Any decision to reduce the level of blinding in order to use a particular method of randomisation requires careful justification.

8. PRIMARY AND SECONDARY VARIABLES

The primary variable ('target' variable, primary endpoint) should be the variable capable of providing the most relevant and convincing evidence directly related to the primary objective of the trial. There should generally be only one primary variable and this will usually be an efficacy criterion, because the primary objective of most confirmatory trials concerns efficacy. The selection of the primary variable must reflect the accepted norms and standards in the relevant field of research. The use of a reliable and validated variable with which experience has been gained either in earlier studies or in published literature is recommended. There should be sufficient evidence that the primary variable will be sensitive to clinically relevant treatment effects in the patient population described by the inclusion and exclusion criteria.

The primary variable should be specified in the protocol. A post-hoc definition will rarely be acceptable, since the biases this introduces are difficult to assess. When the clinical effect defined by the primary objective is to be measured in more than one way, then the protocol should identify one of the measurements as the primary variable, whenever possible. If this is not possible, another useful strategy is to combine the multiple measurements into a single variable, perhaps with suitable weighting; indeed, the primary variable sometimes arises naturally as a combination of multiple clinical measurements (e.g. the rating scales used in arthritis, psychiatric disorders and elsewhere). In either case, the method of combining the multiple measurements should be specified in the protocol, and an interpretation of the new scale should be provided in terms of the size of a clinically relevant benefit. The primary variable is also the most appropriate one to use in the protocol for estimating the sample size (see section 5).

It may sometimes be desirable to use more than one primary variable to cover the range of effects of the therapies, simultaneous evidence of effect on each of them (or a pre specified subset of them) being sought in order to demonstrate efficacy of the product. The planned manner of interpretation of this type of evidence must be carefully spelled out and the effect on the type I error explained, because of the potential for multiple comparison problems (see Section 10.5).

When direct measurement of the clinical benefit to the patient is not feasible, indirect criteria (so-called surrogate variables) may have to be considered. The use of a surrogate variable as a primary variable requires careful justification of the claim that, under the circumstances of the trial, the anticipated or observed treatment effects on the surrogate variable predict clinical benefits of value to the patient.

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.

On some occasions categorisation or dichotomisation of continuous or ordinal variables may be necessary in which case the criteria should be specified in advance in the protocol. 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 a minimum of 'good' on an ordinal rating scale.

9. THE CASE REPORT FORM

The form and the content of the case report forms (CRFs) must be in full accordance with the protocol and must be established in advance of the conduct of the study. Case report forms should only include relevant items which will be evaluated in the final report. Data fields should be designed so that "missing values" can be distinguished from "value zero" or "characteristic absent". Verbal scales should be exhaustive and overlapping categories should be avoided so that unambiguous answers are obtained. It is generally more appropriate to document the absolute value on a rating scale rather than changes from a baseline. If changes *are* documented, then the time point to which the change is related should be made clear and unambiguous.

The CRF should be filled out during the course of the trial and the dates to which the data apply and on which the form was completed should be noted. Subsequent additions or amendments to the form should be avoided, except as a means of correcting errors identified before analysis during a planned process of monitoring or quality control, in which case the changes should be properly documented, signed and dated.

Similar considerations apply to data collected via a medium other than paper, such as the electronic medium employed in remote data entry. Further guidance relating to CRFs can be found in Chapter 3 of GCP.

10. PRE SPECIFIED DATA ANALYSIS

When designing a clinical trial, it is good practice to consider the eventual analysis of the data and to document the resulting plans – the "analysis plan". In the case of an exploratory trial the plan may describe only general principles and directions, but in a confirmatory trial the plan should be precise in a number of respects if the credibility of the findings is to be maximised. All features of the proposed confirmatory analysis of the primary variable(s), including the way in which anticipated analysis problems will be handled, should be covered in the plan normally within the protocol. The principal features of the analysis of all other data should also be outlined.

It is also valuable to review the analysis plan when the study data are complete and available for review but whilst the statistician (amongst others) is still blinded to treatment, and hence before the main analysis. This will be referred to as the "blind review".

10.1 Study populations

The analysis plan should first define the population of patients whose data it is planned to include in the main analyses. In addition in pivotal trials any available data on the wider population screened for entry, or enrolled into a screening phase of the study prior to randomisation, should be summarised within the bounds (if any) imposed by the need for informed consent. The intention of this planned summary is to provide information about the numbers and characteristics of excluded patients, both eligible and ineligible, together with their reasons for exclusion, in order to guide assessment of the potential practical impact of the study results. As a minimum, documentation is required for all patients for whom study procedures were initiated and who have given their informed consent. The content of this patient documentation depends on detailed features of the particular trial, but at least demographic and baseline data on disease status should be collected whenever possible.

If all patients randomised into a clinical trial satisfied all entry criteria, followed all trial procedures perfectly with no losses to follow up, and provided complete data records, then the population of patients to be included in the analysis would be self-evident. The design and conduct of a trial should aim to approach this ideal as closely as possible but, in practice, it is doubtful if it can ever be fully achieved, and hence the analysis plan should address any anticipated problems prospectively in terms of how these affect the patients and data to be analysed. Decisions concerning the analysis population should be guided by the principles underlying the "intention-to-treat" and the "per protocol" strategies.

10.1.1 The intention-to-treat population

The principle of intention-to-treat (ITT) implies that all randomised patients should be included in the analysis. In most clinical trials it provides a conservative approach, and also gives estimates of treatment effects which are more likely to mirror those observed in subsequent practice. In practice the protocol should try to avoid or ameliorate each type of irregularity which is anticipated during the trial and which might impair a satisfactory ITT analysis. Wherever possible, it should also define prospectively how any unavoidable problems will be addressed. Any objective entry criteria, measured before randomisation, which will be used to exclude patients from analysis should be pre-specified and justified, for example, those relating to the presence or absence of the disease under investigation. Any other anticipated reasons for excluding patients from ITT analysis should be justified.

The problems associated with ITT lie in the handling of some irregularities (incorrectly entered patients, errors in treatment assignment, missing data, loss to follow up, and some other protocol violations) and the subtleties which this can involve. For example, patients who fail to satisfy an objective entry criterion measured prior to randomisation, but who erroneously enter the trial, may be excluded from analysis without introducing bias into the treatment comparison. However not all incorrectly entered patients fall into this category. As another example, patients withdrawn from treatment may introduce serious bias and, if they have provided no data after withdrawal, ITT offers no obvious solution. The necessary inclusion of such patients in the analysis may require some re-definition of the primary variable, or some assumptions about the patients' fate.

Measurements of primary variables made at the time of withdrawal (or at the time of the loss of the patient for any other reason), or subsequently collected in accordance with the protocol, are valuable in the context of ITT analysis. Their use in analysis should be described and justified in the analysis plan and their collection described elsewhere in the protocol. However, the use of techniques such as "last observation carried forward" (LOCF) should recognise the fact that it can lead to biased estimates of treatment effects when the likelihood of the loss of a patient is related to treatment or response. Any other methods to be employed to ensure the availability of measurements of the primary variable for every patient in the ITT analysis should be described.

Because of the complexity of the problems connected with ITT it may sometimes be preferable to defer detailed consideration of the manner of dealing with irregularities until the blind review of the data at the end of the study and, if so, this should be stated in the analysis plan.

10.1.2 The per protocol population

The 'per protocol' population, sometimes described as the 'valid cases', the 'efficacy' population or the 'evaluable patient' population, is a subset of the intention-to-treat population and is characterised by the following criteria:

- the completion of a certain pre-specified minimal exposure to the treatment regimen,
- the availability of measurements of the primary variable(s) at relevant and pre-specified time(s),
- the absence of any major protocol violation including the violation of entry criteria (where the nature of and reasons for these protocol violations must be defined and documented before breaking the blind).

This population generally maximises the opportunity for a new treatment to show additional efficacy in analysis, and most closely reflects the scientific model underlying the protocol. However it is not conservative, and is more vulnerable to biases than ITT.

10.1.3 The roles of the intention-to-treat and the per protocol analyses

In general, it is advantageous to demonstrate a lack of sensitivity of the principal trial results to alternative choices of patient population for analysis. In pivotal trials it is usually appropriate to plan to conduct both ITT and per protocol analyses, so that any differences between them can be the subject of explicit discussion and interpretation subsequently. In some cases, it may be desirable to plan further exploration of the sensitivity of conclusions to the choice of population analysed. When the ITT and the per protocol analyses come to essentially the same conclusions, confidence in the study results is increased, bearing in mind, however, that the need to exclude a *substantial* proportion of the ITT population from the per protocol analysis throws some doubt on the overall validity of the study.

ITT and per-protocol analyses play different roles in comparative trials (which seek to show one active substance to be superior), and in equivalence trials (which seek to show medicinal products to be similar – see Section 4.3). In comparative studies the ITT analysis usually tends to avoid the optimistic estimate of efficacy which may result from a per protocol analysis, since the noncompliers included in an ITT analysis will generally diminish the overall treatment effect. However in an equivalence trial ITT no longer provides a conservative strategy and its role should be considered very carefully. Hence in comparative trials, particularly those claiming to be pivotal with regard to efficacy, the ITT analysis will normally be of higher priority, whereas in equivalence trials this is not the case.

10.2 Missing values and outliers.

Missing values represent a potential source of bias in a clinical trial. Hence, every effort should be undertaken to fulfil all the requirements of the protocol concerning the collection of data and their subsequent management. However, in reality there will almost always be some missing data. A study may be regarded as valid, none-the-less, provided the methods of dealing with missing values are sensible, and particularly if those methods are pre-defined in the analysis plan of the protocol. Pre-definition of methods may be facilitated by updating this aspect of the analysis plan during the blind review. Unfortunately, no universally applicable methods of handling missing values can be recommended. An investigation

should be made concerning the sensitivity of the results of analysis to the method of handling missing values, especially if the number of missing values is substantial.

A similar approach should be adopted to exploring the influence of outliers, the statistical definition of which is, to some extent, arbitrary. Clear identification of a particular value as an outlier is most convincing when justified medically as well as statistically, and the medical context will then often define the appropriate action. Any outlier procedure set out in the protocol should be such as not to favour any treatment group *a priori*. Once again, this aspect of the analysis plan can be usefully updated during blind review. If no procedure for dealing with outliers was foreseen in the study protocol, one analysis with the actual values and at least one other analysis eliminating or reducing the outlier effect should be performed and differences between their results discussed.

10.3 Data transformation/modification

The decision to transform key variables prior to analysis is best made during the design of the trial on the basis of similar data from earlier clinical trials. Transformations (e.g. square root, logarithm) should be specified in the protocol and a rationale provided, especially for the primary variable(s). The general principles guiding the use of transformations are to be found in standard texts; conventions for particular variables have been developed in a number of specific clinical areas. The decision on whether and how to transform a variable should be influenced by the preference for a scale which facilitates clinical interpretation.

Similar considerations apply to other data modifications sometimes used to create a key variable for analysis, such as the use of change from baseline, percentage change from baseline, the 'area under the curve' of repeated measures, or the ratio of two different variables. Subsequent clinical interpretation should be carefully considered, and the modification should be justified in the protocol. Closely related points are made in Section 8.

10.4 Estimation, confidence intervals and hypothesis testing.

The statistical analysis plan should specify the hypotheses which are to be tested and/or the treatment effects which are to be estimated in order to satisfy the objectives of the trial. The statistical methods to be used to accomplish these tasks should be described for the primary (and preferably the secondary) variables, and the underlying statistical model should be made clear. Estimates of treatment effects should be accompanied by confidence intervals, whenever possible, and the way in which these will be calculated should be identified. The plan should also describe any intentions to use baseline data to improve precision and to adjust estimates for potential baseline differences, for example by means of analysis of covariance. The reporting of precise numerical probabilities (e.g."P=0.034") should be envisaged in the plan, rather than exclusive reference to critical values (e.g. "P<0.05"). It is important to clarify whether one- or two-sided tests of statistical significance will be used, and in particular to justify prospectively the use of one-sided tests. If formal hypothesis tests are not considered appropriate, then the alternative process for arriving at statistical conclusions should be given.

The particular statistical model chosen should reflect the current state of the art. All effects to be fitted in the analysis (for example in analysis of variance 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 10.6). Due attention should be paid to the statistical distribution of both primary and secondary variables. The pre-planned use of non-parametric methods to avoid distribution difficulties is acceptable. When making this choice it is important to bear in mind the need to provide statistical estimates of the size of treatment effects, together with confidence intervals, as this may influence the choice of method in cases of doubt. Any essential changes to the pre-planned statistical modelling procedure should be made during blind review.

The analysis of the primary variable by which the primary objective of the trial will be judged must be clearly distinguished from supporting analyses of the primary or secondary variables. In these other analyses, the number of significance tests should be minimised and preference given to descriptive methods and estimation procedures whenever possible. Within the analysis plan there should also be an outline of the way in which data other than the primary and secondary variables will be summarised and reported. This should include a reference to any approaches adopted for the purpose of achieving consistency of analysis across a range of studies, for example for safety data.

10.5 Adjustment of type I error

When multiplicity is present, the usual frequentist approach to the analysis of clinical trial data may require an adjustment to the type I error. Multiplicity may arise from multiple primary variables (see Section 8), multiple comparisons of treatments, repeated evaluation over time and/or interim analyses (see Section 11.4). Methods to avoid or reduce multiplicity are sometimes preferable when available, such as the identification of the key primary variable (multiple variables), the choice of a critical treatment contrast (multiple comparisons), the use of a summary measure such as “area under the curve” (repeated measures), and the use of extreme type I errors (interim analysis). 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 should be set out in the analysis plan.

10.6 Subgroups, interactions and covariates

The primary variable(s) is often systematically related to other influences apart from treatment. For example, there may be relationships to covariates such as age and sex, or there may be differences between specific subgroups of patients such as those treated at the 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 analysis plan and hence should be set out in the protocol. Pre-study deliberations should identify those covariates and factors expected to have an important influence on the primary variable(s), and should consider how to account for these in the analysis in order to improve precision and to compensate for any lack of balance between treatment groups. When the potential value of an adjustment is in doubt, it is often advisable to nominate the unadjusted analysis as the one for primary attention, the adjusted analysis being supportive. Special attention should be paid to centre effects and to the role of baseline measurements of the primary variable. It is not advisable to adjust the main analyses for covariates measured after randomisation.

The treatment effect itself may also vary with subgroup or covariate – for example, the effect may decrease with age or may be larger in a particular diagnostic category of patient. In some cases such interactions are anticipated, and hence a subgroup analysis, or a statistical model including interactions, is part of the confirmatory analysis plan. In most cases,

however, subgroup or interaction analyses are exploratory and should be clearly identified as such; they should explore the uniformity of any treatment effects found overall. In general, such analyses should proceed through the addition of interaction terms to the statistical model in question, rather than through multiple separate analyses within each subgroup of patients, or within strata defined by the covariates. When exploratory, these analyses should be carried out informally and interpreted cautiously; any claims of treatment efficacy based solely on exploratory subgroup analyses are unlikely to be accepted.

11. THE CONDUCT AND MONITORING PHASE

Careful conduct of a trial according to the protocol has a major impact on the credibility of the study results. Careful monitoring can ensure that difficulties are noticed early and their recurrence minimised. Both help to permit a straightforward and unambiguous analysis. The period of conduct and monitoring starts with the selection of the study sites and ends with the collection and cleaning of the last patient's data. In some cases, employees of the sponsor may suffice to monitor the quality of the trial activities and data, while in other cases an independent and experienced body of specialists (a monitoring committee) will also be required. The latter is appropriate if formal interim analyses are planned, with some lifting of the blindness being necessary, and/or if regular and independent reviews of safety are required. The employees of the sponsor and the monitoring committee both require access to statistical expertise for advice and assistance. This may relate to the handling of unexpected protocol violations to minimise effects on the analysis (sponsor), or to the execution, interpretation or modification of interim analyses (monitoring committee).

11.1 Changes in inclusion and exclusion criteria

Inclusion and exclusion criteria should, if possible, remain constant, as specified in the protocol, throughout the period of patient recruitment. However, occasionally this may not prove possible; in long-term studies, for example, growing medical knowledge either from outside the trial or from interim analyses may suggest a change of entry criteria. Changes may also result from the discovery by monitoring staff that regular violations of the entry criteria are occurring, or that seriously low recruitment rates are due to over-restrictive criteria. The protocol amendment implementing such changes should cover any statistical consequences, such as sample size adjustments arising from different event rates, or modifications to the analysis plan.

11.2 Accrual rates

In studies with an appreciable time-scale for the accrual of patients, the rate of accrual should be monitored and, if it falls appreciably below the projected level, the reasons should be identified and remedial actions taken in order to protect the power of the trial and allay concerns about selective entry.

In a multicentre trial the accrual rate of patients at the various study sites should also be monitored to seek to ensure that the centres do not diverge excessively in the numbers of patients entered (see Section 4.1).

11.3 Checking the design assumptions

In large trials there will usually be an opportunity to check the assumptions which underlay the original design and sample size calculations. This may be particularly important if the trial specifications have been made on preliminary and/or uncertain information. An interim check conducted on the blinded data may reveal that overall response variances, event rates or survival experience are not as anticipated. A revised sample size may then be calculated using suitably modified assumptions, and should be justified and documented in a protocol amendment and in the final report. The steps taken to preserve blindness and the consequences, if any, for the type I error and the width of confidence intervals should be explained. The potential need for re-estimation of the sample size should be envisaged in the protocol whenever possible (see Section 5).

Substantial differences between the anticipated and observed average level of the primary variable(s) may pose additional design problems for a trial, because it may alter judgement concerning the treatment difference to be detected, or the size of the non-negligible difference to be excluded in the case of an equivalence trial. In addition, the validity of an equivalence trial is judged by its similarity to earlier trials of the positive control, and substantial differences in the level of the primary variable do not support the idea of similarity.

11.4 Interim analyses and early stopping

When interim analysis is planned, this is usually accomplished by the use of a group sequential design (see Section 4.5). The goal of such an interim analysis is to stop the trial early if the superiority of the treatment under test is already clearly established, if the demonstration of a relevant treatment difference has become unlikely, or if unacceptable adverse effects are apparent. The schedule of analyses, or at least the considerations which will govern its generation, should be stated in the protocol or appended to it before the time of the first interim analysis; the stopping rules and their properties should also be clearly stated there. This material should be written or approved by the monitoring committee, when the study has one. Deviations from the planned procedure always bear the potential of invalidating the study results. If it becomes necessary to make changes to the trial, any consequent amendments to the statistical procedures should be specified in the protocol at the earliest opportunity. The procedures selected should always ensure that the overall type I error is, at least, maintained.

The execution of an interim analysis must be a completely confidential process because unblinded data and results are potentially involved. All staff 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 applies equally to the investigators and their staff and to staff employed by the applicant. Investigators should only be informed about the decision to continue or to discontinue the trial.

Any interim analysis planned for administrative purposes only, with no adjustment of the type I error, should also be described in the protocol and subsequently reported. The purposes of an analysis of this type should be clearly stated, and should specifically exclude any possibility of early stopping or other major changes in design. The blind should not be broken.

Any unplanned or incorrect interim analysis (with or without the consequence of stopping the trial early) can flaw the results of a trial, and weaken confidence in the conclusions

drawn. Such analyses should therefore be avoided if at all possible. If an unplanned interim analysis cannot be avoided, the study report must later explain carefully why it was necessary, the degree to which blindness had to be lifted, and what resulting actions were possible in principle.

12. 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 must therefore be based on thorough and effective standard operating procedures (SOPs). The computer software used for data management and statistical analysis must be completely reliable, and documentation of appropriate validation procedures should be available. The computer systems and packages to be used should preferably be identified in the protocol.

13. EVALUATION AND REPORTING

As stated in the Introduction, the structure and content of clinical reports is the subject of another note for guidance on *Structure and Content of Clinical Study Reports*. This note for guidance fully covers the reporting of statistical work, appropriately integrated with clinical and other material. The current section is therefore relatively brief.

During the planning phase of a trial the main features of the analysis should have been specified in an analysis plan and included in the protocol as described in Section 10. When the conduct of the trial is over and the data are assembled and available for preliminary inspection, it is valuable to carry out the blind review of the planned analysis also described in Section 10. This pre-analysis review, blinded to treatment, should cover decisions concerning the exclusion of patients or data from the analysis population(s); possible transformations may also be checked, and outliers defined; important covariates identified in other recent research may be added to the model; the use of parametric or non-parametric methods may be reconsidered. Decisions made at this time should be described in reports, and should be distinguished from those made after the statistician has had access to the treatment codes, as blind decisions will generally carry greater weight.

Many of the more detailed aspects of presentation and tabulation, particularly of the less critical data, may be finalised at or about the time of the blind review so that by the time of the actual analysis full plans exist for all its aspects including patient selection, data selection and modification, data summary and tabulation, estimation and hypothesis testing. Once data validation is complete, the analysis should proceed according to the pre-defined plans; the more these plans are adhered to, the greater the credibility of the results. Particular attention should be paid to any differences between the planned analysis and the actual analysis, and a careful explanation should be provided.

All patients who entered the trial should be accounted for in the report, whether or not they are part of an analysis population. All reasons for exclusion from analysis should be documented; for any patient included in the intention-to-treat population but not in the per-protocol population, the reasons for exclusion from the latter must also be documented. Similarly, for all patients included in an analysis population, the measurements of all

important variables should be accounted for at all relevant time-points. Further available information on additional patients screened for entry but not randomised should also be summarised (see Section 10.1).

The effect of all losses of patients or data, withdrawals from treatment and major protocol violations on the main analyses of the primary variable(s) should be considered carefully. Patients lost to follow up or withdrawn from treatment should be identified, and a descriptive analysis of them provided, including the reasons for their loss and its relationship to treatment and outcome.

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

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

The play of chance may lead to unforeseen imbalances between the treatment groups in terms of baseline measurements not pre-defined as covariates in the analysis plan but having some prognostic importance nevertheless. This is best dealt with by showing that a subsidiary analysis which 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 quantified.

In general, sparing use should be made of unplanned subsidiary analyses. Subsidiary analyses are often carried out when it is thought that the treatment effect may vary according to some other factor or factors. An attempt may then be made to identify subgroups of patients for whom the effect is particularly beneficial. The potential dangers of over-interpretation of unplanned subgroup analyses are well known (see also Section 10.4), and should be carefully avoided. Although similar problems of interpretation arise if a treatment appears to have no benefit, or an adverse effect, in a subgroup of patients, such possibilities must be properly assessed and should therefore always be reported.

Finally statistical judgement should always 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 be a signatory to it.

14. OUTLINE OF SAFETY ISSUES

Controlled clinical trials are rarely sufficiently powerful to detect infrequent adverse effects of treatments. Nevertheless, it is important to monitor Phase III clinical trials for those adverse effects which do become apparent. Few Phase III clinical trials nominate a safety variable as a primary variable in the protocol but, if claims of superior safety or tolerance are made for a product, they should, in general, be supported by relevant evidence in confirmatory studies.

In any clinical trial the methods and measurements chosen to establish the safety of a substance will depend on a number of factors, including knowledge of the adverse effects of closely related substances, information from pre clinical and earlier clinical studies, and possible consequences of the pharmacodynamic/ pharmacokinetic properties of the particular product, the mode of administration, the type of patients to be studied, and the duration of the study. Laboratory variables, vital signs, and adverse events form the main body of the safety data.

For the purposes of overall safety assessment, the study population is usually defined as those patients who received at least one dose of the product. Safety variables should subsequently be collected as comprehensively as possible from these patients, whatever their fate. Additional safety analyses may be required in specific sub populations, such as the elderly, the severely ill, or those who have a common concomitant treatment. These analyses may need to address more specific safety issues.

All safety variables will require attention during analysis, and the broad approach should be indicated in the protocol. All available data in the study population should be included in the analysis. All adverse events should be reported, whether or not they are considered to be related to treatment. Definitions of measurement units and reference ranges of laboratory variables should be made with care; if different units or different reference ranges appear in the same trial (e.g. if more than one laboratory is involved), then measurements should be appropriately standardised to allow a unified analysis.

The investigation of safety and tolerability is a multi-dimensional problem. Although some specific adverse effects of any substance can usually be anticipated and specifically monitored, the range of possible adverse effects is very large, and new and unforeseeable reactions are always possible. This background underlies the statistical difficulties associated with the analytical evaluation of safety and tolerability of products, and means that confirmatory information from Phase III clinical trials is the exception rather than the rule in the safety field. In most trials the safety implications are best addressed by applying descriptive statistics to the safety data, supplemented by confidence intervals wherever this aids interpretation. Hypothesis tests are sometimes useful either as an aid to evaluating a specific difference of interest, or as a "flagging" device applied to a large number of safety variables to highlight differences worth further attention. If hypothesis tests *are* used, statistical adjustments for multiplicity to maintain the type I error are usually inappropriate, because the type II error is of more concern. In the majority of studies investigators are seeking to establish that there are no notable differences in safety compared with either a comparator substance or a placebo and, in principle, a firm conclusion that no differences exists cannot be drawn from the failure to demonstrate them through hypothesis tests. The low power attached to most safety variable comparisons exacerbates this problem. These considerations, which are similar to those which apply to equivalence trials, generally favour the use of confidence intervals in preference to hypothesis tests for the analysis of safety data.

For more details of the reporting requirements relating to safety see Chapter 12 of the note for guidance on *Structure and Content of Clinical Study Reports.*

15. SUMMARISING THE CLINICAL DATABASE

Regulatory review inevitably involves the summary and synthesis of the evidence on safety and efficacy from all the reported clinical trials. When carrying out this task, applicants

should conform to the requirements of the relevant section of the Notice to Applicants. An overall summary may be accompanied by a formal meta-analysis, if appropriate.

Within the summary a number of areas of specific statistical interest arise: describing the demography and clinical features of the population treated during the course of the clinical trial programme; addressing the key questions of efficacy by considering the results of the relevant (usually controlled) trials and highlighting the degree to which they reinforce or contradict each other; working up the safety information which is available from the combined database of all the studies whose results contribute to the marketing authorisation application and identifying potential safety issues. During the design of a clinical programme careful attention should be paid to the uniform definition and collection of measurements which will facilitate subsequent interpretation of the series of trials, particularly if they are likely to be combined across trials. A common dictionary for recording the details of medication, medical history and adverse events should be selected and used. A common definition of the primary and secondary variables is nearly always worthwhile, and essential for meta-analysis. The manner of measuring key efficacy variables, the timing of assessments relative to randomisation/entry, the handling of protocol violators and deviators, and perhaps the definition of prognostic factors, should all be kept compatible unless there are compelling reasons not to do so, at least in the pivotal studies.

Any statistical procedures used to combine data across trials must be described in detail. Attention should be paid to the possibility of bias associated with the selection of trials, to the homogeneity of their results, and to the proper modelling of the various sources of variation. The sensitivity of conclusions to the assumptions and selections made should be explored.

15.1 Efficacy data

Individual clinical trials should always be large enough to satisfy their objectives. Additional valuable information may also be gained by summarising a series of clinical trials which address essentially identical key efficacy questions. The main results of such a set of studies should always be presented in an identical form to permit comparison, usually in tables or graphs which focus on estimates plus confidence limits. The use of meta-analytic techniques to combine these estimates is often a useful addition, because it allows a more precise overall estimate of the size of the treatment effects to be generated, and provides a complete and concise summary of the results of the trials. Under *exceptional* circumstances a meta-analytic approach may also be the most appropriate way, or the only way, of providing sufficient overall evidence of efficacy via an overall hypothesis test.

15.2 Safety data

In summarising safety data it is important to examine the safety database thoroughly for any indications of potential toxicity, and to follow up any indications by looking for an associated supportive pattern of observations. The combination of the safety data from all human exposure to the product should provide the main source of information, because its larger sample size provides the best chance of detecting the rarer adverse events and, perhaps, of estimating their approximate incidence. However, incidence data from this database are difficult to evaluate without a natural comparator group, and data from comparative studies are especially valuable in overcoming this difficulty. The results from studies which use a common comparator (placebo or specific active comparator) should be combined and presented separately for each comparator providing sufficient data.

All indications of potential toxicity arising from statistical exploration of the data should be reported. The evaluation of the reality of these potential adverse effects should take account of the issue of multiplicity arising from the numerous statistical comparisons made. The evaluation should also make appropriate use of survival analysis methods to exploit the potential relationship of the incidence of adverse events to duration of exposure and/or follow-up. The risks associated with identified adverse effects should be appropriately quantified to allow a proper assessment of the risk/benefit relationship.