

ICH E9

- [ICH E9 Statistical Principles for Clinical Trials](#)
- [ICH E9 \[REDACTED\]](#)
- [ICH E9\(R1\) Addendum: Statistical Principles for Clinical Trials](#)
- [ICH E9\(R1\) \[REDACTED\] E9\[REDACTED\]](#)

ICH E9 Statistical Principles for Clinical Trials

III	
ICH E9	IIIIIIIIII

I. INTRODUCTION

1.1 Background and Purpose

The efficacy and safety of medicinal products should be demonstrated by clinical trials which follow the guidance in 'Good Clinical Practice: Consolidated Guideline' (ICH E6) adopted by the ICH, 1 May 1996. The role of statistics in clinical trial design and analysis is acknowledged as essential in that ICH guideline. The proliferation of statistical research in the area of clinical trials coupled with the critical role of clinical research in the drug approval process and health care in general necessitate a succinct document on statistical issues related to clinical trials. This guidance is written primarily to attempt to harmonise the principles of statistical methodology applied to clinical trials for marketing applications submitted in Europe, Japan and the United States.

As a starting point, this guideline utilised the CPMP (Committee for Proprietary Medicinal Products) Note for Guidance entitled 'Biostatistical Methodology in Clinical Trials in Applications for Marketing Authorisations for Medicinal Products' (December, 1994). It was also influenced by 'Guidelines on the Statistical Analysis of Clinical Studies' (March, 1992) from the Japanese Ministry of Health and Welfare and the U.S. Food and Drug Administration document entitled 'Guideline for the Format and Content of the Clinical and Statistical Sections of a New Drug Application' (July, 1988). Some topics related to statistical principles and methodology are also embedded within other ICH guidelines, particularly those listed below. The specific guidance that contains related text will be identified in various sections of this document.

E1A:	The Extent of Population Exposure to Assess Clinical Safety
E2A:	Clinical Safety Data Management: Definitions and Standards for Expedited Reporting
E2B:	Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports
E2C:	Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs
E3:	Structure and Content of Clinical Study Reports
E4:	Dose-Response Information to Support Drug Registration

E5:	Ethnic Factors in the Acceptability of Foreign Clinical Data
E6:	Good Clinical Practice: Consolidated Guideline
E7:	Studies in Support of Special Populations: Geriatrics
E8:	General Considerations for Clinical Trials
E10:	Choice of Control Group in Clinical Trials
M1:	Standardisation of Medical Terminology for Regulatory Purposes
M3:	Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals.

This guidance is intended to give direction to sponsors in the design, conduct, analysis, and evaluation of clinical trials of an investigational product in the context of its overall clinical development. The document will also assist scientific experts charged with preparing application summaries or assessing evidence of efficacy and safety, principally from clinical trials in later phases of development.

1.2 Scope and Direction

The focus of this guidance is on statistical principles. It does not address the use of specific statistical procedures or methods. Specific procedural steps to ensure that principles are implemented properly are the responsibility of the sponsor. Integration of data across clinical trials is discussed, but is not a primary focus of this guidance. Selected principles and procedures related to data management or clinical trial monitoring activities are covered in other ICH guidelines and are not addressed here.

This guidance should be of interest to individuals from a broad range of scientific disciplines. However, it is assumed that the actual responsibility for all statistical work associated with clinical trials will lie with an appropriately qualified and experienced statistician, as indicated in ICH E6. The role and responsibility of the trial statistician (see Glossary), in collaboration with other clinical trial professionals, is to ensure that statistical principles are applied appropriately in clinical trials supporting drug development. Thus, the trial statistician should have a combination of education/training and experience sufficient to implement the principles articulated in this guidance.

For each clinical trial contributing to a marketing application, all important details of its design and conduct and the principal features of its proposed statistical analysis should be clearly specified in a protocol written before the trial begins. The extent to which the procedures in the protocol are followed and the primary analysis is planned a priori will contribute to the degree of confidence in the final results and conclusions of the trial. The protocol and subsequent amendments should be approved by the responsible personnel, including the trial statistician. The trial statistician should ensure that the protocol and any amendments cover all relevant statistical issues clearly and accurately, using technical terminology as appropriate.

The principles outlined in this guidance are primarily relevant to clinical trials conducted in the later phases of development, many of which are confirmatory trials of efficacy. In addition to efficacy,

confirmatory trials may have as their primary variable a safety variable (e.g. an adverse event, a clinical laboratory variable or an electrocardiographic measure), a pharmacodynamic or a pharmacokinetic variable (as in a confirmatory bioequivalence trial). Furthermore, some confirmatory findings may be derived from data integrated across trials, and selected principles in this guidance are applicable in this situation. Finally, although the early phases of drug development consist mainly of clinical trials that are exploratory in nature, statistical principles are also relevant to these clinical trials. Hence, the substance of this document should be applied as far as possible to all phases of clinical development.

Many of the principles delineated in this guidance deal with minimising bias (see Glossary) and maximising precision. As used in this guidance, the term 'bias' describes the systematic tendency of any factors associated with the design, conduct, analysis and interpretation of the results of clinical trials to make the estimate of a treatment effect (see Glossary) deviate from its true value. It is important to identify potential sources of bias as completely as possible so that attempts to limit such bias may be made. The presence of bias may seriously compromise the ability to draw valid conclusions from clinical trials.

Some sources of bias arise from the design of the trial, for example an assignment of treatments such that subjects at lower risk are systematically assigned to one treatment. Other sources of bias arise during the conduct and analysis of a clinical trial. For example, protocol violations and exclusion of subjects from analysis based upon knowledge of subject outcomes are possible sources of bias that may affect the accurate assessment of the treatment effect. Because bias can occur in subtle or unknown ways and its effect is not measurable directly, it is important to evaluate the robustness of the results and primary conclusions of the trial. Robustness is a concept that refers to the sensitivity of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis. Robustness implies that the treatment effect and primary conclusions of the trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches. The interpretation of statistical measures of uncertainty of the treatment effect and treatment comparisons should involve consideration of the potential contribution of bias to the p-value, confidence interval, or inference.

Because the predominant approaches to the design and analysis of clinical trials have been based on frequentist statistical methods, the guidance largely refers to the use of frequentist methods (see Glossary) 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 Glossary) and other approaches may be considered when the reasons for their use are clear and when the resulting conclusions are sufficiently robust.

II. CONSIDERATIONS FOR OVERALL CLINICAL DEVELOPMENT

2.1 Trial Context

2.1.1 Development Plan

The broad aim of the process of clinical development of a new drug is to find out whether there is a dose range and schedule at which the drug can be shown to be simultaneously safe and effective, to the extent that the risk-benefit relationship is acceptable. The particular subjects who may benefit from the drug, 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 (see ICH E8). 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. A marketing application 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 7.2). 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 statistical summary, overview or meta-analysis (see Glossary) 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 necessary common features of their designs are specified in advance. Other major statistical issues (if any) that are expected to affect a number of trials in a common plan should be addressed in that plan.

2.1.2 Confirmatory Trial

A confirmatory trial is an adequately controlled trial in which the hypotheses are stated in advance and evaluated. As a rule, confirmatory trials are necessary to provide firm evidence of efficacy or safety. In such trials the key hypothesis of interest follows directly from the trial's primary objective, is always pre-defined, and is the hypothesis that is subsequently tested when the trial is complete. 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 intended to provide firm evidence in support of claims and hence adherence to protocols and standard operating procedures is particularly important; unavoidable changes should be explained and documented, and their effect examined. A justification of the design of each such trial, and of other important statistical aspects such as the principal features of the planned analysis, should be set out in the protocol. Each trial should address only a limited number of questions.

Firm evidence in support of claims requires that the results of the confirmatory trials demonstrate that the investigational product under test has clinical benefits. The confirmatory trials should therefore be sufficient to answer each key clinical question relevant to the efficacy or safety claim clearly and definitively. In addition, it is important that the basis for generalisation (see Glossary) to the intended patient population is understood and explained; this may also influence the number and type (e.g. specialist or general practitioner) of centres and/or trials needed. The results of the confirmatory trial(s) should be robust. In some circumstances the weight of evidence from a single confirmatory trial may be sufficient.

2.1.3 Exploratory Trial

The rationale and design of confirmatory trials 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 trials, their objectives may not always lead to simple tests of pre-defined hypotheses. In addition, exploratory trials 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 trials cannot be the basis of the formal proof of efficacy, although they may contribute to the total body of relevant evidence.

Any individual trial may have both confirmatory and exploratory aspects. For example, in most confirmatory trials 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 trial which will be used for confirmatory proof and the aspects which will provide data for exploratory analysis.

2.2 Scope of Trials

2.2.1 Population

In the earlier phases of drug 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 subgroup of the total patient population for which the drug may eventually be indicated. However by the time the confirmatory trials are undertaken, the subjects in the trials should more closely mirror the target population. Hence, in these trials it is generally helpful to relax the inclusion and exclusion criteria as much as possible within the target population, while maintaining sufficient homogeneity to permit precise estimation of treatment effects. No individual clinical trial can be expected to be totally representative of future users, because of the possible influences of 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.

2.2.2 Primary and Secondary Variables

The primary variable ('target' variable, primary endpoint) should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial. There should generally be only one primary variable. This will usually be an efficacy variable, because the primary objective of most confirmatory trials is to provide strong scientific evidence regarding efficacy. Safety/tolerability may sometimes be the primary variable, and will always be an important consideration. Measurements relating to quality of life and health economics are further potential primary variables. The selection of the primary variable should 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 can provide a valid and reliable measure of some clinically relevant and important treatment benefit in the patient population described by the inclusion and exclusion criteria. The primary variable should generally be the one used when estimating the sample size (see section 3.5).

In many cases, the approach to assessing subject outcome may not be straightforward and should be carefully defined. For example, it is inadequate to specify mortality as a primary variable without further clarification; mortality may be assessed by comparing proportions alive at fixed points in time, or by comparing overall distributions of survival times over a specified interval. Another common example is a recurring event; the measure of treatment effect may again be a simple dichotomous variable (any occurrence during a specified interval), time to first occurrence, rate of occurrence (events per time units of observation), etc. The assessment of functional status over time in studying treatment for chronic disease presents other challenges in selection of the primary variable. There are many possible approaches, such as comparisons of the assessments done at the beginning and end of the interval of observation, comparisons of slopes calculated from all assessments throughout the interval, comparisons of the proportions of subjects exceeding or declining beyond a specified threshold, or comparisons based on methods for repeated measures data. To avoid multiplicity concerns arising from post hoc definitions, it is critical to specify in the protocol the precise definition of the primary variable as it will be used in the statistical analysis. In addition, the clinical relevance of the specific primary variable selected and the validity of the associated measurement procedures will generally need to be addressed and justified in the protocol.

The primary variable should be specified in the protocol, along with the rationale for its selection. Redefinition of the primary variable after unblinding will almost always be unacceptable, since the biases this introduces are difficult to assess. When the clinical effect defined by the primary objective is to be measured in more than one way, the protocol should identify one of the measurements as the primary variable on the basis of clinical relevance, importance, objectivity, and/or other relevant characteristics, whenever such selection is feasible.

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. The number of secondary variables should be limited and should be related to the limited number of questions to be answered in the trial.

2.2.3 Composite Variables

If a single primary variable cannot be selected from multiple measurements associated with the primary objective, another useful strategy is to integrate or combine the multiple measurements into a single or 'composite' variable, using a pre-defined algorithm. Indeed, the primary variable sometimes arises as a combination of multiple clinical measurements (e.g. the rating scales used in arthritis, psychiatric disorders and elsewhere). This approach addresses the multiplicity problem without requiring adjustment to the type I error. The method of combining the multiple measurements should be specified in the protocol, and an interpretation of the resulting scale should be provided in terms of the size of a clinically relevant benefit. When a composite variable is used as a primary variable, the components of this variable may sometimes be analysed separately, where clinically meaningful and validated. When a rating scale is used as a primary variable, it is especially important to address such factors as content validity (see Glossary), inter- and intra-rater reliability (see Glossary) and responsiveness for detecting changes in the severity of disease.

2.2.4 Global Assessment Variables

In some cases, 'global assessment' variables (see Glossary) are developed to measure the overall safety, overall efficacy, and/or overall usefulness of a treatment. This type of variable integrates objective variables and the investigator's overall impression about the state or change in the state of the subject, and is usually a scale of ordered categorical ratings. Global assessments of overall efficacy are well established in some therapeutic areas, such as neurology and psychiatry.

Global assessment variables generally have a subjective component. When a global assessment variable is used as a primary or secondary variable, fuller details of the scale should be included in the protocol with respect to:

- 1) the relevance of the scale to the primary objective of the trial;
- 2) the basis for the validity and reliability of the scale;
- 3) how to utilise the data collected on an individual subject to assign him/her to a unique category of the scale;
- 4) how to assign subjects with missing data to a unique category of the scale, or otherwise evaluate them.

If objective variables are considered by the investigator when making a global assessment, then those objective variables should be considered as additional primary, or at least important secondary, variables.

Global assessment of usefulness integrates components of both benefit and risk and reflects the decision making process of the treating physician, who must weigh benefit and risk in making product use decisions. A problem with global usefulness variables is that their use could in some cases lead to the result of two products being declared equivalent despite having very different profiles of beneficial and adverse effects. For example, judging the global usefulness of a treatment as equivalent or superior to an alternative may mask the fact that it has little or no efficacy but fewer adverse effects. Therefore it is not advisable to use a global usefulness variable as a primary variable. If global usefulness is specified as primary, it is important to consider specific efficacy and safety outcomes separately as additional primary variables.

2.2.5 Multiple Primary Variables

It may sometimes be desirable to use more than one primary variable, each of which (or a subset of which) could be sufficient to cover the range of effects of the therapies. The planned manner of interpretation of this type of evidence should be carefully spelled out. It should be clear whether an impact on any of the variables, some minimum number of them, or all of them, would be considered necessary to achieve the trial objectives. The primary hypothesis or hypotheses and parameters of interest (e.g. mean, percentage, distribution) should be clearly stated with respect to the primary variables identified, and the approach to statistical inference described. The effect on the type I error should be explained because of the potential for multiplicity problems (see Section 5.6); the method of controlling type I error should be given in the protocol. The extent of intercorrelation among the proposed primary variables may be considered in evaluating the impact

on type I error. If the purpose of the trial is to demonstrate effects on all of the designated primary variables, then there is no need for adjustment of the type I error, but the impact on type II error and sample size should be carefully considered.

2.2.6 Surrogate Variables

When direct assessment of the clinical benefit to the subject through observing actual clinical efficacy is not practical, indirect criteria (surrogate variables - see Glossary) may be considered. Commonly accepted surrogate variables are used in a number of indications where they are believed to be reliable predictors of clinical benefit. There are two principal concerns with the introduction of any proposed surrogate variable. First, it may not be a true predictor of the clinical outcome of interest. For example it may measure treatment activity associated with one specific pharmacological mechanism, but may not provide full information on the range of actions and ultimate effects of the treatment, whether positive or negative. There have been many instances where treatments showing a highly positive effect on a proposed surrogate have ultimately been shown to be detrimental to the subjects' clinical outcome; conversely, there are cases of treatments conferring clinical benefit without measurable impact on proposed surrogates. Secondly, proposed surrogate variables may not yield a quantitative measure of clinical benefit that can be weighed directly against adverse effects. Statistical criteria for validating surrogate variables have been proposed but the experience with their use is relatively limited. In practice, the strength of the evidence for surrogacy depends upon (i) the biological plausibility of the relationship, (ii) the demonstration in epidemiological studies of the prognostic value of the surrogate for the clinical outcome and (iii) evidence from clinical trials that treatment effects on the surrogate correspond to effects on the clinical outcome. Relationships between clinical and surrogate variables for one product do not necessarily apply to a product with a different mode of action for treating the same disease.

2.2.7 Categorised Variables

Dichotomisation or other categorisation of continuous or ordinal variables may sometimes be desirable. 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.

The reduction of diastolic blood pressure below 90mmHg is a common dichotomisation. Categorisations are most useful when they have clear clinical relevance. The criteria for categorisation should be pre-defined and specified in the protocol, as knowledge of trial results could easily bias the choice of such criteria. Because categorisation normally implies a loss of information, a consequence will be a loss of power in the analysis; this should be accounted for in the sample size calculation.

2.3 Design Techniques to Avoid Bias

The most important design techniques for avoiding bias in clinical trials are blinding and randomisation, and these should be normal features of most controlled clinical trials intended to be included in a marketing application. 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 subject 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 subject, not even as a code letter. This approach will be assumed in Section 2.3.1 and most of Section 2.3.2, exceptions being considered at the end.

Bias can also be reduced at the design stage by specifying procedures in the protocol aimed at minimising any anticipated irregularities in trial 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 also to handle the problems that do occur in the analysis of data.

2.3.1 Blinding

Blinding or masking 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 subjects, their subsequent care, the attitudes of subjects 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 have passed.

A double-blind trial is one in which neither the subject nor any of the investigator or sponsor staff who are involved in the treatment or clinical evaluation of the subjects are aware of the treatment received. This includes anyone determining subject eligibility, evaluating endpoints, or assessing compliance with the protocol. This level of blinding is maintained throughout the conduct of the trial, and only when the data are cleaned to an acceptable level of quality will appropriate personnel be unblinded. If any of the sponsor staff who are not involved in the treatment or clinical evaluation of the subjects are required to be unblinded to the treatment code (e.g. bioanalytical scientists, auditors, those involved in serious adverse event reporting), the sponsor should have adequate standard operating procedures to guard against inappropriate dissemination of treatment codes. In a single-blind trial the investigator and/or his staff are aware of the treatment but the subject is not, or vice versa. 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 (appearance, taste, etc.) either before or during administration, and that the blind is maintained appropriately during the whole trial.

Difficulties in achieving the double-blind ideal can arise: the treatments may be of a completely different nature, for example, surgery and drug therapy; two drugs may have different formulations and, although they could be made indistinguishable by the use of capsules, changing the formulation might also change the pharmacokinetic and/or pharmacodynamic properties and hence require that bioequivalence of the formulations 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' (see Glossary) technique. This technique may sometimes force an administration scheme that is sufficiently unusual to influence adversely the motivation and compliance of the subjects. Ethical difficulties may also interfere with its use when, for example, it entails dummy operative procedures. Nevertheless, extensive efforts should be made to overcome these difficulties.

The double-blind nature of some clinical trials may be partially compromised by apparent treatment induced effects. In such cases, blinding may be improved by blinding investigators and relevant sponsor staff to certain test results (e.g. selected clinical laboratory measures). Similar approaches (see below) to minimising bias in open-label trials should be considered in trials where unique or specific treatment effects may lead to unblinding individual patients.

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. Single-blind and open-label trials provide additional flexibility, but it is particularly important that the investigator's knowledge of the next treatment should not influence the decision to enter the subject; this decision should precede knowledge of the randomised treatment. For these trials, consideration should be given to the use of a centralised randomisation method, such as telephone randomisation, to administer the assignment of randomised treatment. In addition, clinical assessments should be made by medical staff who are not involved in treating the subjects 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 and primary variables should be as objective as possible. The reasons for the degree of blinding adopted should be explained in the protocol, together with steps taken to minimise bias by other means. For example, the sponsor should have adequate standard operating procedures to ensure that access to the treatment code is appropriately restricted during the process of cleaning the database prior to its release for analysis.

Breaking the blind (for a single subject) should be considered only when knowledge of the treatment assignment is deemed essential by the subject's physician for the subject's care. 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 and timing for revealing the treatment assignments should be documented.

In this document, the blind review (see Glossary) of data refers to the checking of data during the period of time between trial completion (the last observation on the last subject) and the breaking of the blind.

2.3.2 Randomisation

Randomisation introduces a deliberate element of chance into the assignment of treatments to subjects in a clinical trial. During subsequent analysis of the trial data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. It also tends to produce treatment groups in which the distributions of prognostic factors, known and unknown, are similar. In combination with blinding, randomisation helps to avoid possible bias in the selection and allocation of subjects arising from the predictability of treatment assignments.

The randomisation schedule of a clinical trial documents the random allocation of treatments to subjects. In the simplest situation it is a sequential list of treatments (or treatment sequences in a crossover trial) or corresponding codes by subject number. The logistics of some trials, such as those with a screening phase, may make matters more complicated, but the unique pre-planned assignment of treatment, or treatment sequence, to subject should be clear. Different trial designs will require different procedures for generating randomisation schedules. The randomisation schedule should be reproducible (if the need arises).

Although unrestricted randomisation is an acceptable approach, some advantages can generally be gained by randomising subjects in blocks. This helps to increase the comparability of the treatment groups, particularly when subject 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 crossover trials it provides the means of obtaining balanced designs with their greater efficiency and easier interpretation. Care should be taken to choose block lengths that are sufficiently short to limit possible imbalance, but that are long enough to avoid predictability towards the end of the sequence in a block. Investigators and other relevant staff 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. (Theoretically, in a double-blind trial predictability does not matter, but the pharmacological effects of drugs may provide the opportunity for intelligent guesswork.)

In multicentre trials (see Glossary) the randomisation procedures should be organised centrally. It is advisable to have a separate random scheme for each centre, i.e. to stratify by centre or to allocate several whole blocks to each 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. The use of a dynamic allocation procedure (see below) may help to achieve balance across a number of stratification factors simultaneously provided the rest of the trial procedures can be adjusted to accommodate an approach of this type. Factors on which randomisation has been stratified should be accounted for later in the analysis.

The next subject to be randomised into a trial 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 for the next subject should only be allocated when entry of that subject to the randomised part of the trial has been confirmed. Details of the randomisation that facilitate predictability (e.g. block length) should not be contained in the trial protocol. The randomisation schedule itself should be filed securely by the sponsor or an independent party in a manner that ensures that blindness is properly maintained throughout the trial. Access to the randomisation schedule during the trial should take into account the possibility that, in an emergency, the blind may have to be broken for any subject. The procedure to be followed, the necessary documentation, and the subsequent treatment and assessment of the subject should all be described in the protocol.

Dynamic allocation is an alternative procedure in which the allocation of treatment to a subject is influenced by the current balance of allocated treatments and, in a stratified trial, by the stratum to which the subject belongs and the balance within that stratum. Deterministic dynamic allocation procedures should be avoided and an appropriate element of randomisation should be incorporated for each treatment allocation. Every effort should be made to retain the double-blind status of the trial. For example, 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, features that can be valuable in certain types of multicentre trial. The usual system of pre-packing and labelling drug supplies for double-blind trials can then be followed, but the order of their use is no longer sequential. It is desirable to use appropriate computer algorithms to keep personnel at

the central trial office blind to the treatment code. The complexity of the logistics and potential impact on the analysis should be carefully evaluated when considering dynamic allocation.

III. TRIAL DESIGN CONSIDERATIONS

3.1 Design Configuration

3.1.1 Parallel Group Design

The most common clinical trial design for confirmatory trials is the parallel group design in which subjects 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 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, as with other designs, there may be additional features of the trial that complicate the analysis and interpretation (e.g. covariates, repeated measurements over time, interactions between design factors, protocol violations, dropouts (see Glossary) and withdrawals).

3.1.2 Crossover Design

In the crossover design, each subject is randomised to a sequence of two or more treatments, and hence acts as his own control for treatment comparisons. This simple manoeuvre is attractive primarily because it reduces the number of subjects and usually the number of assessments needed to achieve a specific power, sometimes to a marked extent. In the simplest 2×2 crossover design each subject 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 subject receiving all n treatments. Numerous variations exist, such as designs in which each subject receives a subset of $n(>2)$ treatments, or ones in which treatments are repeated within a subject.

Crossover designs have a number of problems that 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 2×2 design the carryover effect cannot be statistically distinguished from the interaction between treatment and period and the test for either of these effects lacks power because the corresponding contrast is 'between subject'. This problem is less acute in higher order designs, but cannot be entirely dismissed.

When the crossover 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 fully within the treatment period. The washout 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.

There are additional problems that need careful attention in crossover trials. The most notable of these are the complications of analysis and interpretation arising from the loss of subjects. Also, the potential for carryover leads to difficulties in assigning adverse events which occur in later treatment periods to the appropriate treatment. These, and other issues, are described in ICH E4. The crossover design should generally be restricted to situations where losses of subjects from the trial are expected to be small.

A common, and generally satisfactory, use of the 2×2 crossover design is to demonstrate the bioequivalence of two formulations of the same medication. In this particular application in healthy volunteers, carryover effects on the relevant pharmacokinetic variable are most 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 drug is detectable at the start of each period.

3.1.3 Factorial Designs

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

Another important use of the factorial design is to establish the dose-response characteristics of the simultaneous use of treatments C and D, especially when the efficacy of each monotherapy has been established at some dose in prior trials. 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 m×n 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 (see ICH E4).

In some cases, the 2×2 design may be used to make efficient use of clinical trial subjects by evaluating the efficacy of the two treatments with the same number of subjects as would be required to evaluate the efficacy of either one alone. This strategy has proved to be particularly valuable for very large mortality trials. The efficiency and validity of this approach depends upon the absence of 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 crossover 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.

3.2 Multicentre Trials

Multicentre trials are carried out for two main reasons. Firstly, a multicentre trial is an accepted way of evaluating a new medication more efficiently; under some circumstances, it may present the only practical means of accruing sufficient subjects to satisfy the trial objective within a

reasonable time-frame. Multicentre trials of this nature may, in principle, be carried out at any stage of clinical development. They may have several centres with a large number of subjects 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 trial may be designed as a multicentre (and multi-investigator) trial primarily to provide a better basis for the subsequent generalisation of its findings.

This arises from the possibility of recruiting the subjects from a wider population and of administering the medication in a broader range of clinical settings, thus presenting an experimental situation that 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. Such a trial would be a confirmatory trial in the later phases of drug development and would be likely to involve a large number of investigators and centres. It might sometimes be conducted in a number of different countries in order to facilitate generalisability (see Glossary) even further.

If a multicentre trial is to be meaningfully interpreted and extrapolated, then the manner in which the protocol is implemented should be clear and similar at all centres. Furthermore the usual sample size and power calculations depend upon the assumption that the differences between the compared treatments in the centres are unbiased estimates of the same quantity. It is important to design the common protocol and to conduct the trial with this background in mind. Procedures should be standardised as completely as possible. Variation of evaluation criteria and schemes can be reduced by investigator meetings, by the training of personnel in advance of the trial and by careful monitoring during the trial. Good design should generally aim to achieve the same distribution of subjects to treatments within each centre and good management should maintain this design objective. Trials that avoid excessive variation in the numbers of subjects per centre and trials that avoid a few very small centres have advantages if it is later found necessary to take into account the heterogeneity of the treatment effect from centre to centre, because they reduce the differences between different weighted estimates of the treatment effect. (This point does not apply to trials in which all centres are very small and in which centre does not feature in the analysis.) 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 sponsor's claims.

In the simplest multicentre trial, each investigator will be responsible for the subjects recruited at one hospital, so that 'centre' is identified uniquely by either investigator or hospital. In many trials, however, the situation is more complex. One investigator may recruit subjects from several hospitals; one investigator may represent a team of clinicians (subinvestigators) who all recruit subjects from their own clinics at one hospital or at several associated hospitals. Whenever there is room for doubt about the definition of centre in a statistical model, the statistical section of the protocol (see Section 5.1) should clearly define the term (e.g. by investigator, location or region) in the context of the particular trial. In most instances centres can be satisfactorily defined through the investigators and ICH E6 provides relevant guidance in this respect. In cases of doubt the aim should be to define centres so as to achieve homogeneity in the important factors affecting the measurements of the primary variables and the influence of the treatments. Any rules for combining centres in the analysis should be justified and specified prospectively in the protocol

where possible, but in any case decisions concerning this approach should always be taken blind to treatment, for example at the time of the blind review.

The statistical model to be adopted for the estimation and testing of treatment effects should be described in the protocol. The main treatment effect may be investigated first using a model which allows for centre differences, but does not include a term for treatment-by-centre interaction. If the treatment effect is homogeneous across centres, the routine inclusion of interaction terms in the model reduces the efficiency of the test for the main effects. In the presence of true heterogeneity of treatment effects, the interpretation of the main treatment effect is controversial.

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

If positive treatment effects are found in a trial with appreciable numbers of subjects per centre, there should generally be an exploration of the heterogeneity of treatment effects across centres, as this may affect the generalisability of the conclusions. Marked heterogeneity may be identified by graphical display of the results of individual centres or by analytical methods, such as a significance test of the treatment-by-centre interaction. When using such a statistical significance test, it is important to recognise that this generally has low power in a trial designed to detect the main effect of treatment.

If heterogeneity of treatment effects is found, this should be interpreted with care and vigorous attempts should be made to find an explanation in terms of other features of trial management or subject characteristics. Such an explanation will usually suggest appropriate further analysis and interpretation. In the absence of an explanation, heterogeneity of treatment effect as evidenced, for example, by marked quantitative interactions (see Glossary) implies that alternative estimates of the treatment effect may be required, giving different weights to the centres, in order to substantiate the robustness of the estimates of treatment effect. It is even more important to understand the basis of any heterogeneity characterised by marked qualitative interactions (see Glossary), and failure to find an explanation may necessitate further clinical trials before the treatment effect can be reliably predicted.

Up to this point the discussion of multicentre trials has been based on the use of fixed effect models. Mixed models may also be used to explore the heterogeneity of the treatment effect. These models consider centre and treatment-by-centre effects to be random, and are especially relevant when the number of sites is large.

3.3 Type of Comparison

3.3.1 Trials to Show Superiority

Scientifically, efficacy is most convincingly established by demonstrating superiority to placebo in a placebo-controlled trial, by showing superiority to an active control treatment or by demonstrating a dose-response relationship. This type of trial is referred to as a 'superiority' trial (see Glossary). Generally in this guidance superiority trials are assumed, unless it is explicitly stated otherwise.

For serious illnesses, when a therapeutic treatment which has been shown to be efficacious by superiority trial(s) exists, a placebo-controlled trial may be considered unethical. In that case the scientifically sound use of an active treatment as a control should be considered. The appropriateness of placebo control vs. active control should be considered on a trial by trial basis.

3.3.2 Trials to Show Equivalence or Non-inferiority

In some cases, an investigational product is 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 (see Glossary) and the other is a 'non-inferiority' trial (see Glossary).

Many active control trials are designed to show that the efficacy of an investigational product is no worse than that of the active comparator, and hence fall into the latter category. Another possibility is a trial in which multiple doses of the investigational drug are compared with the recommended dose or multiple doses of the standard drug. The purpose of this design is simultaneously to show a dose-response relationship for the investigational product and to compare the investigational product with the active control.

Active control equivalence or non-inferiority trials may also incorporate a placebo, thus pursuing multiple goals in one trial; for example, they may establish superiority to placebo and hence validate the trial design and simultaneously evaluate the degree of similarity of efficacy and safety to the active comparator. There are well known difficulties associated with the use of the active control equivalence (or non-inferiority) trials that do not incorporate a placebo or do not use multiple doses of the new drug. These relate to the implicit lack of any measure of internal validity (in contrast to superiority trials), thus making external validation necessary. The equivalence (or non-inferiority) trial is not conservative in nature, so that many flaws in the design or conduct of the trial will tend to bias the results towards a conclusion of equivalence. For these reasons, the design features of such trials should receive special attention and their conduct needs special care. For example, it is especially important to minimise the incidence of violations of the entry criteria, non-compliance, withdrawals, losses to follow-up, missing data and other deviations from the protocol, and also to minimise their impact on the subsequent analyses.

Active comparators should be chosen with care. An example of a suitable active comparator would be a widely used therapy whose efficacy in the relevant indication has been clearly established and quantified in well designed and well documented superiority trial(s) and which can be reliably expected to exhibit similar efficacy in the contemplated active control trial. To this end, the new trial should have the same important design features (primary variables, the dose of the active comparator, eligibility criteria, etc.) as the previously conducted superiority trials in which the active comparator clearly demonstrated clinically relevant efficacy, taking into account advances in medical or statistical practice relevant to the new trial.

It is vital that the protocol of a trial designed to demonstrate equivalence or non-inferiority contain a clear statement that this is its explicit intention. An equivalence margin should be specified in the protocol; this margin is the largest difference that can be judged as being clinically acceptable and should be smaller than differences observed in superiority trials of the active comparator. For the active control equivalence trial, both the upper and the lower equivalence margins are needed, while only the lower margin is needed for the active control non-inferiority trial. The choice of equivalence margins should be justified clinically.

Statistical analysis is generally based on the use of confidence intervals (see Section 5.5). For equivalence trials, two-sided confidence intervals should be used. Equivalence is inferred when the entire confidence interval falls within the equivalence margins. Operationally, this is equivalent to the method of using two simultaneous one-sided tests to test the (composite) null hypothesis that the treatment difference is outside the equivalence margins versus the (composite) alternative hypothesis that the treatment difference is within the margins. Because the two null hypotheses are disjoint, the type I error is appropriately controlled. For non-inferiority trials a one-sided interval should be used. The confidence interval approach has a one-sided hypothesis test counterpart for testing the null hypothesis that the treatment difference (investigational product minus control) is equal to the lower equivalence margin versus the alternative that the treatment difference is greater than the lower equivalence margin. The choice of type I error should be a consideration separate from the use of a one-sided or two-sided procedure. Sample size calculations should be based on these methods (see Section 3.5).

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

There are also special issues in the choice of analysis sets. Subjects who withdraw or dropout of the treatment group or the comparator group will tend to have a lack of response, and hence the results of using the full analysis set (see Glossary) may be biased toward demonstrating equivalence (see Section 5.2.3).

3.3.3 Trials to Show Dose-response Relationship

How response is related to the dose of a new investigational product is a question to which answers may be obtained in all phases of development, and by a variety of approaches (see ICH E4). Dose-response trials 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 should be addressed using the data collected at a number of doses under investigation, including a placebo (zero dose) wherever appropriate. For this purpose the application of procedures to estimate the relationship between dose and response, including the construction of confidence intervals and the use of graphical methods, is as important as the use of statistical tests. The hypothesis tests that are used may need to be tailored to the natural ordering of doses or to particular questions regarding the shape of the dose-response curve (e.g. monotonicity). The details of the planned statistical procedures should be given in the protocol.

3.4 Group Sequential Designs

Group sequential designs are used to facilitate the conduct of interim analysis (see section 4.5 and Glossary). While group sequential designs are not the only acceptable types of designs permitting interim analysis, they are the most commonly applied because it is more practicable to assess grouped subject outcomes at periodic 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 of the availability of information on treatment outcomes and subject treatment assignments (i.e. blind breaking, see Section 4.5). An Independent Data Monitoring Committee (see Glossary) may be used to review or to conduct the interim analysis of data arising from a group sequential design (see Section 4.6). While the design has been most widely and successfully used in large, long-term trials of mortality or major non-fatal endpoints, its use is growing in other circumstances. In particular, it is recognised that safety must be monitored in all trials and therefore the need for formal procedures to cover early stopping for safety reasons should always be considered.

3.5 Sample Size

The number of subjects in a clinical trial should always be large enough to provide a reliable answer 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 or important secondary objectives may need larger numbers of subjects than a trial sized on the basis of the primary efficacy question (see, for example, ICH E1a).

Using the usual method for determining the appropriate sample size, the following items should be specified: a primary variable, the test statistic, the null hypothesis, the alternative ('working') hypothesis at the chosen dose(s) (embodying consideration of the treatment difference to be detected or rejected at the dose and in the subject 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 violations. 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 trial.

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. It is important to investigate the sensitivity of the sample size estimate to a variety of deviations from these assumptions and this may be facilitated by providing a range of sample sizes appropriate for a reasonable range of deviations from assumptions. In confirmatory trials, assumptions should normally be based on published data or on the results of earlier trials. 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 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 may be 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 10% to 20%; it is in the sponsor's interest to keep this figure as low as feasible especially in the case of trials that are difficult or impossible to repeat. Alternative values to the conventional levels of type I and type II error may be acceptable or even preferable in some cases.

Sample size calculations should refer to the number of subjects required for the primary analysis. If this is the 'full analysis set', estimates of the effect size may need to be reduced compared to the per protocol set (see Glossary). This is to allow for the dilution of the treatment effect arising from the inclusion of data from patients who have withdrawn from treatment or whose compliance is poor. The assumptions about variability may also need to be revised.

The sample size of an equivalence trial or a non-inferiority trial (see Section 3.3.2) should normally be based on the objective of obtaining a confidence interval for the treatment difference that shows that the treatments differ at most by a clinically acceptable difference. When the power of an equivalence trial is assessed at a true difference of zero, then the sample size necessary to achieve this power is underestimated if the true difference is not zero. When the power of a non-inferiority trial is assessed at a zero difference, then the sample size needed to achieve that power will be underestimated if the effect of the investigational product is less than that of the active control. The choice of a 'clinically acceptable' difference needs justification with respect to its meaning for future patients, and may be smaller than the 'clinically relevant' difference referred to above in the context of superiority trials designed to establish that a difference exists.

The exact 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 guideline and the true treatment difference. The design of the stopping guideline should take into account the consequent distribution of the sample size, usually embodied in the expected and maximum sample sizes.

When event rates are lower than anticipated or variability is larger than expected, methods for sample size re-estimation are available without unblinding data or making treatment comparisons (see Section 4.4).

3.6 Data Capture and Processing

The collection of data and transfer of data from the investigator to the sponsor can take place through a variety of media, including paper case record forms, remote site monitoring systems, medical computer systems and electronic transfer. Whatever data capture instrument is used, the form and content of the information collected should be in full accordance with the protocol and should be established in advance of the conduct of the clinical trial. It should focus on the data necessary to implement the planned analysis, including the context information (such as timing assessments relative to dosing) necessary to confirm protocol compliance or identify important protocol deviations. 'Missing values' should be distinguishable from the 'value zero' or 'characteristic absent'.

The process of data capture through to database finalisation should be carried out in accordance with GCP (see ICH E6, Section 5). Specifically, timely and reliable processes for recording data and rectifying errors and omissions are necessary to ensure delivery of a quality database and the achievement of the trial objectives through the implementation of the planned analysis.

IV. TRIAL CONDUCT CONSIDERATIONS

4.1 Trial Monitoring and Interim Analysis

Careful conduct of a clinical trial according to the protocol has a major impact on the credibility of the results (see ICH E6). Careful monitoring can ensure that difficulties are noticed early and their occurrence or recurrence minimised.

There are two distinct types of monitoring that generally characterise confirmatory clinical trials sponsored by the pharmaceutical industry. One type of monitoring concerns the oversight of the quality of the trial, while the other type involves breaking the blind to make treatment comparisons (i.e. interim analysis). Both types of trial monitoring, in addition to entailing different staff responsibilities, involve access to different types of trial data and information, and thus different principles apply for the control of potential statistical and operational bias.

For the purpose of overseeing the quality of the trial the checks involved in trial monitoring may include whether the protocol is being followed, the acceptability of data being accrued, the success of planned accrual targets, the appropriateness of the design assumptions, success in keeping patients in the trials, etc. (see Sections 4.2 to 4.4). This type of monitoring does not require access to information on comparative treatment effects, nor unblinding of data and therefore has no impact on type I error. The monitoring of a trial for this purpose is the responsibility of the sponsor (see ICH E6) and can be carried out by the sponsor or an independent group selected by the sponsor. The period for this type of monitoring usually starts with the selection of the trial sites and ends with the collection and cleaning of the last subject's data.

The other type of trial monitoring (interim analysis) involves the accruing of comparative treatment results. Interim analysis requires unblinded (i.e. key breaking) access to treatment group assignment (actual treatment assignment or identification of group assignment) and comparative treatment group summary information. This necessitates that the protocol (or appropriate amendments prior to a first analysis) contains statistical plans for the interim analysis to prevent certain types of bias. This is discussed in Sections 4.5 & 4.6.

4.2 Changes in Inclusion and Exclusion Criteria

Inclusion and exclusion criteria should remain constant, as specified in the protocol, throughout the period of subject recruitment. Changes may occasionally be appropriate, for example, in long term trials, where 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. 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 modifications to the planned analysis, such as stratifying the analysis according to modified inclusion/exclusion criteria.

4.3 Accrual Rates

In trials with a long time-scale for the accrual of subjects, 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 alleviate concerns about selective entry and other aspects of quality. In a multicentre trial these considerations apply to the individual centres.

4.4 Sample Size Adjustment

In long term 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 clinical study 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 3.5).

4.5 Interim Analysis and Early Stopping

An interim analysis is any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to formal completion of a trial. Because the number, methods and consequences of these comparisons affect the interpretation of the trial, all interim analyses should be carefully planned in advance and described in the protocol. Special circumstances may dictate the need for an interim analysis that was not defined at the start of a trial. In these cases, a protocol amendment describing the interim analysis should be completed prior to unblinded access to treatment comparison data. When an interim analysis is planned with the intention of deciding whether or not to terminate a trial, this is usually accomplished by the use of a group sequential design which employs statistical monitoring schemes as guidelines (see Section 3.4). The goal of such an interim analysis is to stop the trial early if the superiority of the treatment under study is clearly established, if the demonstration of a relevant treatment difference has become unlikely or if unacceptable adverse effects are apparent. Generally, boundaries for monitoring efficacy require more evidence to terminate a trial early (i.e. they are more conservative) than boundaries for monitoring safety. When the trial design and monitoring objective involve multiple endpoints then this aspect of multiplicity may also need to be taken into account.

The protocol should describe the schedule of interim analyses, or at least the considerations which will govern its generation, for example if flexible alpha spending function approaches are to be employed; further details may be given in a protocol amendment before the time of the first interim analysis. The stopping guidelines and their properties should be clearly described in the protocol or amendments. The potential effects of early stopping on the analysis of other important variables should also be considered. This material should be written or approved by the Data Monitoring Committee (see Section 4.6), when the trial has one. Deviations from the planned procedure always bear the potential of invalidating the trial results. If it becomes necessary to make changes to the trial, any consequent changes to the statistical procedures should be specified in an amendment to the protocol at the earliest opportunity, especially discussing the

impact on any analysis and inferences that such changes may cause. The procedures selected should always ensure that the overall probability of type I error is controlled.

The execution of an interim analysis should 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 the characteristics of patients to be recruited or biases in treatment comparisons. This principle may be applied to all investigator staff and to staff employed by the sponsor except for those who are directly involved in the execution of the interim analysis. Investigators should only be informed about the decision to continue or to discontinue the trial, or to implement modifications to trial procedures.

Most clinical trials intended to support the efficacy and safety of an investigational product should proceed to full completion of planned sample size accrual; trials should be stopped early only for ethical reasons or if the power is no longer acceptable. However, it is recognised that drug development plans involve the need for sponsor access to comparative treatment data for a variety of reasons, such as planning other trials. It is also recognised that only a subset of trials will involve the study of serious life-threatening outcomes or mortality which may need sequential monitoring of accruing comparative treatment effects for ethical reasons. In either of these situations, plans for interim statistical analysis should be in place in the protocol or in protocol amendments prior to the unblinded access to comparative treatment data in order to deal with the potential statistical and operational bias that may be introduced.

For many clinical trials of investigational products, especially those that have major public health significance, the responsibility for monitoring comparisons of efficacy and/or safety outcomes should be assigned to an external independent group, often called an Independent Data Monitoring Committee (IDMC), a Data and Safety Monitoring Board or a Data Monitoring Committee whose responsibilities should be clearly described.

When a sponsor assumes the role of monitoring efficacy or safety comparisons and therefore has access to unblinded comparative information, particular care should be taken to protect the integrity of the trial and to manage and limit appropriately the sharing of information. The sponsor should assure and document that the internal monitoring committee has complied with written standard operating procedures and that minutes of decision making meetings including records of interim results are maintained.

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

4.6 Role of Independent Data Monitoring Committee (IDMC) (see Sections 1.25 and 5.52 of ICH E6)

An IDMC may be established by the sponsor to assess at intervals the progress of a clinical trial, safety data, and critical efficacy variables and recommend to the sponsor whether to continue, modify or terminate a trial. The IDMC should have written operating procedures and maintain records of all its meetings, including interim results; these should be available for review when the trial is complete. The independence of the IDMC is intended to control the sharing of important comparative information and to protect the integrity of the clinical trial from adverse impact resulting from access to trial information. The IDMC is a separate entity from an Institutional Review Board (IRB) or an Independent Ethics Committee (IEC), and its composition should include clinical trial scientists knowledgeable in the appropriate disciplines including statistics.

When there are sponsor representatives on the IDMC, their role should be clearly defined in the operating procedures of the committee (for example, covering whether or not they can vote on key issues). Since these sponsor staff would have access to unblinded information, the procedures should also address the control of dissemination of interim trial results within the sponsor organisation.

V. DATA ANALYSIS CONSIDERATIONS

5.1 Prespecification of the Analysis

When designing a clinical trial the principal features of the eventual statistical analysis of the data should be described in the statistical section of the protocol. This section should include all the principal features of the proposed confirmatory analysis of the primary variable(s) and the way in which anticipated analysis problems will be handled. In case of exploratory trials this section could describe more general principles and directions.

The statistical analysis plan (see Glossary) may be written as a separate document to be completed after finalising the protocol. In this document, a more technical and detailed elaboration of the principal features stated in the protocol may be included (see section 7.1). The plan may include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data. The plan should be reviewed and possibly updated as a result of the blind review of the data (see 7.1 for definition) and should be finalised before breaking the blind. Formal records should be kept of when the statistical analysis plan was finalised as well as when the blind was subsequently broken.

If the blind review suggests changes to the principal features stated in the protocol, these should be documented in a protocol amendment. Otherwise, it will suffice to update the statistical analysis plan with the considerations suggested from the blind review. Only results from analyses envisaged in the protocol (including amendments) can be regarded as confirmatory.

In the statistical section of the clinical study report the statistical methodology should be clearly described including when in the clinical trial process methodology decisions were made (see ICH E3).

5.2 Analysis Sets

The set of subjects whose data are to be included in the main analyses should be defined in the statistical section of the protocol. In addition, documentation for all subjects for whom trial procedures (e.g. run-in period) were initiated may be useful. The content of this subject 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 subjects 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 set of subjects 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. Hence, the statistical section of the protocol should address anticipated problems prospectively in terms of how these affect the subjects and data to be analysed. 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 also to handle the problems that do occur in the analysis of data. Possible amendments to the way in which the analysis will deal with protocol violations should be identified during the blind review. It is desirable to identify any important protocol violation with respect to the time when it occurred, its cause and influence on the trial result. The frequency and type of protocol violations, missing values, and other problems should be documented in the clinical study report and their potential influence on the trial results should be described (see ICH E3).

Decisions concerning the analysis set should be guided by the following principles : 1) to minimise bias, and 2) to avoid inflation of type I error.

5.2.1 Full Analysis Set

The intention-to-treat (see Glossary) 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 intention-to-treat ideal of including all randomised subjects. Preservation of the initial randomisation in analysis is important in preventing bias and in providing a secure foundation for statistical tests. In many clinical trials the use of the full analysis set provides a conservative strategy. Under many circumstances it may also provide estimates of treatment effects which are more likely to mirror those observed in subsequent practice.

There are a limited number of 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. Subjects who fail to satisfy an entry criterion may be excluded from the analysis without the possibility of introducing bias only under the following circumstances:

(i) the entry criterion was measured prior to randomisation;

(ii) the detection of the relevant eligibility violations can be made completely objectively;

(iii) all subjects receive equal scrutiny for eligibility violations; (This may be difficult to ensure in an open-label study, or even in a double-blind study if the data are unblinded prior to this scrutiny, emphasising the importance of the blind review.)

(iv) all detected violations of the particular entry criterion are excluded.

In some situations, it may be reasonable to eliminate from the set of all randomised subjects any subject who took no trial medication. The intention-to-treat 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. In other situations it may be necessary to eliminate from the set of all randomised subjects any subject 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 intention-to-treat principle. Special problems arise in connection with subjects withdrawn from treatment after receiving one or more doses who provide no data after this point, and subjects otherwise lost to follow-up, because failure to include these subjects in the full analysis set may seriously undermine the approach. 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, are valuable in this context; subsequent collection is especially important in studies where the primary variable is mortality or serious morbidity. The intention to collect data in this way should be described in the protocol. Imputation techniques, ranging from the carrying forward of the last observation to the use of complex mathematical models, may also be used in an attempt to compensate for missing data. Other methods employed to ensure the availability of measurements of primary variables for every subject in the full analysis set may require some assumptions about the subjects' outcomes or a simpler choice of outcome (e.g. success / failure). 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 of analysis especially when the strategy in question could itself lead to biased estimates of treatment effects.

Because of the unpredictability of some problems, 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 trial, and, if so, this should be stated in the protocol.

5.2.2 Per Protocol Set

sample or the 'evaluable subjects' sample, defines a subset of the subjects in the full analysis set who are more compliant with the protocol and is characterised by criteria such as the following:

(i) the completion of a certain pre-specified minimal exposure to the treatment regimen;

(ii) the availability of measurements of the primary variable(s);

(iii) the absence of any major protocol violations including the violation of entry criteria.

The precise reasons for excluding subjects from the per protocol set should be fully defined and documented before breaking the blind in a manner appropriate to the circumstances of the specific trial.

The use of the per protocol set may maximise the opportunity for a new treatment to show additional efficacy in the analysis, and most closely reflects the scientific model underlying the protocol. However, the corresponding test of the hypothesis and estimate of the treatment effect may or may not be conservative depending on the trial; the bias, which may be severe, arises from the fact that adherence to the study protocol may be related to treatment and outcome.

The problems that lead to the exclusion of subjects to create the per protocol set, and other protocol violations, should be fully identified and summarised. Relevant protocol violations may include errors in treatment assignment, the use of excluded medication, poor compliance, loss to follow-up and missing data. It is good practice to assess the pattern of such problems among the treatment groups with respect to frequency and time to occurrence.

5.2.3 Roles of the Different Analysis Sets

In general, it is advantageous to demonstrate a lack of sensitivity of the principal trial results to alternative choices of the set of subjects analysed. In confirmatory trials it is usually appropriate to plan to conduct both an analysis of the full analysis set and a per protocol analysis, so that any differences between them can be the subject of explicit discussion and interpretation. In some cases, it may be desirable to plan further exploration of the sensitivity of conclusions to the choice 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, bearing in mind, however, that the need to exclude a substantial proportion of subjects from the per protocol analysis throws some doubt on the overall validity of the trial.

The full analysis set and the per protocol set play different roles in superiority trials (which seek to show the investigational product to be superior), and in equivalence or non-inferiority trials (which seek to show the investigational product to be comparable, see section 3.3.2). In superiority trials the full analysis set is used in the primary analysis (apart from exceptional circumstances) because it tends to avoid over-optimistic estimates of efficacy resulting from a per protocol analysis, since the non-compliers included in the full analysis set will generally diminish the estimated treatment effect. However, in an equivalence or non-inferiority trial use of the full analysis set is generally not conservative and its role should be considered very carefully.

5.3 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 and management of data. In reality, however, there will almost always be some missing data. A trial may be regarded as valid, nonetheless, provided the methods of dealing with missing values are sensible, and particularly if those methods are pre-defined in the protocol. Definition of methods may be refined

by updating this aspect in the statistical 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.

5.4 Data Transformation

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 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. 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 derived variables, 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 derivation should be justified in the protocol. Closely related points are made in Section 2.2.2.

5.5 Estimation, Confidence Intervals and Hypothesis Testing

The statistical section of the protocol should specify the hypotheses that are to be tested and/or the treatment effects which are to be estimated in order to satisfy the primary 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. A description should be given of any intentions to use baseline data to improve precision or to adjust estimates for potential baseline differences, for example by means of analysis of covariance.

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 hypothesis tests are not considered appropriate, then the alternative process for arriving at statistical conclusions should be given. The issue of one-sided or two-sided approaches to inference is controversial and a diversity of views can be found in the statistical literature. The approach of setting type I errors for one-sided tests at half the conventional type I error used in two-sided tests is preferable in regulatory settings. This promotes consistency with the two-sided confidence intervals that are generally appropriate for estimating the possible size of the difference between two treatments.

The particular statistical model chosen should reflect the current state of medical and statistical knowledge about the variables to be analysed as well as the statistical design of the trial. 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 5.7.). In the choice of statistical methods due attention should be paid to the statistical distribution of both primary and secondary variables.

When making this choice (for example between parametric and non-parametric methods) it is important to bear in mind the need to provide statistical estimates of the size of treatment effects together with confidence intervals (in addition to significance tests).

The primary analysis of the primary variable should be clearly distinguished from supporting analyses of the primary or secondary variables. Within the statistical section of the protocol or the statistical 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 trials, for example for safety data.

Modelling approaches that incorporate information on known pharmacological parameters, the extent of protocol compliance for individual subjects or other biologically based data may provide valuable insights into actual or potential efficacy, especially with regard to estimation of treatment effects. The assumptions underlying such models should always be clearly identified, and the limitations of any conclusions should be carefully described.

5.6 Adjustment of Significance and Confidence Levels

When multiplicity is present, the usual frequentist approach to the analysis of clinical trial data may necessitate an adjustment to the type I error. Multiplicity may arise, for example, from multiple primary variables (see Section 2.2.2), multiple comparisons of treatments, repeated evaluation over time and/or interim analyses (see Section 4.5). Methods to avoid or reduce multiplicity are sometimes preferable when available, such as the identification of the key primary variable (multiple variables), the choice of a critical treatment contrast (multiple comparisons), the use of a summary measure such as 'area under the curve' (repeated measures). In confirmatory analyses, any aspects of multiplicity which remain after steps of this kind have been taken should be identified in the protocol; adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan.

5.7 Subgroups, Interactions and Covariates

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 subjects 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 planned analysis and hence should be set out in the protocol. Pre-trial 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. If one or more factors are used to stratify the design, it is appropriate to account for those factors in the analysis. 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 because they may be affected by the treatments.

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 subjects. In some cases such interactions are anticipated or are of particular prior interest (e.g. geriatrics), and hence a subgroup analysis, or a statistical model including interactions, is part of the planned confirmatory analysis. 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 first through the addition of interaction terms to the statistical model in question, complemented by additional exploratory analysis within relevant subgroups of subjects, or within strata defined by the covariates. When exploratory, these analyses should be interpreted cautiously; any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses are unlikely to be accepted.

5.8 Integrity of Data and Computer Software Validity

The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software (both internally and externally written) 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 based on thorough and effective standard operating procedures. The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

VI. EVALUATION OF SAFETY AND TOLERABILITY

6.1 Scope of Evaluation

In all clinical trials evaluation of safety and tolerability (see Glossary) constitutes an important element. In early phases this evaluation is mostly of an exploratory nature, and is only sensitive to frank expressions of toxicity, whereas in later phases the establishment of the safety and tolerability profile of a drug can be characterised more fully in larger samples of subjects. Later phase controlled trials represent an important means of exploring in an unbiased manner any new potential adverse effects, even if such trials generally lack power in this respect.

Certain trials may be designed with the purpose of making specific claims about superiority or equivalence with regard to safety and tolerability compared to another drug or to another dose of the investigational drug. Such specific claims should be supported by relevant evidence from confirmatory trials, similar to that necessary for corresponding efficacy claims.

6.2 Choice of Variables and Data Collection

In any clinical trial the methods and measurements chosen to evaluate the safety and tolerability of a drug will depend on a number of factors, including knowledge of the adverse effects of closely related drugs, information from non-clinical and earlier clinical trials and possible consequences of the pharmacodynamic/pharmacokinetic properties of the particular drug, the mode of administration, the type of subjects to be studied, and the duration of the trial. Laboratory tests

concerning clinical chemistry and haematology, vital signs, and clinical adverse events (diseases, signs and symptoms) usually form the main body of the safety and tolerability data. The occurrence of serious adverse events and treatment discontinuations due to adverse events are particularly important to register (see ICH E2A and ICH E3).

Furthermore, it is recommended that a consistent methodology be used for the data collection and evaluation throughout a clinical trial program in order to facilitate the combining of data from different trials. The use of a common adverse event dictionary is particularly important. This dictionary has a structure which gives the possibility to summarise the adverse event data on three different levels; system-organ class, preferred term or included term (see Glossary). The preferred term is the level on which adverse events usually are summarised, and preferred terms belonging to the same system-organ class could then be brought together in the descriptive presentation of data (see ICH M1).

6.3 Set of Subjects to be Evaluated and Presentation of Data

For the overall safety and tolerability assessment, the set of subjects to be summarised is usually defined as those subjects who received at least one dose of the investigational drug. Safety and tolerability variables should be collected as comprehensively as possible from these subjects, including type of adverse event, severity, onset and duration (see ICH E2B). Additional safety and tolerability evaluations may be needed in specific subpopulations, such as females, the elderly (see ICH E7), the severely ill, or those who have a common concomitant treatment. These evaluations may need to address more specific issues (see ICH E3).

All safety and tolerability variables will need attention during evaluation, and the broad approach should be indicated in the protocol. All adverse events should be reported, whether or not they are considered to be related to treatment. All available data in the study population should be accounted for in the evaluation. 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 evaluation. Use of a toxicity grading scale should be prespecified and justified.

The incidence of a certain adverse event is usually expressed in the form of a proportion relating number of subjects experiencing events to number of subjects at risk. However, it is not always self-evident how to assess incidence. For example, depending on the situation the number of exposed subjects or the extent of exposure (in person-years) could be considered for the denominator. Whether the purpose of the calculation is to estimate a risk or to make a comparison between treatment groups it is important that the definition is given in the protocol. This is especially important if long-term treatment is planned and a substantial proportion of treatment withdrawals or deaths are expected. For such situations survival analysis methods should be considered and cumulative adverse event rates calculated in order to avoid the risk of underestimation.

In situations when there is a substantial background noise of signs and symptoms (e.g. in psychiatric trials) one should consider ways of accounting for this in the estimation of risk for different adverse events. One such method is to make use of the 'treatment emergent' (see

Glossary) concept in which adverse events are recorded only if they emerge or worsen relative to pretreatment baseline.

Other methods to reduce the effect of the background noise may also be appropriate such as ignoring adverse events of mild severity or requiring that an event should have been observed at repeated visits to qualify for inclusion in the numerator. Such methods should be explained and justified in the protocol.

6.4 Statistical Evaluation

The investigation of safety and tolerability is a multidimensional problem. Although some specific adverse effects can usually be anticipated and specifically monitored for any drug, the range of possible adverse effects is very large, and new and unforeseeable effects are always possible. Further, an adverse event experienced after a protocol violation, such as use of an excluded medication, may introduce a bias. This background underlies the statistical difficulties associated with the analytical evaluation of safety and tolerability of drugs, and means that conclusive information from confirmatory clinical trials is the exception rather than the rule.

In most trials the safety and tolerability implications are best addressed by applying descriptive statistical methods to the data, supplemented by calculation of confidence intervals wherever this aids interpretation. It is also valuable to make use of graphical presentations in which patterns of adverse events are displayed both within treatment groups and within subjects.

The calculation of p-values is 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 and tolerability variables to highlight differences worth further attention. This is particularly useful for laboratory data, which otherwise can be difficult to summarise appropriately. It is recommended that laboratory data be subjected to both a quantitative analysis, e.g. evaluation of treatment means, and a qualitative analysis where counting of numbers above or below certain thresholds are calculated.

If hypothesis tests are used, statistical adjustments for multiplicity to quantify the type I error are appropriate, but the type II error is usually of more concern. Care should be taken when interpreting putative statistically significant findings when there is no multiplicity adjustment.

In the majority of trials investigators are seeking to establish that there are no clinically unacceptable differences in safety and tolerability compared with either a comparator drug or a placebo. As is the case for non-inferiority or equivalence evaluation of efficacy the use of confidence intervals is preferred to hypothesis testing in this situation. In this way, the considerable imprecision often arising from low frequencies of occurrence is clearly demonstrated.

6.5 Integrated Summary

The safety and tolerability properties of a drug are commonly summarised across trials continuously during an investigational product's development and in particular at the time of a marketing application. The usefulness of this summary, however, is dependent on adequate and well-controlled individual trials with high data quality.

The overall usefulness of a drug is always a question of balance between risk and benefit and in a single trial such a perspective could also be considered, even if the assessment of risk/benefit usually is performed in the summary of the entire clinical trial program. (See section 7.2.2)

For more details on the reporting of safety and tolerability, see Chapter 12 of ICH E3.

VII. REPORTING

7.1 Evaluation and Reporting

As stated in the Introduction, the structure and content of clinical study reports is the subject of ICH E3. That ICH 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 principal features of the analysis should have been specified in the protocol as described in Section 5. 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 5. This pre-analysis review, blinded to treatment, should cover decisions concerning, for example, the exclusion of subjects or data from the analysis sets; 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 the report, and should be distinguished from those made after the statistician has had access to the treatment codes, as blind decisions will generally introduce less potential for bias. Statisticians or other staff involved in unblinded interim analysis should not participate in the blind review or in making modifications to the statistical analysis plan. When the blinding is compromised by the possibility that treatment induced effects may be apparent in the data, special care will be needed for the blind review.

Many of the more detailed aspects of presentation and tabulation should 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 subject 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 as described in the protocol, protocol amendments or the updated statistical analysis plan based on a blind review of data. A careful explanation should be provided for deviations from the planned analysis.

All subjects who entered the trial should be accounted for in the report, whether or not they are included in the analysis. All reasons for exclusion from analysis should be documented; for any subject included in the full analysis set but not in the per protocol set, the reasons for exclusion from the latter should also be documented. Similarly, for all subjects included in an analysis set, the measurements of all important variables should be accounted for at all relevant time-points.

The effect of all losses of subjects or data, withdrawals from treatment and major protocol violations on the main analyses of the primary variable(s) should be considered carefully. Subjects lost to follow up, withdrawn from treatment, or with a severe protocol violation 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. When reporting the results of significance tests, precise p-values (e.g. 'p=0.034') should be reported rather than making exclusive reference to critical values.

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 planned analysis but having some prognostic importance nevertheless. This is best dealt with by showing that an additional 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 discussed.

In general, sparing use should be made of unplanned analyses. Such 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 subjects for whom the effect is particularly beneficial. The potential dangers of over-interpretation of unplanned subgroup analyses are well known (see also Section 5.7), 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 subjects, such possibilities should be properly assessed and should therefore be reported.

Finally 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 clinical study report, and should approve the clinical report.

7.2 Summarising the Clinical Database

An overall summary and synthesis of the evidence on safety and efficacy from all the reported clinical trials is required for a marketing application (Expert report in EU, integrated summary reports in USA, Gaiyo in Japan). This may be accompanied, when appropriate, by a statistical combination of results.

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; summarising the safety information available from the combined database of all the trials whose results contribute to the marketing 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 valid reasons not to do so.

Any statistical procedures used to combine data across trials should 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.

7.2.1 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 trials should 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. When used for this purpose the meta-analysis should have its own prospectively written protocol.

7.2.2 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 drug provides an important 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 because of the lack of a comparator group, and data from comparative trials are especially valuable in overcoming this difficulty. The results from trials 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 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 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.

GLOSSARY

Glossary	Content
Bayesian Approaches	Approaches to data analysis that provide a posterior probability distribution for some parameter (e.g. treatment effect), derived from the observed data and a prior probability distribution for the parameter. The posterior distribution is then used as the basis for statistical inference.
Bias (Statistical & Operational)	The systematic tendency of any factors associated with the design, conduct, analysis and evaluation of the results of a clinical trial to make the estimate of a treatment effect deviate from its true value. Bias introduced through deviations in conduct is referred to as 'operational' bias. The other sources of bias listed above are referred to as 'statistical'.
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.
Content Validity	The extent to which a variable (e.g. a rating scale) measures what it is supposed to measure.
Double-Dummy	A technique for retaining the blind when administering supplies in a clinical trial, when the two treatments cannot be made identical. Supplies are prepared for Treatment A (active and indistinguishable placebo) and for Treatment B (active and indistinguishable placebo). Subjects then take two sets of treatment; either A (active) and B (placebo), or A (placebo) and B (active).
Dropout	A subject in a clinical trial who for any reason fails to continue in the trial until the last visit required of him/her by 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.
Frequentist Methods	Statistical methods, such as significance tests 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.

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 subjects who participated in the trial to a broader patient 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 a subject.
Independent Data Monitoring Committee (IDMC) (Data and Safety Monitoring Board, Monitoring Committee, Data Monitoring Committee)	An independent data-monitoring committee that may be established by the sponsor to assess at intervals the progress of a clinical trial, the safety data, and the critical efficacy endpoints, and to recommend to the sponsor whether to continue, modify, or stop a trial.
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.
Interaction (Qualitative & Quantitative)	The situation in which a treatment contrast (e.g. difference between investigational product and control) is dependent on another factor (e.g. centre). 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-Rater Reliability	The property of yielding equivalent results when used by different raters on different occasions.
Intra-Rater Reliability	The property of yielding equivalent results when used by the same rater 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 a trial.
Meta-Analysis	The formal evaluation of the quantitative evidence from two or more trials bearing on the same question. This most commonly involves the statistical combination of summary statistics from the various trials, but the term is sometimes also used to refer to the combination of the raw data.
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.

Non-Inferiority Trial	A trial with the primary objective of showing that the response to the investigational product is not clinically inferior to a comparative agent (active or placebo control).
Preferred and Included Terms	In a hierarchical medical dictionary, for example MedDRA, the included term is the lowest level of dictionary term to which the investigator description is coded. The preferred term is the level of grouping of included terms typically used in reporting frequency of occurrence. For example, the investigator text "Pain in the left arm" might be coded to the included term "Joint pain", which is reported at the preferred term level as "Arthralgia".
Per Protocol Set (Valid Cases, Efficacy Sample, Evaluable Subjects Sample)	The set of data generated by the subset of subjects who 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.
Safety & Tolerability	The safety of a medical product concerns the medical risk to the subject, usually assessed in a clinical trial by laboratory tests (including clinical chemistry and haematology), vital signs, clinical adverse events (diseases, signs and symptoms), and other special safety tests (e.g. ECGs, ophthalmology). The tolerability of the medical product represents the degree to which overt adverse effects can be tolerated by the subject.
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.
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).
Surrogate Variable	A variable that provides an indirect measurement of effect in situations where direct measurement of clinical effect is not feasible or practical.
Treatment Effect	An effect attributed to a treatment in a clinical trial. In most clinical trials the treatment effect of interest is a comparison (or contrast) of two or more treatments.
Treatment Emergent	An event that emerges during treatment having been absent pre-treatment, or worsens relative to the pre-treatment state.
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.

ICH E9 ??????????

English Version

ICH E9 Statistical Principles for Clinical Trials

1. ??

1.1 ?????

ICH 1996 5 1

GCP ICH E6 ICH E6

Instrument	1994	1992	1998
CPMP	12	3	7
ICH	28	28	10

E1A:	
E2A:	
E2B:	
E2C:	
E3:	
E4:	
E5:	
E6:	
E7:	
E8:	
E10:	
M1:	
M3:	

1.2 ?????

ICH

/

ICH E6

“ ”

P

/

2. ??????????

2.2.7 ????

“ ” “ ”

“ ”

90mmHg

2.3 ?????????

2.3.1 2.3.2

2.3.1 ??

/

/ “ ”

3. ???????

3.1 ????

3.1.1 ?????

3.1.2 ????

ICH E4

2x2

/

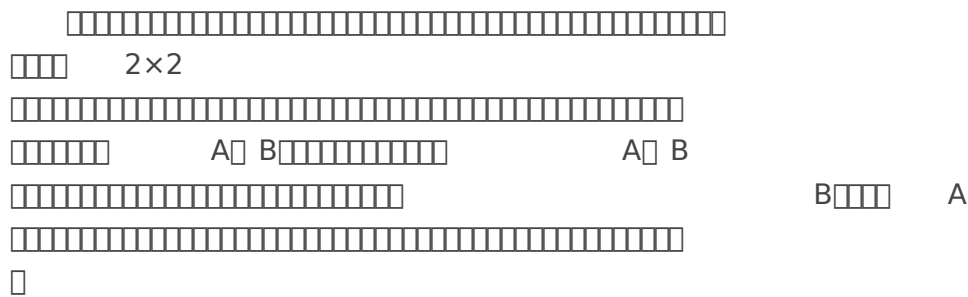
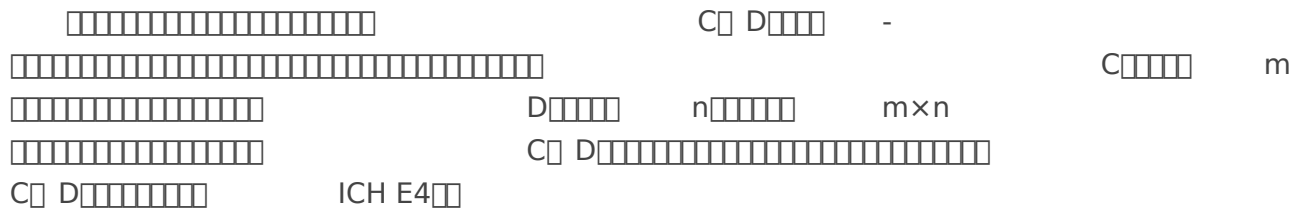
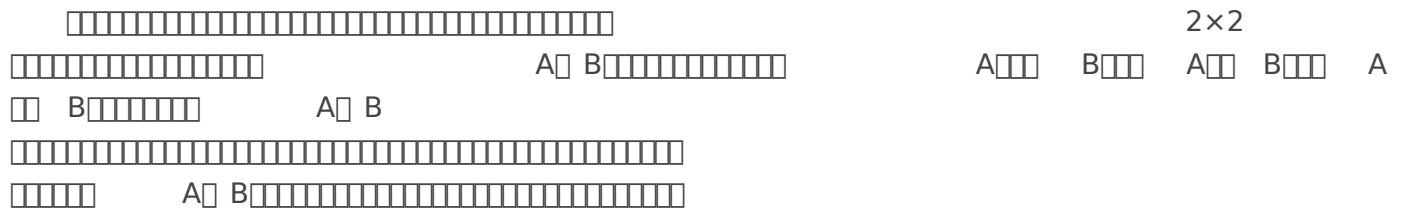
2x2

$n \leq 2$
 $n > 2$
n

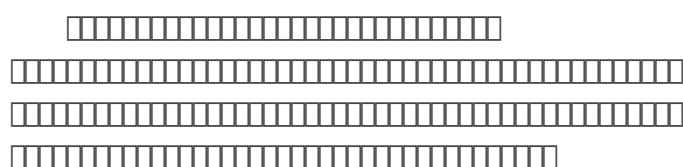
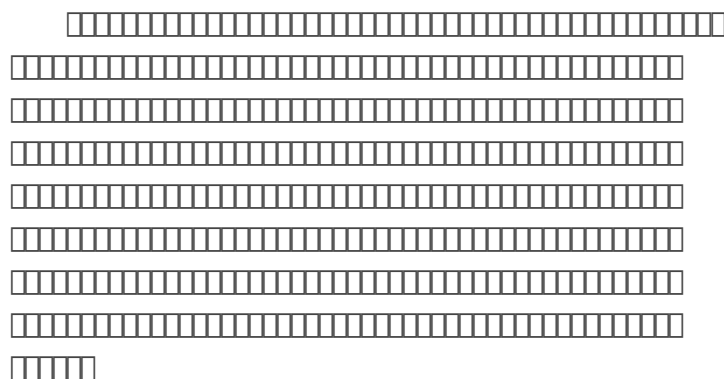
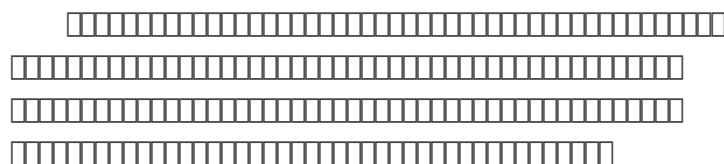
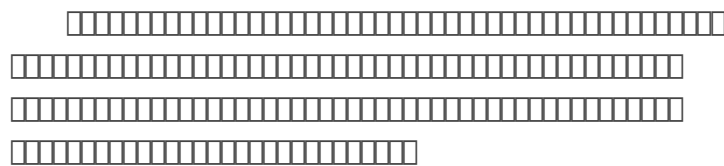
2x2

“ ”

3.1.3 ????



3.2 ?????



“ ”

Diagram illustrating a subtraction problem using boxes:

- Row 1: 20 boxes (representing 20)
- Row 2: 18 boxes (representing 18)
- Row 3: 2 boxes, followed by a minus sign, and 10 boxes (representing $2 - 10$)

A diagram consisting of 10 rows of small squares. Each row contains 10 squares, for a total of 100 squares. The first row is shaded gray, representing 10% of the total.

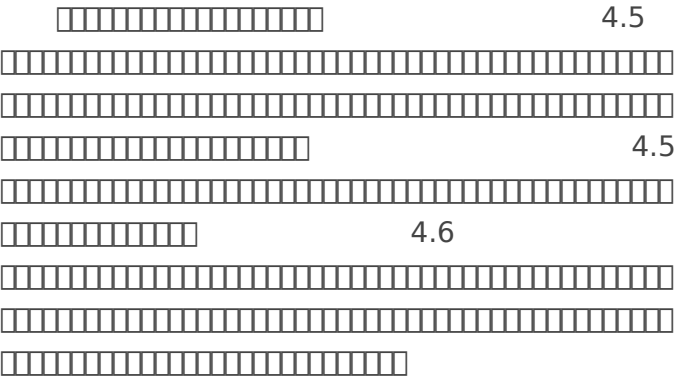
Grade	Number of Students (approximate)
1st	1.5
2nd	2.5
3rd	3.5
4th	4.5
5th	5.5
6th	4.5
7th	3.5
8th	3.5

5.2.3

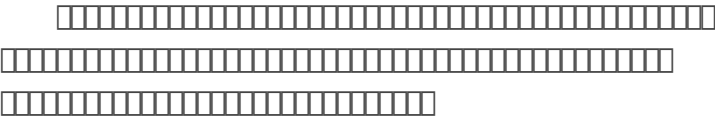
3.3.3 ??-???????

	ICH
E4	-

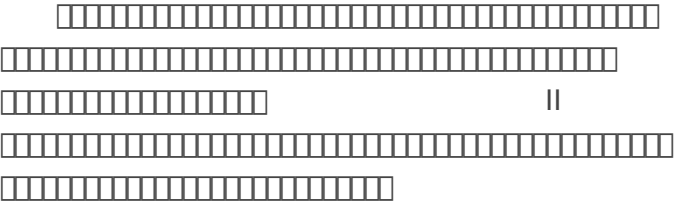
3.4 ??????



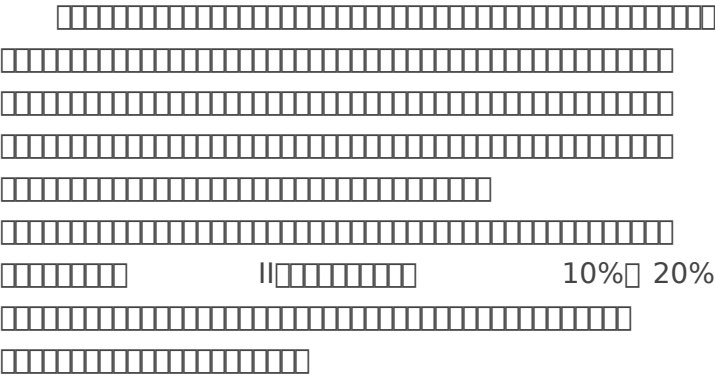
3.5 ???



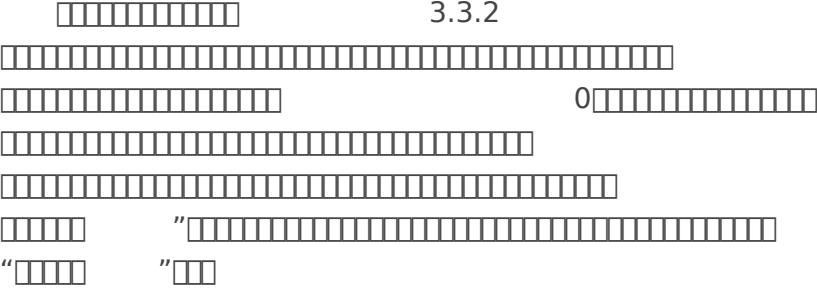
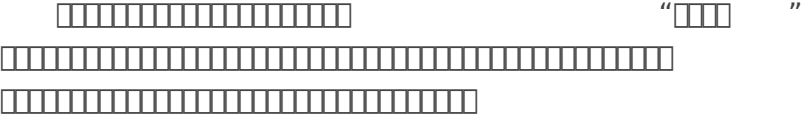
ICH E1a



“ ”
I



I 5%
I
I II



0
0
“



4.4

3.6

“ ” “0 ” “ ”

GCP ICH E6 5

4.

4.1

ICH E6

4.2 4.4
I
ICH E6

4.5 4.6

4.2

/

4.3

4.4

/

I

3.5

4.5

3.4

4.6

I

α

/

ICH E3

5.2.1 ????

A diagram consisting of 10 rows of small squares. Each row contains 10 squares, for a total of 100 squares. The first row is shaded gray, representing 10% of the total.

“ ”

The diagram shows three horizontal bars, each composed of small squares. The top bar is the longest, followed by the middle bar, and the bottom bar is the shortest. These bars represent the lengths of the three longest DNA fragments.

☐ 1

☐ 2 

□ 3

☐ 4

[illegible]

5.4

2.2.2

5.5

5.6

/

ICH E2AICH E3

-

- ICH M1

6.3 ????????????

ICH E2B

ICH E3

ICH E7

“”

6.4 ????

[illegible]

7.2.2 ?????

--	--	--	--	--	--

[illegible]

/

/					
---	--	--	--	--	--

???

[illegible]

[illegible]

□□□□	□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□
□□□□□□	□□□□□□□□□□□□□□□□□□□□ □□□□
□□□□□	□□□□□□□□□□ / □□□□□□□□□□□□□□□□□□□□ □□□□

ICH E9(R1) Addendum: Statistical Principles for Clinical Trials

□□

[ICH E9\(R1\)](#) □□□□□□□□□□

[E9](#) □□□□□□□□

A.1. PURPOSE AND SCOPE

To properly inform decision making by pharmaceutical companies, regulators, patients, physicians and other stakeholders, clear descriptions of the benefits and risks of a treatment (medicine) for a given medical condition should be made available. Without such clarity, there is a concern that the reported “treatment effect” will be misunderstood. This addendum presents a structured framework to strengthen the dialogue between disciplines involved in the formulation of clinical trial objectives, design, conduct, analysis and interpretation, as well as between sponsor and regulator regarding the treatment effect(s) of interest that a clinical trial should address.

Precision in describing a treatment effect of interest is facilitated by constructing the “estimand” (see Glossary; A.3.) corresponding to a clinical question of interest. Clarity requires a thoughtful envisioning of “intercurrent events” (see Glossary; A.3.1.) such as discontinuation of assigned treatment, use of an additional or alternative treatment and terminal events such as death. The description of an estimand should reflect the clinical question of interest in respect of these intercurrent events, and this addendum introduces strategies to reflect different questions of interest that might be posed. The choice of strategies can influence how more conventional attributes of a trial are reflected when describing the clinical question, for example the treatments, population or the variable (endpoint) of interest.

The statistical analysis of clinical trial data should be aligned to the estimand. This addendum clarifies the role of “sensitivity analysis” (see Glossary) to explore robustness of conclusions from the main statistical analysis.

Throughout the addendum, references to the original ICH E9 are made using x.y. References within this addendum are made using A.x.y.

This addendum clarifies and extends ICH E9 in respect of the following topics. Firstly, ICH E9 introduced the Intention-To-Treat (ITT) principle in connection with the effect of a treatment policy in a randomised controlled trial, whereby subjects are followed, assessed and analysed irrespective of their compliance to the planned course of treatment, indicating that preservation of

randomisation provides a secure foundation for statistical tests. Multiple consequences arising from the ITT principle can be distinguished. Firstly, that the trial analysis should include all subjects relevant for the research question. Secondly, that subjects should be included in the analysis as randomised. Taken directly from the definition of the ITT principle (see ICH E9 Glossary), a third consequence is that subjects should be followed-up and assessed regardless of adherence to the planned course of treatment and that those assessments should be used in the analysis. It remains undisputed that randomisation is a cornerstone of controlled clinical trials and that analysis should aim at exploiting the advantages of randomisation to the greatest extent possible. However, the question remains whether estimating an effect in accordance with the ITT principle always represents the treatment effect of greatest relevance to regulatory and clinical decision making. The framework outlined in this addendum gives a basis for describing different treatment effects and some points to consider for the design and analysis of trials to give estimates of these treatment effects that are reliable for decision making.

Secondly, issues considered generally under data handling and “missing data” (see Glossary) are re-visited. Two important distinctions are made. Firstly, the addendum distinguishes discontinuation of randomised treatment from study withdrawal. The former represents an intercurrent event, to be addressed in the precise specification of the trial objective through the estimand. The latter gives rise to missing data to be addressed in the statistical analysis. Consider, for example, a subject switching treatments in an oncology trial, and a subject for whom no outcome event can be observed because the trial is completed. The former represents an intercurrent event and the clinical question of interest in respect of that should be clear. The latter is administrative censoring which needs to be addressed as a missing data problem in the statistical analysis. Having clarity in the estimand gives a basis for planning which data need to be collected and hence which data, when not collected, present a missing data problem to be addressed in the statistical analysis. In turn, methods to address the problem presented by missing data can be selected to align with the estimand. Secondly, the addendum highlights the distinct consequences of different intercurrent events. Events such as discontinuation of treatment, switching between treatments, or use of an additional medication may render the later measurements of the variable irrelevant or difficult to interpret even when they can be collected. Measurements after a subject dies do not exist.

Thirdly, issues related to the concept of analysis sets are considered in the framework. Section 5.2. strongly recommends that analysis of superiority trials be based on the full analysis set, defined to be as close as possible to including all randomised subjects. However, trials often include repeated measurements on the same subject. Elimination of some planned measurements on some subjects, perhaps because the measurement is considered irrelevant or difficult to interpret, can have similar consequences to excluding subjects altogether from the full analysis set, i.e. that the initial randomisation is not fully preserved. A consequence of this is that the theoretical benefits that randomisation confers on testing hypotheses about treatment effects and the practical benefits of balancing confounding factors at baseline can be diminished. In addition, a meaningful value of the outcome variable might not exist, as when the subject dies. Section 5.2. does not directly address these issues. Clarity is introduced by carefully defining the treatment effect of interest in a way that determines both the population of subjects to be included in the estimation of that treatment effect and the observations from each subject to be included in the analysis considering the occurrence of intercurrent events. The meaning and role of an analysis of the per protocol set is also re-visited in this addendum; in particular whether the need to explore the impact of protocol

violations and deviations can be addressed in a way that is less biased and more interpretable than naïve analysis of the per protocol set.

Finally, the concept of robustness (see 1.2.) is given expanded discussion under the heading of sensitivity analysis. A distinction is made between the sensitivity of inference to the assumptions of a chosen method of analysis and the sensitivity to the choice of analytic approach more broadly. With precise specification of an agreed estimand and a method of analysis that is both aligned to the estimand and pre-specified to a level of detail that it can be replicated precisely by a third party, regulatory interest can focus on sensitivity to deviations from assumptions and limitations in the data in respect of a particular analysis.

The principles outlined in this addendum are relevant whenever a treatment effect is estimated, or a hypothesis related to a treatment effect is tested, whether related to efficacy or safety. While the main focus is on randomised clinical trials, the principles are also applicable for single arm trials and observational studies. The framework applies to any data type, including longitudinal, time-to-first event, and recurrent event data. Regulatory interest in the application of the principles outlined will be greater for confirmatory clinical trials and, where used to generate confirmatory conclusions, for data integrated across trials.

A.2. A FRAMEWORK TO ALIGN PLANNING, DESIGN, CONDUCT, ANALYSIS AND INTERPRETATION

Trial planning should proceed in sequence (Figure 1). Clear trial objectives should be translated into key clinical questions of interest by defining suitable estimands. An estimand defines the target of estimation for a particular trial objective (i.e. “what is to be estimated”, see A.3.). A suitable method of estimation (i.e. the analytic approach, referred to as the main “estimator”, see Glossary) can then be selected (see A.5.1.). The main estimator will be underpinned by certain assumptions. To explore the robustness of inferences from the main estimator to deviations from its underlying assumptions, a sensitivity analysis should be conducted, in the form of one or more analyses, targeting the same estimand (see A.5.2.).

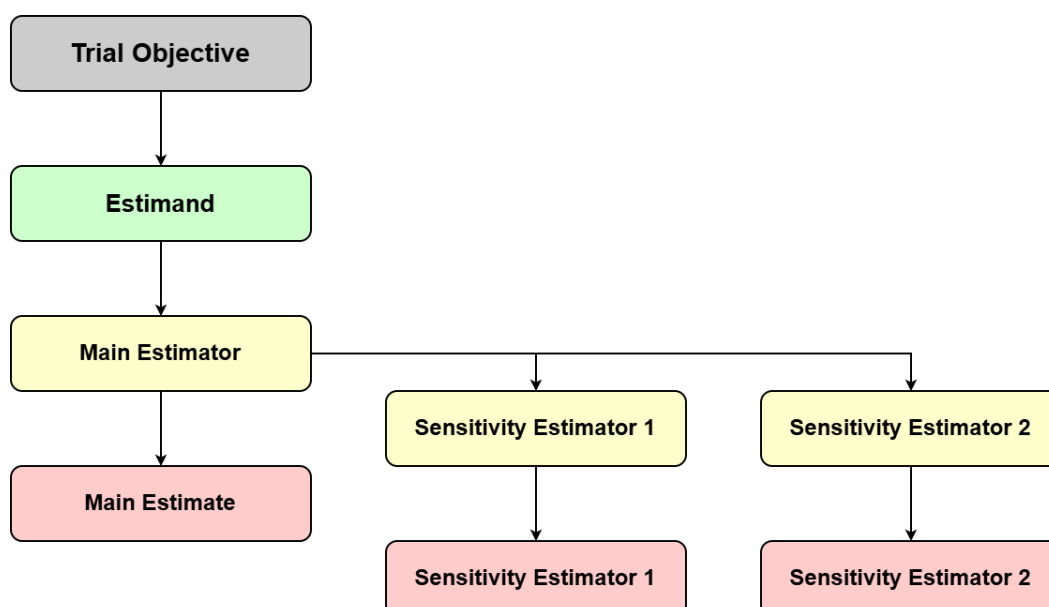


Figure 1: Aligning target of estimation, method of estimation, and sensitivity analysis, for a given trial objective

This framework enables proper trial planning that clearly distinguishes between the target of estimation (trial objective, estimand), the method of estimation (estimator), the numerical result (“estimate”, see Glossary), and a sensitivity analysis. This will assist sponsors in planning trials, regulators in their reviews, and will enhance the interactions between these parties when discussing the suitability of clinical trial designs, and the interpretation of clinical trial results.

The specification of appropriate estimands (see A.3.) will usually be the main determinant for aspects of trial design, conduct (see A.4.) and analysis (see A.5.).

A.3. ESTIMANDS

Central questions for drug development and licensing are to establish the existence, and to estimate the magnitude, of treatment effects: how the outcome of treatment compares to what would have happened to the same subjects under alternative treatment (i.e. had they not received the treatment, or had they received a different treatment). An estimand is a precise description of the treatment effect reflecting the clinical question posed by a given clinical trial objective. It summarises at a population level what the outcomes would be in the same patients under different treatment conditions being compared. The targets of estimation are to be defined in advance of a clinical trial. Once defined, a trial can be designed to enable reliable estimation of the targeted treatment effect.

The description of an estimand involves precise specifications of certain attributes, which should be developed based not only on clinical considerations but also on how intercurrent events are reflected in the clinical question of interest. Section A.3.1. introduces intercurrent events. Section A.3.2. introduces strategies to describe the question of interest in respect of intercurrent events. Section A.3.3. describes the attributes of an estimand and Section A.3.4. gives considerations for

its construction. It is critically important to understand the differences between the strategies and to precisely articulate which are used in constructing the estimand.

A.3.1. Intercurrent Events to be Reflected in the Clinical Question of Interest

Intercurrent events are events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. It is necessary to address intercurrent events when describing the clinical question of interest in order to precisely define the treatment effect that is to be estimated.

Intercurrent events need to be considered in the description of a treatment effect because measurements of the variable can be influenced by the intercurrent event and the occurrence of the intercurrent event may depend on treatment. For example, two patients might be exposed initially to the same treatment and provide the same measure of outcome, but if one patient has received additional medication, the information that the two measures give about the treatment differs between the two patients. Furthermore, whether a patient needs to take additional medication, and whether or not a patient can continue taking treatment, may depend on the treatment to which they are exposed. Unlike missing data, intercurrent events are not to be thought of as a drawback to be avoided in clinical trials. Discontinuation of prescribed treatment, use of additional medication, and other such events may occur in clinical practice as they do in clinical trials, and their occurrence needs to be considered explicitly when defining the clinical question of interest.

Examples of intercurrent events that can affect interpretation of the measurements include discontinuation of assigned treatment and use of an additional or alternative therapy. Use of an additional or alternative therapy can take multiple forms, including change to background or concomitant therapy and switching between treatments of interest. Examples of intercurrent events that would affect the existence of the measurements include terminal events such as death and leg amputation (when assessing symptoms of diabetic foot ulcers), when these events are not part of the variable itself. Certain clinical events can also be intercurrent events, when their occurrence, or non-occurrence, defines a principal stratum of interest (see A.3.2.). Examples include tumour shrinkage defining objective response when assessing a treatment effect on duration of response in oncology and occurrence of infection when assessing a treatment effect on severity of infections occurring after vaccination of initially uninfected subjects.

An intercurrent event might be identified solely by the event itself, such as discontinuation of treatment, or might be more granular. For example, the reason for the event might be specified, such as discontinuation of treatment due to toxicity, or due to lack of efficacy; the event might require to be of certain magnitude or degree, such as use of additional medication exceeding a specified duration or dose; or the timing of the event might be specified, perhaps in relation to its proximity to the assessment of the variable. Some events will affect interpretation of the outcome measurements indefinitely, such as discontinuation of treatment, whilst others will affect interpretation only temporarily, such as short-term use of additional treatment. Indeed, additional or alternative treatments can be diverse; either replacing or supplementing a treatment on which the subject is experiencing inadequate benefit, as an alternative where a subject is not tolerating their assigned treatment, or as a short-term acute treatment to manage a temporary flare in

disease symptoms. In a clinical trial, additional or alternative treatments are often identified as e.g. background treatment, rescue medication, prohibited medication, distinguishing their different roles and allowing them to be considered separately. The additional granularity, identifying different intercurrent events, is required if different strategies are to be used. If the intercurrent event for which a strategy needs to be selected depends not only on, for example, failure to continue with treatment, but also on the reason, magnitude or timing associated with that failure, this additional information should be defined and recorded accurately in the clinical trial. The description of intercurrent events might in theory reflect very specific details of treatment and follow-up, such as a single missed dose of a chronic treatment or a dose taken at the wrong time of day. Where such specific criteria are not expected to affect interpretation of the variable, they would not need to be addressed as intercurrent events.

As indicated above, consideration of intercurrent events is required when constructing the estimand. Because the estimand is to be defined in advance of trial design, neither study withdrawal nor other reasons for missing data (e.g. administrative censoring in trials with survival outcomes) are in themselves intercurrent events. Subjects who withdraw from the trial may have experienced an intercurrent event before withdrawal.

A.3.2. Strategies for Addressing Intercurrent Events when Defining the Clinical Question of Interest

Descriptions of various strategies are listed below, each reflecting a different clinical question of interest in respect of a particular intercurrent event. Whether or not the naming convention is used, it is required that the choices of strategy are unambiguously clear once the estimand is constructed. It is not necessary to use the same strategy to address all intercurrent events. Indeed, different strategies will often be used to reflect the clinical question of interest in respect of different intercurrent events. Section A.3.4. gives some considerations on selecting strategies to construct an estimand.

Treatment policy strategy

The occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest: the value for the variable of interest is used regardless of whether or not the intercurrent event occurs. For example, when specifying how to address use of additional medication as an intercurrent event, the values of the variable of interest are used whether or not the patient takes additional medication.

If applied in relation to whether or not a patient continues treatment, and whether or not a patient experiences changes in other treatments (e.g. background or concomitant treatments), the intercurrent event is considered to be part of the treatments being compared. In that case, this reflects the comparison described in the ICH E9 Glossary (under ITT Principle) as the effect of a treatment policy.

In general, the treatment policy strategy cannot be implemented for intercurrent events that are terminal events, since values for the variable after the intercurrent event do not exist. For example, an estimand based on this strategy cannot be constructed with respect to a variable that cannot be measured due to death.

Hypothetical strategies

A scenario is envisaged in which the intercurrent event would not occur: the value of the variable to reflect the clinical question of interest is the value which the variable would have taken in the hypothetical scenario defined.

A wide variety of hypothetical scenarios can be envisaged, but some scenarios are likely to be of more clinical or regulatory interest than others. For example, it may be of clinical or regulatory importance to consider the effect of a treatment under different conditions from those of the trial that can be carried out. Specifically, when additional medication must be made available for ethical reasons, a treatment effect of interest might concern the outcomes if the additional medication was not available. A very different hypothetical scenario might postulate that intercurrent events would not occur, or that different intercurrent events would occur. For example, for a subject that will suffer an adverse event and discontinue treatment, it might be considered whether the same subject would not have the adverse event or could continue treatment in spite of the adverse event. The clinical and regulatory interest of such hypotheticals is limited and would usually depend on a clear understanding of why and how the intercurrent event or its consequences would be expected to be different in clinical practice than in the clinical trial.

If a hypothetical strategy is proposed, it should be made clear what hypothetical scenario is envisaged. For example, wording such as “if the patient does not take additional medication” might lead to confusion as to whether the patient hypothetically does not take additional medication because it is not available or because the particular patient is supposed not to require it.

Composite variable strategies

This relates to the variable of interest (see A.3.3.). An intercurrent event is considered in itself to be informative about the patient’s outcome and is therefore incorporated into the definition of the variable. For example, a patient who discontinues treatment because of toxicity may be considered not to have been successfully treated. If the outcome variable was already success or failure, discontinuation of treatment for toxicity would simply be considered another mode of failure. Composite variable strategies do not need to be limited to dichotomous outcomes, however. For example, in a trial measuring physical functioning, a variable might be constructed using outcomes on a continuous scale, with subjects who die being attributed a value reflecting the lack of ability to function. Composite variable strategies can be viewed as implementing the intention-to-treat principle in some cases where the original measurement of the variable might not exist or might not be meaningful, but where the intercurrent event itself meaningfully describes the patient’s outcome, such as when the patient dies.

Terminal events, such as death, are perhaps the most salient examples of the need for the composite strategy. If a treatment saves lives, its effect on various measures in surviving patients may be of interest, but it would be inappropriate to say that the summary measure of interest was only the average value of some numerical measure in survivors. The outcome of interest is survival along with the numerical measures. For example, progression-free survival in oncology trials measures the treatment effect on a combination of the growth of the tumour and survival.

While on treatment strategies

For this strategy, response to treatment prior to the occurrence of the intercurrent event is of interest. Terminology for this strategy will depend on the intercurrent event of interest; e.g. “while alive”, when considering death as an intercurrent event.

If a variable is measured repeatedly, its values up to the time of the intercurrent event may be considered relevant for the clinical question, rather than the value at the same fixed timepoint for all subjects. The same applies to the occurrence of a binary outcome of interest up to the time of the intercurrent event. For example, subjects with a terminal illness may discontinue a purely symptomatic treatment because they die, yet the success of the treatment can be measured based on the effect on symptoms before death. Alternatively, subjects might discontinue treatment and, in some circumstances, it will be of interest to assess the risk of an adverse drug reaction while the patient is exposed to treatment.

Like the composite variable strategy, the while on treatment strategy can hence be thought of as impacting the definition of the variable, in this case by restricting the observation time of interest to the time before the intercurrent event. Particular care is required if the occurrence of the intercurrent event differs between the treatments being compared (see A.3.3.).

Principal stratum strategies

This relates to the population of interest (see A.3.3.). The target population might be taken to be the “principal stratum” (see Glossary) in which an intercurrent event would occur. Alternatively, the target population might be taken to be the principal stratum in which an intercurrent event would not occur. The clinical question of interest relates to the treatment effect only within the principal stratum. For example, it might be desired to know a treatment effect on severity of infections in the principal stratum of patients becoming infected after vaccination. Alternatively, a toxicity might prevent some patients from continuing the test treatment, but it would be desired to know the treatment effect among patients who are able to tolerate the test treatment.

It is important to distinguish “principal stratification” (see Glossary), which is based on potential intercurrent events (for example, subjects who would discontinue therapy if assigned to the test product), from subsetting based on actual intercurrent events (subjects who discontinue therapy on their assigned treatment). The subset of subjects who experience an intercurrent event on the test treatment will often be a different subset from those who experience the same intercurrent event on control. Treatment effects defined by comparing outcomes in these subsets confound the effects of the different treatments with the differences in outcomes possibly due to the differing characteristics of the subjects.

A.3.3. Estimand Attributes

The attributes below are used to construct the estimand, defining the treatment effect of interest.

The treatment condition of interest and, as appropriate, the alternative treatment condition to which comparison will be made (referred to as “treatment” through the remainder of this document). These might be individual interventions, combinations of interventions administered concurrently, e.g. as add-on to standard of care, or might consist of an overall regimen involving a complex sequence of interventions. (see Treatment Policy and Hypothetical strategies under A.3.2.).

The population of patients targeted by the clinical question. This will be represented by the entire trial population, a subgroup defined by a particular characteristic measured at baseline, or a principal stratum defined by the occurrence (or non-occurrence, depending on context) of a specific intercurrent event (see Principal Stratum strategies under A.3.2.).

The variable (or endpoint) to be obtained for each patient that is required to address the clinical question. The specification of the variable might include whether the patient experiences an intercurrent event (see Composite Variable and While on Treatment strategies under A.3.2.).

Precise specifications of treatment, population and variable are likely to address many of the intercurrent events considered in sponsor and regulator discussions of the clinical question of interest. The clinical question of interest in respect of any other intercurrent events will usually be reflected using the strategies introduced as treatment policy, hypothetical or while on treatment.

Finally, a population-level summary for the variable should be specified, providing a basis for comparison between treatment conditions.

When defining a treatment effect of interest, it is important to ensure that the definition identifies an effect due to treatment and not due to potential confounders such as differences in duration of observation or patient characteristics.

A.3.4. Considerations for Constructing an Estimand

The clinical questions of interest and associated estimands should be specified at the initial stages of planning any clinical trial. Precise specification of objectives for most trials will need to reflect discontinuation of treatment and use of additional or alternative treatments. In some settings terminal events, such as death, should be addressed. Some trial objectives can only be described with reference to clinical events, for example the duration of response in subjects who achieve a response.

The construction of an estimand should consider what is of clinical relevance for the particular treatment in the particular therapeutic setting. Considerations include the disease under study, the clinical context (e.g. the availability of alternative treatments), the administration of treatment (e.g. one-off dosing, short-term treatment or chronic dosing) and the goal of treatment (e.g. prevention, disease modification, symptom control). Also important is whether an estimate of the treatment effect can be derived that is reliable for decision making. For example, a clinical question on the treatment effect on clinical outcome regardless of which other therapies are to be used before that outcome is experienced differs to a clinical question on the treatment effect had no additional medication been available. Depending on the setting, either might represent a clinical question of interest. However, in both cases, a clinical trial designed to estimate these treatment effects will often include the possibility to use additional medications if medically required. For the former question, values after the use of additional treatment will be relevant. For the latter question, values after the additional treatment are not directly relevant since the values also reflect the impact of that additional medication. It should be agreed that reliable estimation is possible before the choice of estimand is finalised. This includes, for the latter question, the methods to replace observations that are not to be used in the analysis.

When constructing the estimand it is necessary to have a clear understanding of the treatment to which the clinical question of interest pertains (see A.3.3.). Clear specifications for the treatments of interest might already reflect multiple relevant intercurrent events. Specifically, a treatment might already reflect the clinical question of interest in respect of changes in background treatment, concomitant medications, use of additional or later-line therapies, treatment-switching and conditioning regimens. For example, it is possible to specify treatment as intervention A added to background therapy B, dosed as required. In that case, changes to the dose of background therapy B would not need to be considered as an intercurrent event. However, the use of an additional therapy would need to be considered as an intercurrent event. If use of any additional medication is also reflected, using the treatment policy strategy for example, then treatment might be specified as intervention A added to background therapy B, dosed as required, and with additional medication, as required. Alternatively, if the treatment is specified as intervention A, then both changes in background therapy and use of additional therapy would be addressed as intercurrent events.

Discussions should also consider whether specifications for the population and variable attributes should be used to reflect the clinical question of interest in respect of any intercurrent events. Strategies can then be considered for any other intercurrent events. Usually an iterative process will be necessary to reach an estimand that is of clinical relevance for decision making, and for which a reliable estimate can be made. Some estimands, in particular those for which the measurements taken are relevant to the clinical question, can often be robustly estimated making few assumptions. Other estimands may require methods of analysis with more specific assumptions that may be more difficult to justify and that may be more sensitive to plausible changes in those assumptions (see A.5.1.). Where significant issues exist to develop an appropriate trial design or to derive an adequately reliable estimate for a particular estimand, an alternative estimand, trial design and method of analysis would need to be considered.

Avoiding or over-simplifying the process of discussing and constructing an estimand risks misalignment between trial objectives, trial design, data collection and method of analysis. Whilst an inability to derive a reliable estimate might preclude certain choices of strategy, it is important to proceed sequentially from the trial objective and an understanding of the clinical question of interest, and not for the choice of data collection and method of analysis to determine the estimand.

The experimental situation should also be considered. If the management of subjects (e.g. dose adjustment for intolerance, rescue treatment for inadequate response, burden of clinical trial assessments) under a clinical trial protocol is justified to be different to that which is anticipated in clinical practice, this might be reflected in the construction of the estimand.

Once constructed, the estimand should define a target of estimation clearly and unambiguously. Consider an intercurrent event of discontinuation of treatment; it is of utmost importance to distinguish between treatment effects of interest based on the principal stratum of patients who would be able to continue if administered the test treatment and the effect during continued treatment. Furthermore, neither of these should be taken to represent an effect if all patients can continue with treatment.

As stated above, when using the hypothetical strategy, some conditions are likely to be more acceptable for regulatory decision making than others. The hypothetical conditions described should therefore be justified for the quantification of an interpretable treatment effect that is relevant to inform the decisions to be taken by regulators, and use of the medicine in clinical practice. The question of what the values for the variable of interest would have been if rescue medication had not been available may be an important one. In contrast, the question of what the values for the variable of interest would have been under the hypothetical condition that subjects who discontinued treatment because of adverse drug reaction had in fact continued with treatment, might not be justifiable as being of clinical or regulatory interest. A clinical question of interest based on the effect if all subjects had been able to continue with treatment is not well-defined without a thorough discussion of the hypothetical conditions under which it is supposed that they would have continued. The inability to tolerate a treatment may constitute, in itself, evidence of an inability to achieve a favourable outcome.

Characterising beneficial effects using estimands based on the treatment policy strategy might also be more generally acceptable to support regulatory decision making, specifically in settings where estimands based on alternative strategies might be considered of greater clinical interest, but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference. An estimand based on the treatment policy strategy might offer the possibility to obtain a reliable estimate of a treatment effect that is still relevant. In this situation, it is recommended to also include those estimands that are considered to be of greater clinical relevance and to present the resulting estimates along with a discussion of the limitations, in terms of trial design or statistical analysis, for that specific approach. When constructing estimands based on the treatment policy strategy, inference can be complemented by defining an additional estimand and analysis pertaining to each intercurrent event for which the strategy is used; for example, contrasting both the treatment effect on a symptom score and the proportion of subjects using additional medication under each treatment. Similarly, an estimand using a while on treatment strategy should usually be accompanied by the additional information on the time to intercurrent event distributions, and an estimand based on a principal stratum would usefully be accompanied by information on the proportion of patients in that stratum, if available.

The considerations informing the construction of estimand to support regulatory decision making based on a non-inferiority or equivalence objective may differ to those for the choice of estimand for a superiority objective. As explained in ICH E9, the problem facing the regulator in their decision making is different when based on non-inferiority or equivalence studies compared to superiority studies. In Section 3.3.2. it is stated that such trials are not conservative in nature and the importance of minimising the number of protocol violations and deviations, non-adherence and study withdrawals is indicated. In Section 5.2.1. it is described that the result of the Full Analysis Set (FAS) is generally not conservative and that its role in such trials should be considered very seriously. Estimands that are constructed with one or more intercurrent events accounted for using the treatment policy strategy present similar issues for non-inferiority and equivalence trials as those related to analysis of the FAS under the ITT principle. Responses in both treatment groups can appear more similar following discontinuation of randomised treatment or use of another medication for reasons that are unrelated to the similarity of the initially randomised treatments. Estimands could be constructed to directly address those intercurrent events which can lead to the attenuation of differences between treatment arms (e.g. discontinuations from treatment and use

of additional medications). When selecting strategies, it might be important to distinguish between trials designed to detect whether differences exist between treatments containing the same or similar active substance (e.g. comparison of a biosimilar to a reference treatment) and trials where a non-inferiority or equivalence hypothesis is used in order to establish and quantify evidence of efficacy. An estimand can be constructed to target a treatment effect that prioritises sensitivity to detect differences between treatments, if appropriate for regulatory decision making.

A.4. IMPACT ON TRIAL DESIGN AND CONDUCT

The design of a trial needs to be aligned to the estimands that reflect the trial objectives. A trial design that is suitable for one estimand might not be suitable for other estimands of potential importance. Clear definitions for the estimands on which quantification of treatments effects will be based should inform the choices that are made in relation to trial design. This includes determining the inclusion and exclusion criteria that identify the target population, the treatments, including the medications that are allowed and those that are prohibited in the protocol, and other aspects of patient management and data collection. If interest lies, for example, in understanding the treatment effect regardless of whether a particular intercurrent event occurs, a trial in which the variable is collected for all subjects is appropriate. Alternatively, if the estimands that are required to support regulatory decision making do not require the collection of the variable after an intercurrent event, then the benefits of collecting such data for other estimands should be weighed against any complications and potential drawbacks of the collection.

Efforts should be made to collect all data that are relevant to support estimation, including data that inform the characterisation, occurrence and timing of intercurrent events. Data cannot always be collected. Certainly, subjects cannot be retained in a trial against their will, and in some trials missing data for some subjects is inevitable by design, such as administrative censoring in trials with survival outcomes. On the contrary, the occurrence of intercurrent events such as discontinuation of treatment, treatment switching, or use of additional medication, does not imply that the variable cannot be measured thereafter, though the measures may not be relevant. For terminal events such as death, the variable cannot be measured after the intercurrent event, but neither should these data generally be regarded as missing.

Not collecting any data needed to assess an estimand results in a missing data problem for subsequent statistical inference. The validity of statistical analyses may rest upon untestable assumptions and, depending on the proportion of missing data, this may undermine the robustness of the results (see A.5.). A prospective plan to collect informative reasons for why data intended for collection are missing may help to distinguish the occurrence of intercurrent events from missing data. This in turn may improve the analysis and may also lead to a more appropriate choice of sensitivity analysis. For example, “loss to follow-up” may more accurately be recorded as “treatment discontinuation due to lack of efficacy”. Where that has been defined as an intercurrent event, this can be reflected through the strategy chosen to account for that intercurrent event and not as a missing data problem. To reduce missing data, measures can be implemented to retain subjects in the trial. However, measures to reduce or avoid intercurrent events that would normally occur in clinical practice risk reducing the external validity of the trial. For example, selection of the

trial population or use of titration schemes or concomitant medications to mitigate the impact of toxicity might not be suitable if those same measures would not be implemented in clinical practice.

Randomisation and blinding remain cornerstones of controlled clinical trials. Design techniques for avoiding bias are addressed in Section 2.3. Certain estimands may necessitate, or may benefit from, use of trial designs such as run-in or enrichment designs, randomised withdrawal designs, or titration designs. It might be of interest to identify the principal stratum of subjects who can tolerate a treatment using a run-in period, in advance of randomising those subjects between test treatment and control. Dialogue between regulator and sponsor would need to consider whether the proposed run-in period is appropriate to identify the target population, and whether the choices made for the subsequent trial design (e.g. washout period, randomisation) supports the estimation of the target treatment effect and associated inference. These considerations might limit the use of these trial designs, and use of that particular strategy.

A precise description of the treatment effects of interest should inform sample size calculations. Particular care should be taken when making reference to historical studies that might, implicitly or explicitly, have reported estimated treatment effects or variability based on a different estimand. Where all subjects contribute information to the analysis, and where the impact of the strategy to reflect intercurrent events is included in the effect size that is targeted and the expected variance, it is not usually necessary to additionally inflate the calculated sample size by the expected proportion of subject withdrawals from the trial.

Section 7.2. addresses issues related to summarising data across clinical trials. The need to have consistent definitions for the variables of interest is highlighted and this can be extended to the construction of estimands. Hence, in situations when synthesising evidence from across a clinical trial programme is envisaged at the planning stage, a suitable estimand should be constructed, included in the trial protocols, and reflected in the choices made for the design of the contributing trials. Similar considerations apply to the design of a meta-analysis, using estimated effect sizes from completed trials to determine non-inferiority margins, or the use of external control groups for the interpretation of single-arm trials. A naïve comparison between data sources, or integration of data from multiple trials without consideration and specification of the estimand that is addressed in each data presentation or statistical analysis, could be misleading.

More generally, a trial is likely to have multiple objectives translated into multiple estimands, each associated with statistical testing and estimation. The multiplicity issues arising should be addressed.

A.5. IMPACT ON TRIAL ANALYSIS

A.5.1. Main Estimation

An estimand for the effect of treatment relative to a control will be estimated by comparing the outcomes in a group of subjects on the treatment to those in a similar group of subjects on the control. For a given estimand, an aligned method of analysis, or estimator, should be implemented that is able to provide an estimate on which reliable interpretation can be based. The method of

analysis will also support calculation of confidence intervals and tests for statistical significance. An important consideration for whether an interpretable estimate will be available is the extent of assumptions that need to be made in the analysis. Key assumptions should be stated explicitly together with the estimand and accompanying main and sensitivity estimators. Assumptions should be justifiable and implausible assumptions should be avoided. The robustness of the results to potential departures from the underlying assumptions should be assessed through an estimand-aligned sensitivity analysis (see A.5.2.). Estimation that relies on many or strong assumptions requires more extensive sensitivity analysis. Where the impact of deviations from assumptions cannot be comprehensively investigated through sensitivity analysis, that particular combination of estimand and method of analysis might not be acceptable for decision making.

All methods of analysis rely on assumptions, and different methods may rely on different assumptions even when aligned to the same estimand. Nevertheless, some kinds of assumption are inherent in all methods of analysis aligned to estimands that use each of the different strategies outlined; for example, the methodology for predicting the outcomes that would have been observed in the hypothetical scenario, or for identifying a suitable target population in a principal stratum strategy. Some examples are given below related to the different strategies used to reflect the occurrence of intercurrent events. The issues highlighted will be key components of discussion between sponsor and regulator in advance of an estimand, main analysis and sensitivity analysis being agreed.

Analysis aligned with a treatment policy strategy to address a given intercurrent event may entail stronger or weaker assumptions depending on the design and conduct of the trial. When most subjects are followed-up even after the respective intercurrent event (e.g. discontinuation of treatment), the remaining problem of missing data may be relatively minor. In contrast, when observation is terminated after an intercurrent event, which is obviously undesirable in respect of this strategy, the assumption that (unobserved) outcomes for discontinuing subjects are similar to the (observed) outcomes for those who remain on treatment will often be implausible. An alternative approach to handle the missing data would need to be justified and sensitivity analysis will be expected.

Analysis aligned to a hypothetical strategy involves outcomes different from those actually observed; for example, outcomes if rescue medication had not been given when in fact it was. Observations before the rescue medication and observations on subjects who did not require rescue medication may be informative, but only under strong assumptions.

A composite variable strategy can avoid statistical assumptions about data after an intercurrent event by considering occurrence of the intercurrent event as a component of the outcome. The potential concern relates less to assumptions for estimation, and more to the interpretation of the estimated treatment effect. For the estimand to be interpretable, if scores are assigned for failure because the intercurrent event occurs, these should meaningfully reflect the lack of benefit to the patient (e.g. death may be reflected differently than discontinuation of treatment due to adverse event).

Estimands constructed based on a while on treatment strategy can be estimated provided outcomes are collected up to the time of the intercurrent event. Again, the crucial assumptions concern interpretation. Take discontinuation of treatment by way of example. Outcomes while on

treatment may be improved but the treatment may also shorten, or lengthen, the treatment period by provoking, or delaying, discontinuations, and both these effects should be considered in interpretation and assessment of clinical benefit.

Analysis aligned to a principal stratum strategy usually requires strong assumptions. For example, some principal stratification methods infer this from baseline characteristics of the subjects, but the correctness of this inference may be difficult to assess. This difficulty cannot be avoided by simplified methods, however. For example, simply comparing subjects who do not have an intercurrent event on the test treatment to those who do not have an event on control, assuming intercurrent events are unrelated to treatment, is very difficult to justify.

Even after defining estimands that address intercurrent events in an appropriate manner and making efforts to collect the data required for estimation (see A.4.), some data may still be missing, including e.g. administrative censoring in trials with survival outcomes. Failure to collect relevant data should not be confused with the choice not to collect, or to collect and not to use, data made irrelevant by an intercurrent event. For example, data that were intended to be collected after discontinuation of trial medication to inform an estimand based on the treatment policy strategy are missing if uncollected; however, the same data points might be irrelevant for another strategy, and thus, for the purpose of that second estimand, are not missing if uncollected. Where those efforts to collect data are not successful it becomes necessary to make assumptions to handle the missing data in the statistical analysis. Handling of missing data should be based on clinically plausible assumptions and, where possible, guided by the strategies employed in the description of the estimand. The approach taken may be based on observed covariates and post-baseline data from individual subjects and from other similar subjects. Criteria to identify similar subjects might include whether or not the intercurrent event has occurred. For example, for subjects who discontinue treatment without further data being collected, a model may use data from other subjects who discontinued treatment but for whom data collection has continued.

A.5.2. Sensitivity Analysis

A.5.2.1. Role of Sensitivity Analysis

Inferences based on a particular estimand should be robust to limitations in the data and deviations from the assumptions used in the statistical model for the main estimator. This robustness is evaluated through a sensitivity analysis. Sensitivity analysis should be planned for the main estimators of all estimands that will be important for regulatory decision making and labelling in the product information. This can be a topic for discussion and agreement between sponsor and regulator.

The statistical assumptions that underpin the main estimator should be documented. One or more analyses, focused on the same estimand, should then be pre-specified to investigate these assumptions with the objective of verifying whether or not the estimate derived from the main estimator is robust to departures from its assumptions. This might be characterised as the extent of departures from assumptions that change the interpretation of the results in terms of their statistical or clinical significance (e.g. tipping point analysis).

Distinct from sensitivity analysis, where investigations are conducted with the intent of exploring robustness of departures from assumptions, other analyses that are conducted in order to more fully investigate and understand the trial data can be termed “supplementary analysis” (see Glossary; A.5.3.). Where the primary estimand(s) of interest is agreed between sponsor and regulator, the main estimator is pre-specified unambiguously, and the sensitivity analysis verifies that the estimate derived is reliable for interpretation, supplementary analyses should generally be given lower priority in assessment.

A.5.2.2. Choice of Sensitivity Analysis

When planning and conducting a sensitivity analysis, altering multiple aspects of the main analysis simultaneously can make it challenging to identify which assumptions, if any, are responsible for any potential differences seen. It is therefore desirable to adopt a structured approach, specifying the changes in assumptions that underlie the alternative analyses, rather than simply comparing the results of different analyses based on different sets of assumptions. The need for analyses varying multiple assumptions simultaneously should then be considered on a case by case basis. A distinction between testable and untestable assumptions may be useful when assessing the interpretation and relevance of different analyses.

The need for sensitivity analysis in respect of missing data is established and retains its importance in this framework. Missing data should be defined and considered in respect of a particular estimand (see A.4.). The distinction between data that are missing in respect of a specific estimand and data that are not directly relevant to a specific estimand gives rise to separate sets of assumptions to be examined in sensitivity analysis.

A.5.3. Supplementary Analysis

Interpretation of trial results should focus on the main estimator for each agreed estimand providing that the corresponding estimate is verified to be robust through the sensitivity analysis. Supplementary analyses for an estimand can be conducted in addition to the main and sensitivity analysis to provide additional insights into the understanding of the treatment effect. They generally play a lesser role for interpretation of trial results. The need for, and utility of, supplementary analyses should be considered for each trial.

Section 5.2.3. indicates that it is usually appropriate to plan for analyses based on both the FAS and the Per Protocol Set (PPS) so that differences between them can be the subject of explicit discussion and interpretation. Consistent results from analyses based on the FAS and the PPS is indicated as increasing confidence in the trial results. It is also described in Section 5.2.2. that results based on a PPS might be subject to severe bias. In respect of the framework presented in this addendum, it may not be possible to construct a relevant estimand to which analysis of the PPS is aligned. As noted above, analysis of the PPS does not achieve the goal of estimating the effect in any principal stratum, for example, in those subjects able to tolerate and continue to take the test treatment, because it may not compare similar subjects on different treatments.

Protocol violations and deviations might exclude subjects from the PPS, for example by having a visit outside a time window, without an intercurrent event necessarily having occurred. Likewise, subjects could experience an intercurrent event, such as death, without having deviated from the

protocol. Notwithstanding the differences between violations and deviations from the protocol and intercurrent events, events likely to affect the interpretation or existence of measurements are considered in the description of the estimand. Estimands might be constructed, with aligned method of analysis, that better address the objective usually associated with the analysis of the PPS. If so, analysis of the PPS might not add additional insights.

A.6. DOCUMENTING ESTIMANDS AND SENSITIVITY ANALYSIS

A trial protocol should define and specify explicitly a primary estimand that corresponds to the primary trial objective. The protocol and the analysis plan should pre-specify the main estimator that is aligned with the primary estimand and leads to the primary analysis, together with a suitable sensitivity analysis to explore the robustness under deviations from its assumptions. Estimands for secondary trial objectives (e.g. related to secondary variables) that are likely to support regulatory decisions should also be defined and specified explicitly, each with a corresponding main estimator and a suitable sensitivity analysis. Additional exploratory trial objectives may be considered for exploratory purposes, leading to additional estimands.

The choice of the primary estimand will usually be the main determinant for aspects of trial design, conduct and analysis. Following usual practices, these aspects should be well documented in the trial protocol. If secondary estimands are of key interest, these considerations may be extended to support these as needed and should be documented as well. Beyond these aspects, the conventional considerations for trial design, conduct and analysis remain the same.

While it is to the benefit of the sponsor to have clarity on what is being estimated, it is not a regulatory requirement to document an estimand for each exploratory objective.

Results from the main, sensitivity and supplementary analyses should be reported systematically in the clinical trial report, specifying whether each analysis was pre-specified, introduced while the trial was still blinded, or performed post hoc. Summaries of the number and timings of each intercurrent event in each treatment group should be reported.

Changes to the estimand during the trial can be problematic and can reduce the credibility of the trial. Addressing intercurrent events that were not foreseen at the design stage, and are identified during the conduct of the trial, should discuss not only the choices made for the analysis, but the effect on the estimand, i.e. on the description of the treatment effect that is being estimated, and the interpretation of the trial results. A change to the estimand should usually be reflected through amendment to the protocol.

GLOSSARY

Term	Content
------	---------

Estimand:	A precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarises at a population-level what the outcomes would be in the same patients under different treatment conditions being compared.
Estimate:	A numerical value computed by an estimator.
Estimator:	A method of analysis to compute an estimate of the estimand using clinical trial data.
Intercurrent Events:	Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. It is necessary to address intercurrent events when describing the clinical question of interest in order to precisely define the treatment effect that is to be estimated.
Missing Data:	Data that would be meaningful for the analysis of a given estimand but were not collected. They should be distinguished from data that do not exist or data that are not considered meaningful because of an intercurrent event.
Principal Stratification:	Classification of subjects according to the potential occurrence of an intercurrent event on all treatments. With two treatments, there are four principal strata with respect to a given intercurrent event: subjects who would not experience the event on either treatment, subjects who would experience the event on treatment A but not B, subjects who would experience the event on treatment B but not A, and subjects who would experience the event on both treatments. In this document a principal stratum refers to any of the strata (or combination of strata) defined by principal stratification.
Sensitivity Analysis:	A series of analyses conducted with the intent to explore the robustness of inferences from the main estimator to deviations from its underlying modelling assumptions and limitations in the data.
Supplementary Analysis:	A general description for analyses that are conducted in addition to the main and sensitivity analysis with the intent to provide additional insights into the understanding of the treatment effect.

ICH E9(R1)

????????????????E9

??????????

English Version

[ICH E9\(R1\) Addendum: Statistical Principles for Clinical Trials](#)

A.1. ?????

“ ”

A.3. “ ”

A.3.1. “ ”

“ ”

ICH E9 x.y A.x.y.

ICH E9 ITT

ITT ICH E9

ITT

“ ”

5.2.

1.2.

A.2. ?????????????????????

1

“ ” A.3.

” A.5.1.

A.5.2.

[Redacted]
 [Redacted]
 [Redacted]
 [Redacted]

[Redacted]
 [Redacted]
 [Redacted]
 [Redacted] A.3.2. [Redacted]
 [Redacted]
 [Redacted]

[Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]
 [Redacted]

[Redacted]
 [Redacted]
 [Redacted]

A.3.2. ??????????????????

[Redacted]
 [Redacted]
 [Redacted] A.3.4.
 [Redacted]

[Redacted]

[Redacted]
 [Redacted]

[Redacted]
 [Redacted] ICH E9 [Redacted] ITT
 [Redacted]

[Redacted]
 [Redacted]

[Redacted]

A

The diagram shows three horizontal bars, each composed of small squares. The top bar is the longest, followed by the middle bar, and the bottom bar is the shortest. These bars represent the relative lengths of the three longest rivers in the world.

A diagram consisting of 10 rows of 10 small squares each, totaling 100 squares. The first row is shaded gray, representing 10% of the total.

Diagram illustrating the structure of the ICH E9 data set, showing a sequence of data elements (represented by boxes) and labels (ICH E9 and FAS).

3.3.2.

5.2.1.

A.4.

A.5.2. ????

A.5.2.1. ????????

“”

A.5.3.

A.5.2.2. ????????

A.4.

A.5.3. ????

[illegible]

項目	内容
1. 概要	本報告は、2023年度における事業の進捗状況と、今後の見通しについてまとめたものである。 1.1 事業の概要 1.2 事業の目的
2. 事業の進捗状況	2.1 計画と実績の比較 2.2 進捗率の推移
3. 今後の見通し	3.1 今後の計画 3.2 課題と対応策

□□□□	□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□ □□□□□□□□
□□□□	□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□
□□□□	□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□ A□□□□□□□□ B □□□□□□□□□□□□□□□□□□ B□□□□□□□□ A □□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□ □□□□
□□□□ /td>	□□□□□□□□□□□□□□□□□□ □□□□□□□□
□□□□	□□□□□□□□□□□□□□□□□□ □□□□□□□□