人用药品注册技术要求国际协调会（ICH）

**ICH 协调指导原则**

# 临床试验的统计原则

# E9

现行第 4 阶段版本

1998年2月5日起实施

*该指导原则由相应的 ICH 专家小组制定，按照 ICH 进程，已递交管理部门讨论。在 ICH 进程第四阶段，最终草案被推荐给欧盟、日本和美国的管理机构采纳。*

# E9

**文件历史**

| 最初编号 | 历史 | 日期 | 新编号<br>**2005 年 11 月** |
|---|---|---|---|
| E9 | 在第 2 阶段由指导委员会批准并公开征求意见。 | 1997 年 1 月 16 日 | E9 |

**现行第 4 阶段版本**

| | | | |
|---|---|---|---|
| E9 | 由指导委员会批准的第 4 阶段，并建议三个 ICH 监管机构采纳。 | 1998 年 1 月 5 日 | E9 |

# 临床试验的统计原则

**ICH 协调指导原则**

ICH 指导委员会在 1998 年 2 月 5 日的会议上达成第 4 阶段，建议 ICH 三个监管机构采纳该指导原则。

## 目　录

# STATISTICAL PRINCIPLES FOR CLINICAL TRIALS

# 临床试验的统计原则

## I. INTRODUCTION

### 1. 引言

### 1.1 Background and Purpose

### 1.1 背景与目的

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.

医药产品的有效性和安全性需由临床试验来论证。所采用的临床试验需遵循 ICH 在 1996 年 5 月 1 日通过的“良好临床实践（GCP）：综合指南”（ICH E6）。ICH E6 已阐明统计在临床试验设计和分析中不可或缺的作用。由于统计学研究在临床试验领域的不断发展，加之临床研究在药物审批流程及一般医疗保健中的重要作用，因此，有必要制订一份关于临床试验统计问题的简明文件。本指南旨在协调在欧洲、日本和美国提交上市申请的临床试验所应用的统计学方法的原则。

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.

作为起点，本指南使用了欧盟专利医药产品委员会（CPMP）在题为《用于申请医药产品上市许可的临床试验生物统计方法》（1994 年 12 月）指南的意见，并参照了日本厚生省的《临床研究中的统计分析指南》（1992 年 3 月）和美国食品药品监督管理局的《新药申请中临床与统计部分的格式与内容指南》（1998 年 7 月）。其他 ICH 指南也包含一些与统计原则和方法有关的主题，特别是下面所

列的指南。本指南将在各个部分中对包含相关内容的特定指南进行标识。

E1A: The Extent of Population Exposure to Assess Clinical Safety

E1A：人群暴露程度对评价临床安全性的影响

E2A: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting

E2A：临床安全性数据管理：快速报告的定义与标准

E2B: Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports

E2B：临床安全性数据管理：个例安全报告传输数据元素

E2C: Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs

E2C：临床安全性数据管理：上市药品的定期安全性更新报告

E3: Structure and Content of Clinical Study Reports

E3：临床研究报告的结构与内容

E4: Dose-Response Information to Support Drug Registration

E4：支持药品注册的剂量反应信息

E5: Ethnic Factors in the Acceptability of Foreign Clinical Data

E5：国外临床数据可接受性的种族因素

E6: Good Clinical Practice: Consolidated Guideline

E6：良好临床实践：综合指南

E7: Studies in Support of Special Populations: Geriatrics

E7：特殊人群的支持性研究：老年医学

E8: General Considerations for Clinical Trials

E8：临床试验的一般考虑

E10: Choice of Control Group in Clinical Trials

E10：临床试验中对照组的选择

M1: Standardisation of Medical Terminology for Regulatory Purposes

M1：用于监管目的的医学术语标准化

M3: Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals.

M3：用于实施药物人体临床试验的非临床安全性研究

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.

本指南旨在为申办者在整体临床研发背景下，对研究产品临床试验的设计、实施、分析和评价提供指导。本指南也将会帮助科学专家准备上市申请总结报告

或者评价主要来自研发后期的临床试验的有效性和安全性证据。

**1.2 Scope and Direction**

**1.2 范围与说明**

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.

本指南的重点是统计原则，并不涉及具体统计步骤或方法的使用。确保这些原则得到正确实施的具体程序性步骤是申办者的职责。本指南对不同临床试验之间的数据整合亦作了讨论，但并不作为重点。其他 ICH 指南涵盖了与数据管理及临床试验监查活动有关的原则和程序，此处不再赘述。

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.

本指南对具有广泛学科背景的人们是有意义的。然而，正如 ICH E6 所示，在此假定所有与临床试验有关的统计工作的真正职责将由具有足够资格和经验的统计专业人员承担。试验统计专业人员（见词汇表）在与其他临床试验专家合作时，其作用和职责是确保在支持药物研发的临床试验中恰当地应用统计原则。因此，试验统计专业人员应同时具备充足的教育/培训和经验以贯彻本指南所阐明的原则。

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.

对于每一个用于上市申请的临床试验，它的设计和实施的所有重要细节以及

所提出的统计分析的主要特征，都应在试验开始前所写的方案中明确规定。方案中的步骤被遵循的程度以及主要分析被预先计划的程度，都将决定试验最终结果和结论的可信度。方案及后续修订应该获得包括试验统计专业人员在内的负责人员的批准。试验统计专业人员应保证方案以及任何修订都能清楚准确地涵盖所有相关的统计问题，并恰当使用技术术语。

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.

本指南概述的原则首先与在研发后期阶段实施的临床试验有关，其中很多是有效性的确证性试验。除有效性之外，确证性试验也可把安全性指标（如，不良事件、临床实验室指标或心电图测量）、药效学或药代动力学指标（如在确证性生物等效性试验中）作为它们的主要指标。其次，有些确证性结果可能来源于不同试验之间整合的数据，本指南中有些原则适用于此种情况。最后，虽然药物研发的早期阶段本质上主要是探索性的临床试验，但统计原则也与这些临床试验有关。因此，本指南的内容应尽可能应用到临床研发的各个阶段。

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.

本指南中所描述的很多原则都用于最小化偏倚（见词汇表）和最大化精度。在本指南中所用的术语“偏倚”是指与临床试验设计、实施、分析和结果解释有关的任何因素所导致的处理效应（见词汇表）估计值与真实值偏离的系统性趋势。重要的是尽可能完全地识别偏倚的潜在来源，以便采取措施限制这些偏倚。偏倚的存在可能严重削弱从临床试验中得出正确结论的能力。

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.

有些偏倚源于试验设计，例如，处理的分配，将风险较低的受试者系统地分配到一个处理中。其他偏倚源于临床试验的实施和分析。例如，违背方案以及基于对受试者结局的了解从分析中排除受试者是偏倚的可能来源，这可能影响处理效应的准确估计。偏倚常在不知不觉中发生，且难以直接测量，因而评价试验结果和主要结论的稳健性是重要的。稳健性是一个概念，是指总体结论对数据的各种限制、假设和数据分析方法的敏感性。稳健性意味着，当基于另一假设或分析方法进行分析时，试验的处理效应和主要结论不会受到实质性的影响。在解释处理效应和处理之间比较的不确定性的统计测量时，应考虑偏倚对 *P* 值、置信区间或推断的潜在影响。

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.

由于临床试验设计和分析的主要方法基于频率学派统计方法，因此在讨论假设检验和/或置信区间时，本指南主要使用频率学派方法（见词汇表）。这并不意味着其它方法不可取，如果理由充分且所得结论是足够稳健的，则贝叶斯方法（见词汇表）及其他方法亦可考虑。

## II. CONSIDERATIONS FOR OVERALL CLINICAL DEVELOPMENT

## 2. 总体临床研发的考虑

### 2.1 Trial Context

### 2.1 试验背景

#### *2.1.1 Development Plan*

### 2.1.1 研发计划

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.

新药临床研发过程的主要目标是寻找是否存在一个剂量范围和给药方案，在风险-获益关系可接受的范围内，药物可以同时显示出安全性和有效性。同时还要定义可能从该药获益的特定受试者及明确的适应症。

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.

满足这些目标通常需要一系列循序渐进的临床试验，每一个临床试验有它自己特定目的（见 ICH E8）。这应该在一个或一系列临床计划中规定，这些计划应具有适当的决策点和随着知识累积而进行修订的灵活性。上市申请应清晰地描述这些计划的主要内容，以及每个试验的作用。对整个试验项目所提供证据的解释和评价涉及对单个试验证据的综合（见第 7.2 章节）。如果确保针对于试验一些特征，如医学术语词典、主要测量的定义与时点、方案违背的处理等采用通用标准，则有助于实现对证据的综合。当医学问题涉及一个以上的试验时，统计汇总、概述或 Meta 分析（见词汇表）可能会提供信息。可能的话，应在计划中考虑这一点，以便预先清晰的确定相关的试验，并规定这些试验在设计上所必要的共同特征。应该在该计划中阐述可能会影响共同计划中若干试验的其他主要统计学问题（如果有的话）。

***2.1.2 Confirmatory Trial***

### 2.1.2 确证性试验

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.

确证性试验是一种预先提出假设并进行评价的具有充分对照的试验。通常，需要通过确证性试验提供有效性或安全性的确凿证据。在这类试验中，感兴趣的关键假设直接来源于试验的主要目的，通常被预先定义，且在试验完成后检验该假设。在确证性试验中，对于感兴趣的处理，以适当的精度估计其效应大小，与把这些效应和临床意义联系起来是同等重要的。

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.

确证性试验旨在为支持主张提供确凿证据，因此，按照方案及标准操作程序进行试验尤为重要；应该解释和书面记录不可避免的改变，并考察它们的影响。每个此类试验的设计合理性，以及其它重要的统计方面如计划分析的主要特征均应在方案中列出。每个试验应仅解决有限数量的问题。

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.

支持主张的确凿证据要求确证性试验的结果证实所检验的研究产品具有临床获益。因此确证性试验应清晰明确地回答每一个与有效性或安全性主张有关的关键临床问题。另外，能够理解并解释推论（见词汇表）到目标患病人群的基础是很重要的，这也会影响到所需研究中心和/或试验的数量和类型（例如，专家或全科医师）。确证性试验的结果应当稳健。在某些情况下，来自于单一确证性

试验所提供证据的权重可能就足够了。

***2.1.3 Exploratory Trial***

**2.1.3 探索性试验**

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.

确证性试验的理论基础和设计几乎总是依赖于一系列早期探索性临床研究工作。如同所有的临床试验，这些探索性研究也应有清晰和明确的目的。但与确证性试验相比，它们的目的并不总是对预先定义的假设进行简单的检验。此外，探索性试验可能有时需要采用更灵活的方法进行设计，以便根据积累的结果对设计进行修改。它们的分析可能仅限于数据探索；可能进行假设检验，但假设的选择可能依赖于数据。尽管这类试验可能对整体有效性证据有贡献，但不能作为证明有效性的正式依据。

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.

任何单个试验可能同时具有确证性和探索性两方面。例如，在大多数确证性试验中，也会对数据进行探索性分析，作为解释和支持研究发现、为后期研究提出进一步假设的基础。方案应明确区分用作确证性依据的试验方面和为探索性分析提供数据方面的内容。

**2.2 Scope of Trials**

**2.2 试验范围**

***2.2.1 Population***

**2.2.1 人群**

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.

在药物研发的早期阶段，临床试验受试者的选择在很大程度上受到希望最大可能地观察到感兴趣的特定临床疗效的主观愿望的影响，因此，研究对象往往是总体患者人群中一个非常局限的亚组，最终药物可能适用于该亚组。但在开展确证性试验的时候，试验受试者应更能反映目标人群。因此在这些试验中，在保持足够的同质性以允许精确估计处理效应的同时，在目标人群中尽可能放宽纳入和排除标准通常是有益的。由于地理位置、实施时间、特定研究者和诊所的医疗实践等因素的影响，没有一个单一临床试验可以期望完全代表将来的使用者。尽管如此，应尽可能减少这些因素的影响，并在解释试验结果时充分讨论。

***2.2.2 Primary and Secondary Variables***

**2.2.2 主要和次要指标**

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).

主要指标（又称“目标”指标，主要终点）应该能够提供与试验主要目的直接相关的最具临床相关性和说服力的证据。通常应只有一个主要指标。因大部分确证性试验的主要目的是提供与有效性相关的强有力的科学证据，所以主要指标通常是有效性指标。安全性/耐受性有时也可能是主要指标，并且总是一个重要的

考虑内容。有关生活质量和卫生经济的测量也是进一步的潜在主要指标。主要指标的选择应反映相关研究领域公认的准则和标准。建议使用在早期研究或已发表文献中获得的已有实践经验的可靠且已验证的指标。在纳入和排除标准所描述的患者人群中，应该有充分的证据说明主要指标能够对一些临床相关的和重要的治疗获益提供有效且可靠的测量。主要指标通常应当是用于样本量估计的指标（见第3.5章节）。

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.

在很多情况下，评价受试者结局的方法可能并不直接，应仔细定义。例如，将死亡率规定为主要指标而无进一步说明是不够的；对死亡率的评价可以是比较某些固定时点的存活比例，也可以是比较在特定时域内生存时间的总体分布。另一个常见的例子是复发事件，处理效应的测量可以是简单的二分类指标（特定时域内任何复发）、到首次复发的时间、复发率（单位时间内观察到的事件数）等等。在慢性病治疗的研究中，功能状态随时间变化的评价对选择主要指标提出了其他挑战。存在很多可能的方法，例如比较观察时域开始和结束时所做的评价、比较由时域内所有评价而计算得出的斜率、比较超过或低于规定阈值的受试者比例、基于重复测量数据方法的比较。为避免因事后定义所产生的多重性担忧，在方案中规定主要指标的精确定义是至关重要的，因为在统计分析中将使用它。另外，所选择的具体主要指标的临床相关性和相关测量程序的合理性通常需要在方案中加以处理和说明。

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.

主要指标及其选择理由应在方案中详细说明。揭盲后重新定义主要指标通常是不可接受的，因为由此引入的偏倚很难评价。当根据主要目的确定的临床效应存在多种测量方法时，应根据临床相关性、重要性、客观性、和/或其它相关特性在方案中确定一种实际可行的测量方法作为主要指标。

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.

次要指标是与主要目的相关的支持性指标，或与次要目的相关的效应指标。在方案中预先定义次要指标，并说明它们的相对重要性以及在解释试验结果时的作用，也都是很重要的。次要指标的数目应当是有限的，且应与试验中待回答的有限数量的问题相关。

***2.2.3 Composite Variables***

**2.2.3 复合指标**

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.

当与主要目的相关的多种测量方法中难以确定单一的主要指标时，另一种有用的策略是按预先确定的计算方法将多个指标组合成一个单一或“复合”指标。事实上，主要指标有时以多种临床测量方法相组合的形式出现（如在关节炎、精神

疾病及其它疾病中的等级量表）。该法虽涉及多重性问题，但不需对 I 类错误进行调整。将多个指标组合的方法应在方案中详细说明，且应以临床获益的大小对结果进行解释。当复合指标被用作主要指标时，可以对复合指标中有临床意义的单个指标进行单独分析。当等级量表被用作主要指标时，对内容效度（见词汇表），评价者内和评价者间信度（见词汇表）及检测疾病严重程度变化的反应性等因素进行处理尤其重要。

***2.2.4 Global Assessment Variables***

**2.2.4 全局评价指标**

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.

在某些情况下，用全局评价指标（见词汇表）来评价某个处理的总体安全性、有效性和/或实用性。这种指标类型整合了客观指标和研究者对受试者的状态或状态变化的总体印象，它通常是一个有序分类等级指标。总体有效性的全局评价方法已经在一些治疗领域建立，如神经病学和精神病学。

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:

全局评价指标一般带有主观成分。使用全局评价指标作为主要或次要指标时，应该在方案中对量表的以下方面进行详细说明：

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.

1）该量表与试验主要目的的相关性；

2）该量表效度和信度的基础；

3）如何根据所收集的单一受试者数据，将他/她归类到量表中的特定类别；

4）如何将有缺失数据的受试者归类到量表中的特定类别，或用其他方法评价。

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.

若研究者选取的全局评价指标中包含客观指标，则这些客观指标应作为附加的主要指标，或至少作为重要的次要指标。

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.

实用性的全局评价由获益与风险综合得出，反映了治疗医生的决策过程，在做出使用产品的决策时，医生必须权衡获益与风险。全局实用性指标的一个问题是，在某些情况下，使用它们会导致将在获益和不良反应方面特征差别很大的两种产品判断为等效。例如，将一种治疗的全局实用性判断为等效于或优效于另一种处理时可能掩盖了它效果甚微或无效但不良反应较少的事实。因此不建议将全局实用性指标作为主要指标。如果全局实用性被用作主要指标，则将特定的有效性和安全性结局分别作为附加的主要指标考虑是非常重要的。

***2.2.5 Multiple Primary Variables***

**2.2.5 多个主要指标**

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.

有时需要使用一个以上的主要指标，每一个（或其中一个子集）都足以涵盖

治疗效果的范围。对解释这类证据的计划方式应当详细说明。例如，应该说明对任一指标，或最少几个指标，或全部指标的影响是否被认为是达到试验目的所必需的。应该针对已定义的主要指标清楚地说明主要假设或感兴趣的假设与参数（如均数、百分数、分布），并清楚地描述统计推断方法。因为存在潜在的多重性问题，所以应解释对 I 类错误的影响（见第 5.6 章节），也应在方案中给出控制 I 类错误的方法。在评价对 I 类错误的影响时，所提出的主要指标之间的相关程度也可以考虑。如果试验目的是证实对所有指定的主要指标的影响，则无须调整 I 类错误，但必须仔细考虑对 II 类错误和样本量的影响。

***2.2.6 Surrogate Variables***

**2.2.6 替代指标**

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.

当通过观察实际临床有效性不能直接评价受试者的临床获益时，可以考虑间接标准（替代指标—见词汇表）。通常可接受的替代指标被用于许多适应症，在这些适应症中，它们被认为是临床获益的可靠预测因子。在引入任何被提出的替代指标时有两个主要关注点。第一，它可能不是感兴趣的临床结局的真正预测因子。例如，它可以测量与一个特定的药理学机制有关的治疗活性，但不能提供治疗的作用范围与最终效果的全部信息，无论是阳性还是阴性。已有很多例子，对

提出的替代指标显示出高度阳性效应的治疗，最终被证明对受试者的临床结局是有害的。与此相反，也有一些例子，有些治疗显示出了临床获益，但对提出的替代指标却无可测量的影响。第二，提出的替代指标可能不会定量测量可直接权衡不良反应的临床获益。验证替代指标的统计标准已经具备，但是使用它们的经验相对有限。事实上，替代证据的强度取决于（1）替代关系的生物学合理性，（2）流行病学研究证明替代指标对临床结局的预后价值，（3）临床试验证明替代指标的处理效应相当于临床结局的效应。一种产品的临床指标和替代指标之间的关系并不一定适用于治疗同一种疾病但具有不同作用模式的另一种产品。

***2.2.7 Categorised Variables***

**2.2.7 分类指标**

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.

有时可能需要将连续或有序指标转化为二分类或其他分类。“成功”和“有应答”的标准是二分类的常见例子，这些分类标准需要明确规定，例如，在连续指标中（相对于基线）最小百分比的改善，或者按照等于或高于有序等级量表中某个阈值水平（如“好”）划分的等级。

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.

舒张压降低于 90mmHg 是一个常见的二分类。当分类有明确的临床相关性时，它们是最有用的。由于知道试验结果很容易在选择分类标准时产生偏倚，因此在方案中应预先定义和详细说明分类标准。由于分类通常意味着信息丢失，因此在分析中将损失把握度，应该在样本量计算时加以考虑。

**2.3 Design Techniques to Avoid Bias**

**2.3 避免偏倚的设计技术**

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.

在临床试验中，用于避免偏倚的最重要的设计技术是盲法和随机化，这些应是上市申请中所包含的大多数对照临床试验的一般特征。大多数这样的试验采用双盲法，按照合适的随机化方案，对治疗药物进行预先包装并提供给试验中心，只标明受试者编号和治疗周期，从而使参与试验的任何人都不知道分配给任何特定受试者的具体治疗药物，甚至不知道编码字母。这种方法将在第 2.3.1 章节和第 2.3.2 章节中的大部分内容中进行介绍，例外情况将在最后考虑。

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.

任何可能损害分析满意程度的可预期的不规范操作，包括各种类型的方案违背、退出和缺失值，通过在方案中规定程序使它们在试验实施过程中最小化，也可以减少设计阶段的偏倚。方案中应考虑一些方法，以减少出现这些问题的频率，以及解决在数据分析中出现的问题。

***2.3.1 Blinding***

### 2.3.1 盲法

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.

盲法或遮蔽是为了控制临床试验的实施过程中以及解释结果时产生的有意或无意的偏倚，这些偏倚来源于知道所采用的治疗后对受试者的招募和分配、受试者的后续护理、受试者对处理的态度、终点评价、对退出的处理、从分析中排除数据等产生的影响。盲法的根本目标是防止知道采用何种治疗以控制产生偏倚的所有机会。

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.

在双盲试验中，所有受试者及所有参与受试者的治疗或临床评价的研究者和申办人员都不知道所接受的治疗。这包括确定受试者资格、评价终点或评价对方案依从性的任何人。在整个试验实施过程中，这种盲态要持续保持，只有当数据被清理到可接受的质量水平时，才可对适当的人员揭盲。如果需要对不参与受试者的治疗或临床评价的申办者揭盲治疗编码（如生物分析学家、稽查员、参与严重不良事件报告的人员），申办者应该制定严格的标准操作程序，以防止治疗编码的不当传播。在单盲试验中，研究者和/或他的成员知道治疗分组信息，但受试者不知道，反之亦然。在开放试验中，所有的人都可能知道治疗分组信息。双盲试验是最优方法。这要求在试验中所采用的治疗在使用前或使用期间均无法被识别出来（外观、味道等），且在整个试验期间均适当地保持盲态。

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.

实现双盲的想法时会出现很多困难：有些处理可能具有完全不同的性质，例如，手术和药物治疗；两种药物可能具有不同的剂型，并且尽管通过使用胶囊可以使它们无法被区分，但改变剂型可能会改变药代动力学和/或药效学的特性，因此需要建立制剂的生物等效性；两种处理的每日用法可能不同。在这些情况下，

使用“双模拟”（见词汇表）技术是实现双盲条件的一种方法。这种技术有时会强制实施一种非同寻常的使用方案，对受试者的积极性和依从性产生负面影响。伦理上的困难也会干扰它的应用，例如，当需要模拟手术过程时。无论如何，应当努力克服这些困难。

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.

某些临床试验的双盲性质可能由于明显的治疗诱导效应而遭到部分破坏。在这种情况下，通过使研究者和有关申办者对某些检验结果（如所选择的临床实验室测量）保持盲态，可以使盲法得到改善。在独特或特定的治疗效应可能导致单个受试者破盲的试验中，应当考虑在开放试验中使偏倚最小化的类似方法（见下文）。

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.

如果双盲试验不可行，则应考虑用单盲方案。在有些情况下，只有开放试验在实践上或伦理上是可能的。单盲和开放试验提供了额外的灵活性，但特别重要的是，研究者知道下一个受试者的治疗不应影响纳入受试者的决定；这个决定应该在知道随机化治疗之前。对这些试验，应考虑使用中央随机化方法，如用电话随机化，以管理随机化处理的分配。此外，应该由不参与治疗受试者并对处理保持盲态的医务人员进行临床评价。在单盲或开放试验中，应尽一切努力使各种已知的偏倚来源降到最低，并且主要指标应尽可能客观。应该在方案中解释所采用

的盲态程度的原因，以及通过其它方法使偏倚最小化的步骤。例如，申办者应当有严格的标准操作程序，以保证在清理数据库以供分析之前，适当限制对治疗编码的获取。

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.

只有当受试者的医师认为知道治疗分配对受试者的护理是必须时，才应考虑破盲（针对单个受试者）。任何有意或无意地破盲，无论它发生的原因，在试验结束时都应该被报告并给予解释。揭示治疗分配的程序及时间都应该记录在案。

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.

在本文件中，数据的盲态审核（见词汇表）是指在试验完成（对最后一位受试者的最后一次观察）到揭盲之间的这段时间内对数据的检查。

***2.3.2 Randomisation***

**2.3.2 随机化**

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.

在临床试验中，随机化将一个慎重的机会元素引入到受试者的治疗分配中。在试验数据的后续分析期间，它为定量评价与处理效应有关的证据提供了合理的统计基础。它倾向于使各处理组的预后因素（已知和未知的）分布相似。与盲法结合，在受试者的选择和分配时，随机化有助于避免因治疗分配的可预测性而可能出现的偏倚。

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).

临床试验的随机化列表记录了随机分配给受试者的治疗。在最简单的情况下，它是治疗的序列表（或者交叉试验中的治疗序列），或者是按受试者编号对应的编码。有些试验的物流，如具有筛选阶段的试验，可能使问题变得更加复杂，但是对于受试者其唯一的预先计划的治疗分配或者治疗序列应当是清楚的。不同的试验设计将需要不同的程序来生成随机化列表。随机化列表应当有重现性（如果需要）。

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.)

虽然无限制条件的随机化是一种可接受的方法，但在区组中随机受试者一般可获得某些优势。这有助于增加处理组间的可比性，特别是当受试者特征可能随时间变化时，例如，由于招募策略的改变。它还能更好地保证各处理组的样本量几乎相等。在交叉试验中，它提供了获得具有更高效率和更易于解释的平衡设计的方法。在选择区组长度时需注意，既要足够短以限制可能的不平衡，但又要足够长以避免对区组序列末尾的可预测性。研究者及其他有关人员通常应对区组长度保持盲态；使用两种或以上的区组长度，对每个区组随机选择长度，可达到同样目的。（理论上，在双盲试验中，可预测性并不重要，但药物的药理作用可能给聪明的猜测提供了机会。）

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.

在多中心试验（见词汇表）中，应按中心组织随机化程序。提倡为每个中心建立一个单独的随机方案，即，按中心分层，或为每个中心分配几个整段的区组。更一般地，按照基线测量的重要预后因素（如疾病的严重程度、年龄、性别等）进行分层，有时可能对促进层内的平衡分配是有价值的；这在小型试验中有更大的潜在获益。很少需要使用两个或三个以上的分层因素，这在实现平衡方面不太容易，而且很麻烦。应用动态分配程序（见下文）可能有助于同时在多个分层因素之间达到平衡，只要可以调整其余试验流程以适应这种类型的方法。应当在后续的分析中对分层随机化的因素加以考虑。

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.

被随机化进入试验的下一个受试者，在适当的随机化列表中（如果随机化是分层的，则在相应的层中），应该总是接受与下一个自由号码对应的治疗。只有当已经确认下一个受试者进入到试验的随机化部分时，才应该为该受试者分配合适的号码和相关治疗。增加可预测性的（如，区组长度）随机化细节不应包含在试验方案中。随机化列表本身应该由申办者或独立方安全存档，以一种能够确保在整个试验过程正确地保持盲态的方式。在试验期间获取随机化列表应该考虑在紧急情况下为任何受试者破盲的可能性。破盲应遵循的程序、必要的文件以及受试者后续的治疗和评价均应在方案中写明。

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.

动态分配是一种备选程序，在该程序中，分配给受试者的治疗受到已分配的治疗的当前平衡情况的影响，并且在分层试验中，受到受试者所属的层以及该层内平衡情况的影响。应当避免确定性的动态分配程序，应当为每个治疗分配纳入适当的随机化要素。应尽一切努力保持试验的双盲状态。例如，可能仅限于中央试验办公室知道治疗编码，该办公室一般通过电话联系来控制动态分配。这反过来又允许对入选标准进行额外检查，并建立进入试验的记录，这些特征在某些类型的多中心试验中可能有价值。那么可以遵循用于双盲试验的预包装和贴标签的药品供应的常用系统，但它们的使用顺序不再是依次的。最好是使用适当的计算机算法使中央试验办公室的人员对治疗编码保持盲态。当考虑动态分配时，应该仔细评价物流的复杂性以及对分析的潜在影响。

## III. TRIAL DESIGN CONSIDERATIONS

## 3. 试验设计的考虑

### 3.1 Design Configuration

### 3.1 设计类型

#### *3.1.1 Parallel Group Design*

#### 3.1.1 平行组设计

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).

对于确证性试验，最常见的临床试验设计是平行组设计，在该设计中受试者被随机分到两个或多个组中的一组，每组被分配不同的处理。这些处理将包括具有一个或多个剂量的研究产品，以及一个或多个对照处理，如安慰剂或/和阳性对照。该设计的假设并不比大多数其它设计复杂。然而，与其它设计一样，可能

会有使分析和解释复杂化的额外的试验特征（如，协变量、随着时间的重复测量、设计因素之间的交互作用、方案违背、脱落（见词汇表）及退出）。

***3.1.2 Crossover Design***

### 3.1.2 交叉设计

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.

在交叉设计中，每个受试者被随机分到一个包含两个或多个处理的序列中，因此对于处理的比较他作为他自身的对照。这种简单策略之所以具有吸引力，主要是因为它减少了受试者的数量，以及通常减少了实现特定把握度所需的评价数量，有时到了显著的程度。在最简单的 2×2 交叉设计中，在通常由洗脱期分开的两个连续的处理周期中，每个受试者以随机顺序接受两个处理中的每一个。最常见的扩展是在 n 个周期需要比较 n（>2）个处理，每个受试者接受所有 n 个处理。这类设计存在许多变化，例如每个受试者接受 n（>2）个处理中的一个子集的设计，或者对一个受试者重复给予处理的设计。

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.

交叉设计有很多问题可致它们的结果错误。主要困难在于延滞，即在后继处理周期内的前序处理的残余影响。在一个相加模型中，不相等的延滞效应将使处理间的直接比较产生偏倚。在 2×2 设计中，不能从统计上将延滞效应从处理与周期的交互作用中区分出来，并且因为相应的对比是“受试者之间”，故检验这两个效应中任何一个都缺乏把握度。这一问题在高阶设计中不严重，但不能完全消除。

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.

因此，在使用交叉设计时，重要的是要避免延滞。最好的方法是在充分了解疾病领域和新药的基础上有选择地和谨慎地使用本设计。所研究的疾病应当是慢性的，且病情稳定。治疗期间应充分发挥药物的相关效应。洗脱期应该足够长，以使药物效应完全消退。应该在试验前利用已有信息及数据确定是否可能满足这些条件。

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.

在交叉试验中，有一些额外的问题需要密切注意。其中最值得注意的是，受试者丢失导致的分析和解释的复杂化。另外，延滞的潜在作用导致在后面的处理周期中发生的不良事件很难被对应到适当的处理。这些问题以及其它问题，在ICH E4中已有描述。交叉设计一般应限于在试验中预期仅有少数受试者丢失的情形。

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.

2×2交叉设计的一个常用且通常令人满意的用途是，用以验证相同药物的两种制剂的生物等效性。在健康志愿者的这一特定应用中，如果两个周期之间的洗脱时间足够长，最不可能发生对相关的药代动力学指标的延滞效应。然而，在分析期间基于获得的数据核实这一假设仍然是很重要的，例如，通过证明在每个周期开始时未检测到药物。

***3.1.3 Factorial Designs***

### 3.1.3 析因设计

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.

在析因设计中，通过使用不同的处理组合，可以同时评价两个或多个处理。最简单的例子是2×2析因设计，在该设计中受试者被随机分配到两个处理A和B的四种可能组合的其中之一。四种可能组合为单独A、单独B、既有A又有B以及既无A又无B。在很多情况下，该设计用于检查A和B的交互作用的特定目的。如果样本量计算是基于检验主要效应，则交互作用的统计检验可能会缺乏发现交互作用的把握度。当该设计被用于检查A和B的联合效应时，特别是如果两者可能被一起使用，这种考虑很重要。

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).

析因设计的另一个重要用途是，建立同时使用处理C和D时的剂量-反应特征，特别是当在先前试验中每种单一疗法的有效性已经在某个剂量上被确定时。选择C的剂量数m，通常包括零剂量（安慰剂），和相似的D的剂量数n。那么整个设计由m×n个处理组构成，每个处理组接受一种不同的C和D的剂量组合。那么响应面的结果估计可用于帮助确定临床使用的C和D剂量的适当组合（见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.

在某些情况下，2×2设计评价两种处理的有效性所用的受试者数量与单独评价任一种处理的有效性所需的受试者数量相同，因此可用于高效使用临床试验受

试者。对于死亡率非常高的试验，这一策略已经被证实是非常有价值的。这一方法的效率和效度取决于处理A和B之间不存在交互作用以至于A和B对主要有效性指标的效应遵循相加模型，因此，无论是否追加B的效应，A的效应实际是一样的。对于交叉试验，应当在试验之前利用先前的信息和数据确立很可能会满足这种条件的证据。

**3.2 Multicentre Trials**

**3.2 多中心试验**

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.

开展多中心试验主要有两个原因。首先，多中心试验是更加高效地评价新药的一个可接受的方法；在某些情况下，它可能是唯一可行的方法，可在合理的时间框架内获得足够的受试者以满足试验目的。原则上，可在临床研发的任何阶段开展这种性质的多中心试验。它们可能有几个中心，每个中心都大量的受试者；或者，就罕见病来说，它们可能有很多中心，每个中心只有很少的受试者。

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.

第二，可把试验设计成多中心（和多个研究者）试验，主要是为随后推广它的研究发现提供更好的基础。这源于从更广泛的人群中招募受试者并在更广泛的临床环境中施用药物的可能性，从而呈现出更典型的未来使用的实验情境。在这种情况下，许多研究者的参与也可提供关于药物价值的更广泛的临床判断的潜力。这样的试验在药物研发后期将成为确证性试验，可能涉及大量的研究者和中心。为了更进一步增进可推论性（见词汇表），有时会在许多不同国家实施。

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.

如果要对多中心试验进行有意义的解释和推断，那么在所有中心实施该方案的方式应该是明确的和相似的。此外，通常的样本量和把握度计算取决于各个中心被比较的处理之间的差异是相同数量的无偏估计的假设。心怀这样的背景设计共用方案并开展试验是很重要的。应该尽可能完全地对程序进行标准化。通过研究者会议，通过试验前的人员培训，以及通过试验期间的严密监查，可以减少评价标准和方案的变异。良好的设计一般应旨在实现每个中心内各处理组的受试者分布相同，而且良好的管理应维持这一设计目的。如果后期发现有必要考虑中心之间处理效应的异质性，则防止每个中心受试者数量过度变异的试验以及避开一些非常小的中心的试验具有优势，因为它们减小了不同加权估计的处理效应之间的差异。（这一点并不适用于所有中心都非常小的试验以及在分析时中心不发挥主要作用的试验。）未能采取这些预防措施，再加上对结果同质性的怀疑，在严重的情况下，可能将多中心试验的价值降低到不能为申办者的主张提供令人信服的证据的程度。

In the simplest multicenter trial, each investigator will be responsible for the subjects recruited at one hospital, so that 'center' 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 (sub investigators) 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 center 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 centers 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 centers 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 centers 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.

在最简单的多中心试验中，每一个研究者将负责在一家医院招募受试者，所以，“中心”是由研究者或医院唯一确定的。然而，在很多试验中，情况要更加复杂。一个研究者可能从几家医院招募受试者；一个研究者可能代表一个临床医生团队（下属的研究者），他们都从一家或几家相关的医院里自己的诊所中招募受试者。无论何时对统计模型中关于中心的定义有疑问，方案中的统计章节（见第5.1 章节）都应当在特定试验的背景下明确地定义该术语（例如，按研究者、场所或地区）。在大多数情况下，通过研究者可以满意地定义中心，在这方面 ICH E6 提供了相关指南。在有疑问的情况下，则目标应该是定义中心，以便使对影响主要指标测量及处理效应的重要因素实现同质性。任何将中心合并起来进行分析的规则均应当在可能的情况下在方案中得到合理的解释并预先规定，但在任何情况下，应用此方法做决定都应该始终对处理保持盲态，例如，在盲态审核时。

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 center differences, but does not include a term for treatment-by-center interaction. If the treatment effect is homogeneous across centers, 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.

在方案中应该描述用于估计和检验处理效应的统计模型。主要处理效应可首先使用允许中心之间差异的模型进行研究，但不包含中心与处理间的交互作用项。如不同中心之间处理效应是同质的，则在模型中常规地包含交互作用项将降低对主要效应的检验效率。当存在处理效应的真正异质性时，对处理效应的解释是有争议的。

In some trials, for example some large mortality trials with very few subjects per center, there may be no reason to expect the centers 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 recognized from the start that the limited numbers of subjects per center will make it impracticable to include the center effects in the statistical model. In these cases it is not appropriate to include a term for center in the model, and it is not necessary to stratify the

randomization by center in this situation.

在一些试验中，例如一些大型死亡率试验，每个中心只有很少受试者，没有理由期望中心对主要或次要指标有任何影响，因为它们不可能代表临床重要性的影响。在其它试验中，可能从一开始就意识到，每个中心有限数量的受试者将使在统计模型中包含中心效应变得不切实际。在这些情况下，模型中包含中心项是不合适的，而且，在这种情况下也没有必要按中心进行分层随机化。

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

如果在一个试验中发现了阳性处理效应，而每个中心有相当数量的受试者，则通常应该探索不同中心之间处理效应的异质性，因为这可能影响结论的可推论性。通过对单个中心结果的图示，或通过分析方法，如中心与处理间交互作用的显著性检验，可确定显著的异质性。当使用这种统计显著性检验时，重要的是要认识到，在为检出处理的主要效应而设计的试验中，这种检验通常具有低把握度。

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 centers, in order to substantiate the robustness of the estimates of treatment effect. It is even more important to understand the basis of any heterogeneity characterized 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.

如果发现处理效应的异质性，则应当谨慎地加以解释，并应积极尝试从试验管理的其他特征或受试者特征方面来寻找原因。这样的原因通常会提示适当的进一步分析和解释。在缺乏原因的情况下，处理效应的异质性被证实，例如，通过显著的定量交互作用（见词汇表），意味着可能需要处理效应的替代估计，给中心赋予不同权重，以证实处理效应估计的稳健性。理解任何以显著的定性交互作用（见词汇表）为特征的异质性的基础甚至更为重要，而未能找到原因时，要想可靠地预测处理效应，可能需要进一步开展临床试验。

Up to this point the discussion of multicenter 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 center and treatment-by-center effects to be random, and are especially relevant when the number of sites is large.

到目前为止，针对多中心试验的讨论是基于固定效应模型的使用。混合模型也可用于探索处理效应的异质性。混合模型把中心效应和中心与处理间的交互效应看作是随机的，且当现场数量特别多时，该模型尤其适用。

**3.3 Type of Comparison**

**3.3 比较的类型**

***3.3.1 Trials to Show Superiority***

**3.3.1 显示优效性的试验**

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.

从科学上讲，通过在安慰剂-对照试验中显示优于安慰剂，通过显示优于阳性对照处理，或通过显示剂量-反应关系，所证实有效性是最可信的。这种类型的试验被称为"优效性"试验（见词汇表）。除非另有明确说明，一般情况下，本指南中都假定为优效性试验。

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.

对于严重疾病来说，当存在通过优效性试验已经显示出有效的治疗处理时，安慰剂-对照试验可能被认为是有悖伦理的。在这种情况下，应当考虑科学合理地使用阳性处理作为对照。安慰剂对照和阳性对照相比的适用性应当基于各个试验的情况考虑。

***3.3.2 Trials to Show Equivalence or Non-inferiority***

**3.3.2 显示等效性或非劣效性的试验**

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).

在某些情况下，将研究产品与参照处理相比的目的并不是为了显示优效性。这种类型的试验根据它的目的分为两大类，一类是"等效性"试验（见词汇表），

另一类是“非劣效性”试验（见词汇表）。

Bioequivalence trials fall into the former category. In some situations, clinical equivalence trials are also undertaken for other regulatory reasons such as demonstrating the clinical equivalence of a generic product to the marketed product when the compound is not absorbed and therefore not present in the blood stream.

生物等效性试验属于前一类。在某些情况下，出于其他监管原因也进行临床等效性试验，例如，当化合物不被吸收并因此不存在于血液中时，证明仿制产品与已上市产品的临床等效性。

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.

很多阳性对照试验被设计成显示研究产品有效性不差于阳性对照药，因此属于后一类。另一种可能是在试验中将研究药品的多个剂量与推荐剂量或与标准药品的多个剂量进行比较。这种设计的目的是同时显示研究产品的剂量-反应关系，并将研究产品与阳性对照进行比较。

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.

阳性对照等效性或非劣效性试验也可纳入安慰剂，从而在一个试验中寻求多个目标；例如，它们可以建立与安慰剂相比的优效性，从而验证试验设计，并同时评价相对于阳性对照的有效性与安全性的相似程度。众所周知，不包含安慰剂或的或不使用新药多剂量的阳性对照等效性（或非劣效性）试验，在使用时存在困难。这些与隐性缺乏内部效度的任何测量（与优效性试验相比）有关，因此必

须进行外部验证。等效性（或非劣效性）试验在本质上不是保守的，因此在试验设计或实施中的许多缺陷倾向于使结果偏向等效性的结论。由于这些原因，这些试验的设计特点应受到特别关注，它们的实施需要特别小心。例如，特别重要的是尽量减少违反入选标准、不依从、退出、失访、数据缺失和其它偏离方案的发生率，并尽量减少它们对后续分析的影响。

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.

阳性对照应谨慎选择。一个合适的阳性对照的例子将会是一个广泛使用的疗法，它在相关适应症中的有效性已经在精心设计和良好记录的优效性试验中明确地确定和量化，以及在周密计划的阳性对照试验中它能可靠地预期显示出类似的有效性。为此，考虑到与新试验相关的医学或统计实践的进展，新试验应该与以前实施的且明确地显示出了临床相关有效性的优效性试验具有相同的重要设计特征（主要指标、阳性对照的剂量、资格标准等）。

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.

至关重要的是，设计用于证实等效性或非劣效性的试验方案应清楚地表明该试验的明确意图。在方案中，应规定一个等效界值；这个界值是被判断为在临床上可接受的最大差异，并且应当小于在阳性对照优效性试验中所观察到的差异。对于阳性对照等效性试验，需规定等效界值的上限和下限，但对于阳性对照非劣效性试验，仅需规定界值下限。等效界值的选择应该在临床上是合理的。

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 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).

统计分析通常是基于置信区间的使用（见第 5.5 章节）。对于等效性试验，应当使用双侧置信区间。当置信区间完全落在等效界值之内，则推断为等效。在操作上，这等价于同时使用两个单侧检验的方法，以检验处理之间的差异在等效界值之外的（复合）无效假设与在等效界值之内的（复合）备择假设的比较。因为这两个无效假设是不相交的，所以 I 类错误可被适当地控制。一个单侧假设检验可对应于检验处理之间的差异（研究产品减去对照）等于等效界值下限的无效假设与处理之间的差异大于等效界值下限的备择假设的比较，对于含有该假设检验的方法，I 类错误的选择应当与使用单侧或双侧检验分开来考虑。样本量的计算应当基于这些方法（见第 3.5 章节）。

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.

当无效假设为研究产品与阳性对照之间无差异时，基于观察到的该假设的非显著性检验结果而做出等效性或非劣效性结论是不合适的。

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).

在选择分析集时也存在一些特殊问题。从处理组或者对照组退出或脱落的受试者都将倾向于缺乏应答，因此使用全分析集（见词汇表）的结果可能偏向于证实等效性（见第 5.2.3 章节）。

### *3.3.3 Trials to Show Dose-response Relationship*

### 3.3.3 显示剂量-反应关系的试验

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.

反应与新研究产品的剂量如何相关，是一个在研发的所有阶段通过各种方法都可获得答案的问题（见 ICH E4）。剂量反应试验可服务于许多目的，其中以下特别重要：有效性的确证；剂量反应曲线的形态和位置的研究；适宜的初始剂量的估计；个体剂量调整的最优策略的确定；超量不可能出现额外获益的最大剂量的确定。应该使用在研究中收集到的多个剂量的数据，包括适当情况下的安慰剂（零剂量）来解决这些目的。为此，应用估计剂量与反应之间关系的程序，包括构建置信区间以及使用图形法，与使用统计检验同样重要。可能需要根据剂量的自然顺序或关于剂量-反应曲线的形状（如单调性）的特定问题来调整所用的假设检验。应当在方案中提供所计划的统计程序的详细信息。

**3.4 Group Sequential Designs**

**3.4 成组序贯设计**

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.

使用成组序贯设计便于进行期中分析（见第 4.5 章节和词汇表）。尽管成组序贯设计不是可用于期中分析的唯一可接受的设计类型，但它们是最常用的，因

为当来自每个受试者的数据变得可用时，在试验期间以周期性间隔评价分组受试者的结局比在连续的基础上评价更加可行。在获得处理结局和受试者的处理分配（如破盲，见第 4.5 章节）的信息之前，应充分说明统计方法。独立数据监查委员会（见词汇表）可用于对来源于成组序贯设计的数据进行审查或者进行期中分析（见第 4.6 章节）。尽管该设计已被最广泛地、成功地应用于大型、长期的死亡率或主要非致死性终点的试验，但在其它情况下它的使用正在增加。特别是，已经认识到在所有试验中都必须监查安全性，因此应当始终考虑需要出于安全原因提早停止试验的正式程序。

**3.5 Sample Size**

**3.5 样本量**

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).

临床试验中受试者数量应该总是足够大，以对所提出的问题提供可靠的答案。这个数量通常由试验的主要目的来确定。如果样本量是在其它基础上确定的，那么这应该是清楚的和合理的。例如，基于安全性问题或需求或者基于重要的次要目的来确定样本量的试验可能要比基于主要有效性问题来确定样本量的试验需要更多的受试者（例如，见 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.

使用确定适宜样本量的通常方法，应该规定以下各项：主要指标、检验统计量、无效假设、所选剂量下的备择（“工作”）假设（体现了在所选受试者人群中在所选剂量下检测出或拒绝处理之间的差异的考虑）、错误拒绝无效假设的概率（I类错误）、错误地不拒绝无效假设的概率（II类错误），以及应对退出处理和违

背方案的方法。在有些情况下，事件率是估计把握度的主要兴趣，并且为了从所需的事件数量推算出试验的最终样本量，应当做出一些假设。

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.

应该在方案中给出计算样本量的方法，以及在计算中使用的任何量值的估计（如方差、均值、反应率、事件率、待检测的差异）。也应该给出这些估计的依据。重要的是，要研究样本量估计对这些假设的各种偏离的敏感性，通过提供适合于偏离假设的合理范围的样本量范围，可以有助于这样的研究。在确证性研究中，假设通常应基于公开发表的数据或早期试验的结果。待检测的处理之间的差异，可依据在患者管理中对具有临床相关性的最小效应的判断，或者依据对新处理的预期效应的判断，相比之下在依据对新处理的预期效应时更大。通常 I 类错误概率设在 5%或者更小，或者由多重比较考虑所需要的任何调整来决定；所检验的假设的先验合理性以及这些结果的预期效果可能会影响精确选择。II 类错误的概率通常设在 10%到 20%；保持这个数字低到可行的程度符合申办者的利益，特别是在试验很困难或不可能重复的情况下。在某些情况下，I 类和 II 类错误的常规水平的替代值可能是可接受的，甚至是优选的。

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.

样本量计算应参考主要分析所需的受试者数量。如果这是“全分析集”，与符合方案集（见词汇表）相比，可能需要减少对效应量的估计。这是为了稀释纳入退出处理的或者依从性差的受试者的数据所产生的处理效应。关于变异性的假设可能也需要修改。

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.

等效性或非劣效性试验（见第 3.3.2 章节）的样本量通常应基于获得处理之间的差异的置信区间的目的，该差异表明处理之间最多相差到临床上可接受的差异。当在真实差异为 0 的条件下评价等效性试验的把握度时，如果真实差异不为 0，则低估达到这一把握度所需的样本量。当在差异为 0 的条件下评价非劣效性试验的把握度时，如果研究产品的效应低于阳性对照，则低估达到这一把握度所需的样本量。选择“临床上可接受的”差异需要说明它对将来患者的意义的合理性，并且可能小于在上文提到的设计用于确定存在差异的优效性试验的背景下的“临床上有关的”差异。

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.

在成组序贯试验中不能预先确定确切的样本量，因为它依赖于机会作用、所选择的停止试验的准则以及真实的处理之间的差异。设计停止试验的准则应该考虑后续样本量的分布，通常在期望的和最大的样本量中体现。

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).

当事件率低于预期或变异性大于期望时，在不揭盲数据或不进行处理之间比较的情况下，可使用样本量重新估算的方法（见第 4.4 章节）。

**3.6 Data Capture and Processing**

**3.6 数据收集及处理**

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'.

收集数据并从研究者向申办者传送数据可通过各种媒介进行，包括纸质病例报告表、远程现场监查系统、医疗计算机系统和电子传输。无论采用何种数据收集工具，所收集的信息的形式和内容都应完全符合方案，并应在临床试验实施前确定。应当集中在对实施所计划的分析有必要的数据上，包括确认方案依从性或确定重要方违背所需要的背景信息（如与服用剂量有关的定时评价）。"缺失值"应该与"0 值"或"特征缺失"区分开来。

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.

从数据收集到数据库最终确定的过程应该按照 GCP 进行（见 ICH E6，第 5 章节）。具体来说，为了确保通过实施所计划的分析来交付高质量的数据库并实现试验目的，需要及时可靠的过程，用于记录数据和纠正错误与遗漏。

**IV. TRIAL CONDUCT CONSIDERATIONS**

**4. 试验实施的考虑**

**4.1 Trial Monitoring and Interim Analysis**

**4.1 试验监查和期中分析**

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.

按照方案认真实施临床试验，对结果的可靠性具有重大影响（见 ICH E6）。仔细监控可以确保尽早发现困难，并将它们的发生和再发减至最小。

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.

由制药企业资助的确证性临床试验，通常的特征是有两种截然不同的监查类型。一种类型的监查是关注试验质量的监督，另一种涉及破盲以进行处理之间的比较（即期中分析）。这两种类型的试验监查，除需要不同的人员职责之外，还涉及获取不同类型的试验数据和信息，因此需用不同的原则控制潜在的统计和操作偏倚。

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.

出于监督试验质量的目的，试验监查中所涉及的检查可能包括，是否正在遵循方案，所收集数据是否可接受，计划的收集目标是否成功，设计假设是否合适，以及在试验中保留患者是否成功，等（见第 4.2 至 4.4 章节）。这种类型的监查既不需要获取比较处理效应的信息，也不需要对数据进行揭盲，因此对 I 类错误没有影响。出于这一目的对试验进行监查是申办者的职责（见 ICH E6），可由申办者或申办者选择的独立小组来进行。这种类型的监查周期一般是从选择试验现场开始，到收集和清理最后一位受试者的数据结束。

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.

其他类型的试验监查（期中分析）涉及到收集可比较的处理结果。期中分析需要揭盲（即破盲）获取处理组分配（实际的处理分配或者各组分配的标识）以及可比较的处理组的总结信息。这需要在方案（或者首次分析之前的适当修订）中包含期中分析的统计计划，以防止某些类型的偏倚。这将在第 4.5 和 4.6 章节

中讨论。

**4.2 Changes in Inclusion and Exclusion Criteria**

**4.2 纳入与排除标准的更改**

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.

纳入与排除标准应当保持恒定，按照方案中的规定，贯穿受试者招募期。偶尔有些改变也是合适的，例如，在长期试验中，从试验外部或是期中分析所获取的医学知识的更新，可能提示入组标准的改变。通过监查人员发现经常发生违反入组标准的情况，或者由于过度限制的标准导致严重的低招募率，也都可能引起改变。应该在不破盲的情况下进行改变，并应始终通过方案修订来描述改变，它应该涵盖任何统计后果，如不同事件率所致的样本量调整，或者对计划分析的修改，如根据修改的纳入/排除标准进行分层分析。

**4.3 Accrual Rates**

**4.3 入组率**

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.

在长时间入组受试者的试验中，应该监查入组率，如果它明显低于预计水平，应该查明原因并采取补救措施，以确保试验的把握度并减轻对选择性入组和其他质量方面的担忧。在多中心试验中，这些考虑适用于各个中心。

**4.4 Sample Size Adjustment**

**4.4 样本量调整**

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).

在长期试验中，通常有机会对原设计和样本量计算所依据的假设进行检查。如果是根据初步的和/或不确定的信息做出的试验规定，这可能是特别重要的。对盲态数据进行期中检查可能会发现总体反应方差、事件率或生存体验不如预期。那么可以使用适当修改的假设来计算修改的样本量，并且应在方案修订和临床研究报告中说明它的合理性并记录在案。应该解释为保持盲态所采取的措施以及对于 I 类错误和置信区间宽度的后果（如果有）。只要可能，应在方案中设想样本量重新估计的潜在需要（见 3.5 章节）。

**4.5 Interim Analysis and Early Stopping**

**4.5 期中分析和提早停止试验**

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.

期中分析是指，在试验正式完成之前的任何时间，为了比较处理组间的有效性或安全性而进行的任何分析。因为这些比较的次数、方法及结果影响试验的解释，因此所有期中分析都应当预先仔细计划并在方案中阐明。一些特殊情况可能需要开展在试验开始时未定义的期中分析。在这些情况下，在揭盲获取处理比较数据之前，应完成描述期中分析的方案修订。当期中分析被计划用于决定是否停止试验时，这通常是通过使用把统计监查方案作为指南的成组序贯设计来实现的（见第3.4章节）。这种期中分析的目标是，如果研究中处理的优效性已被清楚

地确定，如果已经不可能证实相关处理之间的差异，或者如果有明显的不可接受的不良反应，则提早停止试验。一般来说，与监查安全性的界限相比，监查有效性的界限需要更多的证据来提早停止试验（即，它们更保守）。当试验设计或监查目的涉及多个终点时，那么还需要考虑这方面的多重性。

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.

方案中应当描述期中分析的时间表，或至少描述一些控制它生成的考虑，例如，是否使用灵活的 α 消耗函数方法；在第一次期中分析的时间之前，可在方案修订中提供进一步的细节。在方案或修订中应该清楚地描述停止试验的指南及它们的特性。也应当考虑提早停止对其他重要指标分析的潜在影响。当试验有数据监查委员会时，这个材料应该由他们撰写或批准（见第 4.6 章节）。偏离计划程序总是有可能使试验结果错误。如果需要对试验做出改变，统计程序的任何相应改变应尽早在方案修订中详细说明，特别是讨论这些改变对于任何分析或推断的影响。所选程序应该始终确保 I 类错误的总体概率得到控制。

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.

执行期中分析应该是一个完全保密的过程，因为可能涉及揭盲的数据和结果。参与实施试验的所有人员应当对这些分析的结果保持盲态，因为他们对试验的态度可能会被改变且导致被招募患者的特征改变或者在处理比较中的偏倚。除了直

接参与执行期中分析的人员之外，这一原则可适用于所有研究人员和申办者所雇佣的人员。研究者应仅被告知继续或停止试验的决定或是实施修订试验程序的决定。

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.

大部分支持研究产品有效性和安全性的临床试验应该进展到全部完成计划的样本量收集；只有出于伦理原因，或者如果把握度不再是可接受的，应当提早停止试验。然而，人们认识到，药物研发计划涉及到申办者出于各种原因获取可比较的处理数据的需求，例如计划其他试验。人们还认识到，仅有部分试验将会涉及到研究严重威胁生命的结局或死亡率，出于伦理原因可能需要对收集的可比较的处理效应进行连续监查。在这些情况的任何一种中，为了应对可能引入的潜在的统计偏倚和操作偏倚，应当在揭盲获取可比较的处理数据之前，在方案或方案修订中制定期中统计分析计划。

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.

对于许多研究产品的临床试验，特别是那些具有重大公共卫生意义的临床试验，应将监查有效性和/或安全性结局比较的职责分配给外部独立小组，一般称之为独立数据监查委员会、数据和安全监查委员会或数据监查委员会，应清楚地描述他们的职责。

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.

当申办者承担监查有效性或安全性比较的职责并因此可以获取揭盲的比较信息时，应该特别注意保护试验的完整性并适当地管理和限制信息共享。申办者应当保证内部监查委员会遵守书面标准操作程序，以及保证保存含有期中结果记录的决策会议的纪要，并记录在案。

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.

任何没有适当计划的期中分析（有或没有提早停止试验的结果）都可能使试验结果有缺陷，并可能削弱对所得结论的信心。因此，应该避免这些分析。如果实施非计划的期中分析，临床研究报告应该解释为什么它是有必要的，破盲的程度，提供对引入偏倚的潜在程度的评价，以及对结果解释的影响。

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

**4.6 独立数据监查委员会的作用（见 ICH E6 第 1.25 和 5.52 章节）**

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.

独立数据监查委员会可由申办者组建，每隔一段时间评价临床试验进展、安全性数据和关键有效性指标，并向申办者建议继续、修改或停止试验。该委员会应当有书面操作程序，并保存所有会议记录，包括期中结果；当试验完成时，这些应可供审查。该委员会的独立性旨在控制重要的可比较信息的共享，以及防止临床试验的完整性受到因获取试验信息而造成的不利影响。该委员会是独立于机构审查委员会或独立伦理委员会的实体，它的组成应包括通晓统计等适当学科的临床试验科学家。

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.

当独立数据监查委员会中有申办者代表时，在委员会的操作程序中他们的作用应被明确规定（例如，涉及他们是否能就关键问题进行投票）。由于这些申办者代表将会获得揭盲信息，因此这些操作程序还应解决如何控制期中试验结果在申办者组织内散布。

## V. DATA ANALYSIS CONSIDERATIONS

## 5. 数据分析的思考

### 5.1 Prespecification of the Analysis

### 5.1 分析的预先确定

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.

当设计临床试验时，数据的最终统计分析的主要特征应该在方案的统计章节中进行描述。该章节应包括所提出的主要指标的确证性分析的所有主要特征以及解决预期分析问题的方法。在探索性试验的情况下，该章节可描述更一般性的原则和方向。

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.

统计分析计划（见词汇表）可作为独立文件撰写，可在最终确定方案之后完成。在该文件中，可以更加技术性地和详细地阐述在方案中所述的主要特征（见第 7.1 章节）。该计划可包括对主要和次要指标以及其他数据进行统计分析的详细程序。应审核并按照数据盲态审核的结果可能更新该计划（见第 7.1 章节的定义），但应在破盲之前最终确定。当最终确定统计分析计划时以及当随后破盲时，应保留正式记录。

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.

如果盲态审核建议更改方案中所述的主要特征，则应该在方案修订中记录这些更改。否则，根据盲态审核的建议考虑更新统计分析计划就足够了。只有是在方案（包括修订）中设想的分析，它们的结果才能被认为是确证性的。

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).

在临床研究报告的统计章节中，应该清楚地描述统计方法，包括在临床试验过程中在什么时候做出的方法学决策（见ICH E3）。

**5.2 Analysis Sets**

**5.2 分析集**

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.

受试者集是指受试者的数据被纳入到主要分析中，应在方案的统计章节进行定义。另外，对启动试验程序（例如，导入期）的所有受试者进行文档记录可能是有用的。这个受试者文档的内容取决于特定试验的详细特征，但是，只要可能，至少应收集人口统计学的以及关于疾病状态的基线数据。

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).

如果随机化进入临床试验的所有受试者均满足所有入组标准，完全遵循所有

试验程序且无失访，并提供完整的数据记录，那么将受试者集纳入分析是显而易见的。试验设计和实施的目标应该是尽可能地接近这一理想状态，但在实践中，它是否能完全达到是值得怀疑的。因此，方案的统计章节应该从如何影响被分析的受试者和数据的角度前瞻性地解决预期的问题。方案还应该详细说明旨在尽量减少研究实施中任何预期不规范的程序，这些不规范包括各种类型的方案违背、退出及缺失值，可能会损害令人满意的分析。方案应该考虑降低这些问题发生频率的方法以及解决在数据分析中所发生问题的方法。在盲态审核期间，应该明确说明对应对方案违背的分析方法的可能修订。最好是确定任何重大方案违背的发生时间、原因及对试验结果的影响。方案违背、缺失值以及其它问题的频率和类型应记录在临床试验报告中，并且应当描述它们对试验结果的潜在影响（见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.

关于分析集的决策应该遵循以下原则：（1）使偏倚减到最小；（2）避免I类错误的膨胀。

***5.2.1 Full Analysis Set***

**5.2.1 全分析集**

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.

意向性处理（见词汇表）的原则，是指主要分析应包括所有随机化受试者。遵循这一原则需要对所有随机化受试者进行完整随访以获得研究结局。由于描述的原因，在实践中很难实现这一理想状态。在本文件中，“全分析集”这个术语被用来描述尽可能完整的分析集，以及尽可能接近包括所有随机化受试者的意向性处理的理想状态的分析集。在分析中保持初始随机化对于防止偏倚以及为统计检验提供安全基础是很重要的。在许多临床试验中，全分析集的使用提供了一个保守策略。在许多情况下，它也可以提供处理效应的估计，这些估计更有可能反

映了在随后实践中观察到的效应。

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:

在一些有限的情况下，可能导致将随机化受试者从全分析集中排除，包括未能满足主要入组标准（入选标准违背），未服用过至少一次试验药物以及缺乏随机化后的任何数据。这些排除应该总要说明合理性。只有在以下情况下，未能满足入组标准的受试者可被从分析中排除而不会引入偏倚：

(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.

（1）在随机化之前评判了入组标准；

（2）可以完全客观地检查相关的资格违规；

（3）所有受试者都接受资格违规的同等审查；（在开放试验中或者甚至在双盲试验中，如果在审查之前数据被揭盲，同等审查就很难保证，所以要强调盲态审核的重要性。）

（4）排除所有检查到的特定入组标准的违反。

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.

在某些情况下，从所有随机化受试者集中排除任何未服用试验药物的受试者是合理的。如果，例如，是否开始给予处理的决定不受已知晓所分配的处理的影响，虽然排除这些患者，但意向性处理原则仍将被保持。在其他情况下，可能需要从所有随机化受试者集中排除任何随机化后无数据的受试者。除非来自这些特定排除的或任何其它的潜在偏倚得到解决，否则任何分析都不是完整的。

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.

当使用受试者全分析集时，随机化后发生的方案违背可能会对数据和结论产生影响，特别是如果它们的发生与处理分配相关时。在很多时候，把这些受试者的数据纳入到分析中是合适的，与意向性处理原则相一致。接受一次或多次剂量后从处理中退出并此后不提供数据的受试者，以及否则就失访的受试者导致出现了特殊问题，因为不把这些受试者纳入到全分析集中可能会破坏这个方法。在这种背景下，在受试者因任何原因失访时进行的主要指标测量，或者根据方案中预期的评价时间表随后收集到的主要指标测量，都是有价值的；在主要指标是死亡率或严重发病率的研究中，后续收集尤为重要。应该在方案中描述以这种方法收集数据的意图。从进行末次观察值结转到使用复杂的数学模型，填补技术也可用于尝试补充缺失值。用于确保全分析集中每个受试者的主要指标测量的可用性而采用的其他方法，可能会要求关于受试者结局的一些假设，或者要求对结局的更简单选择（例如，成功或失败）。这些策略的任何使用都应在方案的统计章节中进行描述并说明合理性，并且所用的任何数学模型所依据的假设均应解释清楚。特别是当所述的策略本身可能导致处理效应的有偏估计时，证实相应的分析结果的稳健性也同样重要。

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.

由于一些问题的不可预测性，有时最好把详细考虑应对不规范现象的方式推迟到试验结束对数据进行盲态审核时，如果是这样，则应在方案中加以说明。

***5.2.2 Per Protocol Set***

**5.2.2 符合方案集**

The 'per protocol' set of subjects, sometimes described as the 'valid cases', the 'efficacy' 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:

受试者的“符合方案”集，有时被称为“有效病例”，“有效性”样本或“可评价的受试者”样本，被定义为全分析集的受试者中对方案更具有依从性的子集，并且以符合如下标准为特征：

(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.

（1）完成了对处理方案的某个预先设定的最小暴露量；

（2）可以获得主要指标的测量值；

（3）无任何重大方案违背，包括对入组标准的违反。

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.

在按照适合于特定试验情况的方式破盲之前，应该全面定义并记录将受试者排除在符合方案集之外的确切原因。

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.

使用符合方案集可以使新处理在分析中显示出额外有效性的机会最大化，并且最密切地反映了作为方案基础的科学模型。然而，相应的假设检验和处理效应的估计可能保守也可能不保守，这取决于试验；对研究方案的依从可能与处理和结局有关，由此产生的偏倚可能是严重的。

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.

应该充分识别并总结导致排除受试者以生成符合方案集的问题，以及其它的方案违背。相关的方案违背可能包括处理分配的错误、使用禁忌药物、依从性差、失访和缺失数据。在处理组之间从发生频率和时间方面，评价这些问题的模式，是一种良好做法。

***5.2.3 Roles of the Different Analysis Sets***

**5.2.3 不同分析集的作用**

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.

一般说来，证明主要试验结果对被分析的受试者集的替代选择缺乏敏感性，是有利的。在确证性试验中，计划对全分析集及符合方案集都进行分析，一般来说是恰当的，以便可以明确地讨论和解释它们之间的任何差异。在某些情况下，可能期望进一步探讨被分析的受试者集的选择对结论的敏感性。当全分析集和符合方案集得出实质上相同的结论时，试验结果的可信度增加，但是要记住，从符合方案集中排除相当大比例的受试者的需求，给试验的总体可信性带来了一些疑问。

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.

在优效性试验（试图显示研究产品是更优的）和等效性或非劣效性试验（试图显示研究产品具有可比性，见第 3.3.2 章节）中，全分析集和符合方案集发挥不同的作用。在优效性试验中，全分析集用于主要分析（除了例外情况），因为它倾向于避免符合分析集所导致的对有效性的过度乐观估计，因为包含在全分析

集中的非依从者一般会降低估计的处理效应。然而，在等效性或非劣效性试验中，使用全分析集一般不保守，应非常仔细地考虑它的作用。

**5.3 Missing Values and Outliers**

**5.3 缺失值及离群值**

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.

缺失值是临床试验中的一个潜在偏倚来源。因此，应尽一切努力满足有关数据收集和管理的所有方案要求。然而，实际上，几乎总是会有一些缺失数据。虽然如此，只要应对缺失值的方法是合理的，并且特别是如果在方案中预先定义了这些方法，则试验可以被认为是可信的。通过在盲态审核期间在统计分析计划中更新此方面，可以细化这些方法的定义。遗憾的是，没有普遍适用的应对缺失值的方法可以被推荐。应当研究分析结果对应对缺失值的方法的敏感性，特别是当缺失值的数量很大时。

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

应当采用类似的方法探索离群值的影响，它们的统计定义在某种程度上是主观的。当从医学上和统计上证明是合理的时，把某一特定值明确地确定为异常值才最具有说服力，并且医学背景通常会定义适当的操作。在方案或统计分析计划中设定的任何离群值程序都应当是预先不偏袒任何处理组。同样，在盲态审核期间可以有效地更新这方面的分析。如果在试验方案中未预先规定应对离群值的程序，则应进行一个具有实际值的分析，以及至少进行一个消除或减少离群值效应

的其它分析，并讨论它们结果之间的差异。

**5.4 Data Transformation**

**5.4 数据转换**

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.

最好是在试验设计期间基于早期临床试验的类似数据，做出在分析前对关键指标进行转换的决定。应该在方案中对转换（例如，平方根、对数）进行详细说明，并提供基本原理，特别是对于主要指标。在标准文本中可以找到指导使用转换的一般原则，以确保满足统计方法所依据的假设；在许多特定的临床领域已经形成了针对特定指标的惯例。偏爱便于临床解释的量表应该会影响是否以及如何对指标进行转换的决定。

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.

类似的考虑也适用于其他衍生指标，例如，使用自基线的改变、自基线的百分比改变、重复测量的“曲线下面积”或者两个不同指标的比值。应仔细考虑后续的临床解释，并在方案中说明衍生的合理性。与此密切相关的内容已在第 2.2.2 章节中进行了讨论。

**5.5 Estimation, Confidence Intervals and Hypothesis Testing**

**5.5 估计、置信区间及假设检验**

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.

为了满足试验的主要目的，应该在方案的统计章节中详细说明待检验的假设和/或待估计的处理效应。用于完成这些任务的统计方法应当针对主要指标（以及优选的次要指标）进行描述，并且应当明确所依据的统计模型。只要有可能，处理效应的估计应伴有置信区间，并应确定它们的计算方法。应当说明使用基线数据以提高精度或校正潜在基线差异的估计值的任何打算，例如，通过协方差分析。

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.

重要的是，要阐述清楚是否将使用单侧或双侧的统计显著性检验，特别是要前瞻性地说明使用单侧检验的合理性。如果认为假设检验不合适，那么应该给出得出统计结论的替代过程。关于单侧或双侧推断方法的问题是有争议的，在统计文献中可以找到各种各样的观点。在监管背景下，更可取的方法是将单侧检验的 I 类错误设置为双侧检验中使用的传统 I 类错误的一半。这促进了与双侧置信区间的一致性，双侧置信区间通常适合于估计两种处理之间的差异的可能大小。

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).

所选择的特定统计模型应当反映有关待分析指标的医学和统计知识的现状以及试验的统计设计。应充分说明在分析中待拟合的所有效应（例如在方差模型分析中），并应解释根据初步结果修改这组效应的方式（如果有）。同样的考虑也适用于在协方差分析中所拟合的协变量集合（见第 5.7 章节）。在选择统计方法

时，应注意主要和次要指标的统计分布。当进行这种选择时（例如，在参数和非参数方法之间），重要的是，要记住需要提供处理效应大小的统计估计值及置信区间（除了显著性检验之外）。

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.

应当清楚地区分主要指标的主要分析与主要或次要指标的支持性分析。在方案的统计章节或者统计分析计划中，应当概述总结和报告除主要和次要指标之外的数据的方法。为了在一系列试验中实现分析一致性的目的，例如对于安全数据，应当包括所采用方法的介绍。

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.

关于已知的药理学参数、单个受试者的方案依从程度或其它生物学基础数据，整合这些信息的建模方法可以提供对实际或潜在有效性的有价值的洞察，特别是对于处理效应的估计。应始终清晰地确定这些模型所依据的假设，并仔细描述任何结论的局限性。

**5.6 Adjustment of Significance and Confidence Levels**

**5.6 显著性及置信水准的调整**

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.

当存在多重性时，用于临床试验数据分析的常用的频率论方法可能需要对 I 类错误进行调整。多重性可能来源于，例如，多个主要指标（见第 2.2.2 章节）、

处理的多重比较、随时间的多次评价和/或期中分析（见第 4.5 章节）。在可行的情况下，避免或减少多重性的方法有时候是更可取的，例如确定关键的主要指标（多个指标），选择关键的处理比较（多重比较），使用概括性测量如“曲线下面积”（重复测量）。在确证性分析中，在采取此类步骤之后，剩余的多重性的任何方面也应当在方案中确定；应始终考虑调整，并应在分析计划中列出任何调整程序的细节，或者解释为什么认为没有必要调整。

**5.7 Subgroups, Interactions and Covariates**

**5.7 亚组、交互作用及协变量**

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.

主要指标通常系统性地关联到除处理之外的其它影响。例如，可能存在与诸如年龄和性别等协变量的关系，或者诸如在多中心试验的不同中心接受处理的受试者的特定亚组之间可能存在差异。在有些情况下，对协变量影响的调整或者对亚组效应的调整是分析计划中不可缺少的部分，因此应在方案中列出。应通过试验前仔细考虑，确定这些协变量以及预期对主要指标有重要影响的因素，并应考虑如何在分析中考虑它们，以便提高精度和补偿处理组之间的任何不平衡。如果使用一个或多个因素对设计进行分层，那么在分析中考虑这些因素是合适的。当调整的潜在价值受到怀疑时，通常建议将未调整的分析作为主要关注的分析，把调整的分析作为支持性的。应特别注意中心效应和主要指标基线测量的作用。不建议在主要分析中校正随机化后测量的协变量，因为它们可能受到处理的影响。

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.

处理效应本身也可能随亚组或协变量而变化——例如，效应可能随年龄降低或者可能在特定诊断类别的受试者中更大。在某些情况下，这些交互作用是预期的或者具有特别优先的兴趣（如老年病学），因此，亚组分析或者包含交互作用的统计模型，是计划的确证性分析的一部分。然而，在大多数情况下，亚组分析和交互作用分析是探索性的，并且应当明确地认定为是探索性的；它们应当探索总体上发现的任何处理效应的一致性。一般来说，这些分析应该首先通过向所讨论的统计模型添加交互作用项来进行，辅之以在相关受试者亚组内或者由协变量定义的层内进行额外的探索性分析。当进行探索性分析时，应谨慎解释这些分析；仅仅基于探索性亚组分析的处理有效性（或缺乏有效性）或安全性的任何结论都不太可能被接受。

**5.8 Integrity of Data and Computer Software Validity**

**5.8 数据的完整性与计算机软件的可靠性**

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.

分析的数值结果的可靠程度取决于用于数据管理（数据录入、存储、验证、校正和检索）以及在统计上处理数据的方法和软件（内部和外部编写）的质量和可靠性。因此，数据管理活动应当基于全面和有效的标准操作程序。用于数据管理和统计分析的计算机软件应当是可靠的，并应提供适当的软件测试过程的文件。

**VI. EVALUATION OF SAFETY AND TOLERABILITY**

**6. 安全性与耐受性评价**

**6.1 Scope of Evaluation**

### 6.1 评价的范围

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.

在所有临床试验中，安全性和耐受性（见词汇表）的评价是一个重要方面。在早期阶段，这种评价主要是探索性的，并且只对毒性的直接表达敏感，而在后期阶段，可在更大样本量的受试者中更加全面地描述所建立的药物安全性和耐受性概况。后期阶段的对照试验代表了以无偏的方式探索任何新的潜在不良反应的重要方法，即使这些试验在这方面通常缺乏把握度。

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.

出于对安全性与耐受性的优效性或等效性做出特定主张的目的，可以设计某些试验，与其它药物或与研究药物的其它剂量相比。应当通过来自确认性试验的相关证据支持这些特定主张，类似于对应的有效性主张所需要的证据。

### 6.2 Choice of Variables and Data Collection

### 6.2 指标选择与数据收集

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).

在任何临床试验中，选择用于评价药物安全性和耐受性的方法和测量将取决于许多因素，包括对与药物密切相关的不良反应的了解，来自非临床和早期临床研究的信息以及特定药物的药效/药代动力学特性的可能结果，给药方式，待研究的受试者类型，以及试验持续时间。有关临床化学和血液学、生命体征和临床不良事件（疾病、体征和症状）的实验室检查通常构成安全性和耐受性数据的主

体。发生严重不良事件以及因不良事件导致治疗中止对于注册是特别重要的（见 ICH E2A 和 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).

此外，建议在整个临床试验规划中采用一致的方法来收集和评价数据，以便于合并来自不同试验的数据。使用通用的不良事件词典尤为重要。该词典具有一种结构，提供了在三个不同层级上汇总不良事件数据的可能性，即系统-器官分类、首选术语或收录术语（见词汇表）。首选术语是处在通常汇总不良事件的层级上，在数据的描述性展示中，可以汇集属于同一系统-器官分类的首选术语（见 ICH M1）。

**6.3 Set of Subjects to be Evaluated and Presentation of Data**

**6.3 待评价的受试者集及数据展示**

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).

对于总体安全性和耐受性评价，待汇总的受试者集通常被定义为那些接受至少一个剂量研究药物的受试者。应尽可能全面地从这些受试者中收集安全性和耐受性指标，包括不良事件类型、严重程度、发病和持续时间（见 ICH E2B）。可能需要在特定的亚组人群，如女性、老年人（见 ICH E7）、严重疾病或那些有常见伴随治疗的人们中，进行额外的安全性及耐受性评价。这些评价可能需要解决更加特殊的问题（见 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.

在评价过程中需要注意所有安全性和耐受性指标，并且在方案中应该指出方法。所有不良事件都应报告，无论它们是否被认为与处理有关。在评价中应当考虑研究人群中的所有可用数据。应当谨慎地定义测量值的单位和实验室指标的参考范围，如果在同一试验中出现不同的单位或不同的参考范围（例如，如果涉及一个以上的实验室），则测量值应当被适当标准化，以便进行统一评价。应预先确定毒性分级量表的使用，并说明合理性。

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.

某种不良事件的发生率通常以经历事件的受试者数量与处于危险中的受试者数量之间的比例来表示。然而，如何评价发生率并不总是显而易见的。例如，根据情况，可考虑把暴露的受试者数量或暴露程度（用人年表示）作为分母。无论计算的目的是估计风险还是在处理组之间进行比较，重要的是要在方案中给出定义。如果计划进行长期治疗，并预期有相当比例的退出处理或死亡，则这一点尤其重要。对于这些情况，应考虑生存分析方法，并计算累积不良事件率，以避免低估的危险。

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.

在一些情况下，当存在大量体征和症状的背景噪声时（例如，在精神病试验中），人们在估计不同不良事件的风险时，应该考虑对此进行解释的方法。一种这样的方法是利用“治疗引发事件”（见词汇表）的概念，只有当不良事件出现或相对于处理前基线发生恶化时，才记录它们。

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.

减少背景噪声影响的其他方法，如忽略轻度不良事件，或要求在重复随访时观察到的事件才可计入分子，也是合适的。这些方法应在方案中解释并说明合理性。

**6.4 Statistical Evaluation**

**6.4 统计评价**

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.

安全性与耐受性的研究是一个多维问题。虽然对于任何药物，某些特定不良反应通常可被预计到，并被特别监查，但是可能的不良反应范围非常大，并且总是可能出现新的和不可预见的反应。此外，在违背方案之后经历的不良事件，例如使用被排除的药物，可能引入偏倚。这个背景构成了与药物安全性和耐受性分析评价有关的统计上的困难，并且意味着来自确证性临床试验的结论性信息是一种例外而不是通例。

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.

在大多数试验中，通过将描述性统计方法应用于数据，只要有助于解释，还可辅之以计算置信区间，从而最好地解决安全性和耐受性的影响。利用图形展示处理组之间及受试者之间的不良事件模式也是有价值的。

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.

计算 *P* 值有时是有用的，或者作为辅助手段，评价感兴趣的特殊差异，或者作为“标记”手段，用于大量安全性与耐受性指标以强调值得进一步关注的差异。这对于实验室数据尤其有用，否则可能难以适当地进行汇总。建议对实验室数据既要进行定量分析，例如处理组均数的评价，又要进行定性分析，计算高于或低于某些阈值的计数。

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.

如果使用假设检验，则对多重性进行统计调整以量化 I 类错误是合适的，但是 II 类错误通常更值得关注。当没有多重性调整时，在解释推定的统计显著性发现时应该谨慎。

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.

在大多数试验中，研究者正在试图确定，与阳性对照药物及安慰剂相比，在安全性及耐受性方面未出现临床上不可接受的差异。与有效性的非劣性或等效性评价的情况一样，在这种情况下，使用置信区间比假设检验更可取。通过这种方式，清楚地显示了由于低发生频率而经常引起的相当大的不准确性。

**6.5 Integrated Summary**

**6.5 综合性总结**

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.

在研究产品的开发过程中，特别是在上市申请时，通常在不同试验之间不断地总结药物安全性与耐受性的特性。然而，这个总结的可用性取决于充分的和控制良好的单个试验，具有高数据质量。

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)

药物的总体可用性始终是风险与获益之间的平衡问题，即使风险/获益的评

价通常是在整个临床试验规划的总结中进行，但在单个试验中也可考虑这一观点（见第 7.2.2 章节）。

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

有关安全性与耐受性报告的更多细节，见 ICH E3 第 12 章。

## VII. REPORTING

## 7. 研究报告

### 7.1 Evaluation and Reporting

### 7.1 评价与报告

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.

如引言所述，临床研究报告的结构与内容是 ICH E3 的主题。该 ICH 指南充分地涵盖了适当整合临床及其它资料的统计工作报告。因此，本章节相对简短。

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.

如第 5 章节所述，在试验的计划阶段中，分析的主要特征应在方案中确定。当试验结束，收集数据并进行初步检查时，对也在第 5 章节中所描述的计划分析进行盲态审核是有价值的。这种对处理保持盲态的分析前审核应当包括有关下列方面的决定，例如，从分析集中排除受试者或数据；还可以检查可能的转换，并定义离群值；将在其它近期研究中确定的重要协变量添加到模型中；可以重新考虑使用参数或非参数方法。此时所作的决定应当在报告中加以描述，并且应当与统计专业人员获得处理编码之后做出的决定加以区别，因为盲态下的决定通常会减少产生偏倚的可能性。参与揭盲的期中统计分析的统计专业人员或其他人员不应参与盲态审核或修改统计分析计划。在数据中可能会显示出处理诱导效应，当

这种可能性会削弱盲法时，盲态审核将需要特别谨慎。

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.

在接近或正当盲态审核时，应完成报告内容和表格的许多更详细的方面，以便在实际分析时对它的所有方面都存在完整的计划，包括受试者选择、数据选择与修改、数据汇总与列表、估计与假设检验。一旦完成数据验证，应按照预先设定的计划进行分析，越遵循这些计划，结果的可信度就越高。应特别注意在方案、方案修订以及基于数据盲态审核更新的统计分析计划中所描述的计划分析与实际分析之间的任何差异。应对偏离计划的分析提供详细的解释。

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.

进入试验的所有受试者，无论是否纳入分析，都应在报告中说明。排除在分析之外的所有原因都应被记录，对于纳入全分析集但未纳入符合方案集的受试者，排除在后者之外的原因也应被记录。类似地，对于纳入分析集的所有受试者，所有重要指标的测量应该在所有相关时间点进行说明。

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.

应仔细考虑受试者或数据的所有缺失、退出处理和重要方案违背对主要指标的主要分析的影响。应确定失访、退出处理或严重方案违背的受试者，并对他们进行描述性分析，包括他们缺失的原因及其与处理和结局的关系。

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.

描述性统计是报告不可缺少的部分。合适的表格和/或图形展示应清楚地说明主要和次要指标以及关键预后指标和人口统计学指标的重要特征。应当特别仔细地描述与试验目的有关的主要分析的结果。当报告显著性检验结果时，应当报告精确的 *P* 值（如“P=0.034”）而不是只参考临界值。

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.

尽管临床试验分析的主要目标应当是回答它的主要目的提出的问题，但在揭盲分析过程中，基于观察数据的新问题很可能会出现。后果是，可能需要额外的或许复杂的统计分析。在报告中，应该严格区分这种额外工作与方案中计划的工作。

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.

对于在计划的分析中未被预先定义为协变量但仍然具有某些预后重要性的基线测量，机率作用可能会导致它们在处理组之间无法预料的不平衡。最好的解决办法是，证明用于解释这些不平衡的补充分析得出了与计划分析基本相同的结论。如果不是这种情况，则应讨论不平衡对结论的影响。

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.

一般而言，应少使用计划外分析。当认为处理效应可能根据某个或某些其他因素而变化时，经常进行这样的分析。那么可以尝试确定该效应对其特别有益的受试者亚组。众所周知，计划外亚组分析的过度解释有潜在风险（见第 5.7 章节），

应谨慎避免。虽然如果在受试者亚组中处理未显示出获益或具有不良反应时会出现类似的解释问题，但应该恰当地评价这些可能性并因此予以报告。

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.

最后，应根据临床试验结果的分析、解释及介绍做出统计判断。为此，试验统计专业人员应当是负责临床研究报告的小组成员之一，并核准临床报告。

**7.2 Summarising the Clinical Database**

**7.2 临床数据库的总结**

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.

上市申请需要对所有报告的临床试验的安全性和有效性证据进行全面总结和综合（欧盟的专家报告、美国的综合总结报告、日本的概要）。在适当的时候，这可能伴随着结果的统计组合。

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.

在总结中，有一些特定的统计关注的领域：描述在临床试验规划期间受试人群的人口统计学和临床特征；通过考虑相关（通常是有对照组的）试验的结果并强调它们相互加强或相互矛盾的程度来解决有效性的关键问题；对于其结果有助于上市申请的所有试验，总结从它们的合并数据库中可获得的安全信息，并确定潜在的安全问题。在设计临床规划期间，应认真关注测量值的统一定义和收集，

这将有助于随后一系列试验的解释，特别是如果不同试验之间的测量值可能被合并时。应该选择和使用可记录用药细节、病史和不良事件的通用词典。对主要和次要指标采用通用定义几乎总是有价值的，并且它对 Meta 分析是极其重要的。关键有效性指标的测量方式、相对于随机化/入组的评价时机、方案违背和偏离的应对以及可能的预后因素定义都应该保持一致，除非有合理的理由不这么做。

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.

应当详细描述用于不同试验之间数据合并的任何统计程序。应注意与试验选择有关的偏倚的可能性、试验结果的同质性、以及各种变异来源的适当模型。应探索结论对假设和选择的敏感性。

***7.2.1 Efficacy Data***

**7.2.1 有效性数据**

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.

单个临床试验应该总是大到足以满足它们的目的。通过总结一系列解决基本相同的关键有效性问题的临床试验，也可以获得额外的有价值的信息。为了便于比较，应该以相同的形式，通常是以聚焦于估计值加上置信限的表格和图形，呈现这一系列试验的主要结果。使用 Meta 分析技术来合并这些估计值常常是一个有用的补充，因为它允许对处理效应的大小生成更精确的总体估计，并提供完整而简明的试验结果总结。在一些特殊情况下，Meta 分析方法也可能是通过总体假设检验提供充分的有效性总体证据的最适当方式，或者唯一方式。当用于此目的时，Meta 分析应该有它自己的预期书面方案。

***7.2.2 Safety Data***

### 7.2.2 安全性数据

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.

在总结安全性数据时，重要的是要彻底检查安全性数据库以寻找潜在毒性的任何迹象，并通过寻找相关的支持性观察模式来跟踪任何迹象。合并来自所有暴露于药物的人们的安全数据，能提供重要的信息来源，因为它的较大的样本量能提供发现更罕见的不良事件的最佳机会，并且可能提供估计它们的近似发生率的最佳机会。然而，由于缺乏对照组，难以评价来自该数据库的发生率数据，来自对照试验的数据在克服这种困难方面特别有价值。应合并具有相同对照组（安慰剂或特定阳性对照）的研究的结果，并分开展示每个提供充足数据的对照组的结果。

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.

所有通过数据探索发现的潜在毒性的迹象都应报告。评价这些潜在不良反应的现实情况应考虑到由于多次比较而产生的多重性问题。还应适当地使用生存分析方法进行评价，以探索不良事件的发生率与暴露时间和/或随访时间的潜在关系。应适当地量化已知不良反应相关的风险，以便正确评价风险/获益关系。

**GLOSSARY**

### 词汇表

**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.

**贝叶斯方法（Bayesian Approaches）**

是指为某些参数（例如处理效应）提供后验概率分布的数据分析方法。后验概率分布由该参数的观测数据和先验概率分布衍生而来，被用作统计推断的基础。

**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'.

**偏倚（统计和操作）（Bias (Statistical & Operational)）**

是指与临床试验的设计、实施、分析和结果评价有关的任何因素导致的处理效应估计值偏离它的真实值的系统趋势。由于实施中的偏离所引入的偏倚称为“操作”偏倚。上述其他来源的偏倚称为“统计”偏倚。

**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.

**盲态审核（Blind Review）**

是指在试验完成（最后一位受试者的最后一次观察）到揭盲之间的这段时间内检查和评价数据，目的是最终确定计划的分析。

**Content Validity**

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

**内容效度（Content Validity）**

是指一个指标（如等级量表）测量它应该测量的内容的程度。

**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).

**双模拟（Double-Dummy）**

是指在临床试验中当两种处理不能做到完全相同时，使试验处理仍能保持盲态的一种技术。先准备处理 A（阳性药和不能区分的安慰剂）和处理 B（阳性药和不能区分的安慰剂），然后受试者接受两套处理：A（阳性药）和 B（安慰剂），或者 A（安慰剂）和 B（阳性药）。

**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.

**脱落（Dropout）**

是指临床试验的受试者由于任何原因不能继续按研究方案进行到所要求的最后一次随访。

**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.

**等效性试验（Equivalence Trial）**

是指主要目的为确认两种或多种处理的差别大小在临床上并无重要意义的试验。通常以真实的处理之间的差异落在临床上可接受的等效性界值上下限之间来表明等效性。

**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.

**频率论方法（Frequentist Methods）**

是指在假设重复实现相同实验情境时，用某些结局的发生频率来解释的统计方法，例如显著性检验和置信区间。

**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.

**全分析集（Full Analysis Set）**

是指尽可能接近符合意向性处理原则的理想的受试者集。该数据集是从所有随机化的受试者中以最少的和合理的方法排除受试者后得到的。

**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.

**推论性（Generalisability, Generalisation）**

是指把临床试验的研究发现从参与试验的受试者可靠地推断到更广泛的患者人群和临床环境的程度。

**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.

**全局评价指标（Global Assessment Variable）**

是指将客观指标和研究者对受试者的状态或状态变化的总体印象综合起来所设定的一个单一指标，通常是一个有序分类等级量表。

**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.

**独立数据监查委员会（数据和安全监查委员会、监查委员会、数据监查委员会）（Independent Data Monitoring Committee (IDMC) (Data and Safety Monitoring Board, Monitoring Committee, Data Monitoring Committee)）**

独立数据监查委员会由申办者建立，可用于定期评价临床试验进度、安全性数据以及关键有效性终点，并可向申办者建议是否继续、修改或停止试验。

**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.

**意向性处理原则（Intention-To-Treat Principle）**

是指基于受试者意向的处理（即计划的处理方案）而不是实际给予的处理进行评价的原则，该原则可以对处理策略的效应做出最佳评价。它的结果是，计划分配到每一个处理组的受试者即应作为该组的成员被随访、评价和分析，而无论他们是否依从于所计划的处理过程。

**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.

**交互作用（定性和定量）（Interaction (Qualitative & Quantitative)）**

是指处理间的对比（如研究产品与对照之间的差异）依赖于另一因素（如中心）的情况。定量交互作用是指对比差异的大小在因素的不同水平之间不同；定性交互作用是指对比差异的方向至少在因素的一个水平上不同。

**Inter-Rater Reliability**
The property of yielding equivalent results when used by different raters on different occasions.

**评价者间信度（Inter-Rater Reliability）**

是指不同评价者在不同场合使用时产生同样结果的特性。

**Intra-Rater Reliability**
The property of yielding equivalent results when used by the same rater on different occasions.

**评价者内信度（Intra-Rater Reliability）**

是指同一评价者在不同场合使用时产生同样结果的特性。

**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.

**期中分析（Interim Analysis）**

是指正式完成临床试验前，比较处理组间的有效性或安全性所做的任何分析。

**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.

**Meta 分析（Meta-Analysis）**

是指对同一个问题的两个或多个试验的量化证据进行的规范评价。这常是将不同试验的汇总统计量进行统计合并，但此术语有时也用于对原始数据的合并。

**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.

**多中心试验（Multicentre Trial）**

是指由多个现场的多个研究者按同一个方案进行的临床试验。

**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).

**非劣效性试验（Non-Inferiority Trial）**

是指主要目的为显示研究产品的反应在临床上不劣于对照（阳性药或安慰剂对照）的试验。

**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”.

**首选术语和收录术语（Preferred and Included Terms）**

在分层级医学词典中，例如 MedDRA，收录术语是词典术语的最低层级，以研究者的描述进行编码。首选术语是收录术语的分组层级，通常用于报告发生频率。例如，研究者写的是“左臂疼痛”，收录术语编码为：“关节疼痛”，在首选术语层级上报告为“关节痛”。

**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.

**符合方案集（有效病例，有效性样本，可评价的受试者样本）（Per Protocol Set (Valid Cases, Efficacy Sample, Evaluable Subjects Sample)）**

是指由充分依从于方案的受试者子集所产生的数据集，以确保这些数据按照所依据的科学模型可能会展现出处理效应。依从性包括以下一些考虑：如暴露于处理、可获得测量值以及无重大方案违背等。

**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.

**安全性和耐受性（Safety & Tolerability）**

医疗产品的安全性是指受试者的医学风险，通常在临床试验中由实验室检查（包括临床生化和血液学）、生命体征、临床不良事件（疾病、体征和症状），以及其他特殊的安全性检查（如心电图、眼科检查）等来评价。医疗产品的耐受性是指受试者能耐受明显不良反应的程度。

**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.

**统计分析计划（Statistical Analysis Plan）**

是指更技术性地和更详细地阐述方案中描述的分析要点的文件，包括对主要

和次要指标及其他数据进行统计分析的详细程序。

**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).

**优效性试验（Superiority Trial）**

是指主要目的为显示研究产品的反应优于对照（阳性药或安慰剂对照）的试验。

**Surrogate Variable**

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

**替代指标（Surrogate Variable）**

是指在直接测量临床效应是不可能的或不实际的情况下，用于间接测量临床效应的指标。

**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.

**处理效应（Treatment Effect）**

是指在临床试验中归因于处理的效应。在大多数临床试验中，感兴趣的处理效应是两个或多个处理的比较（或对比）。

**Treatment Emergent**

An event that emerges during treatment having been absent pre-treatment, or worsens relative to the pre-treatment state.

**治疗引发事件（Treatment Emergent）**

是指在处理期间出现的事件，它在处理前未发生或相对于处理前状态恶化。

**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.

**试验统计专业人员（Trial Statistician）**

是指同时具备充足的教育/培训和经验，可以实施本指南中的原则并负责临床试验统计方面的统计专业人员。