

药物临床试验的生物统计学指导原则

二〇一六年六月

附件

药物临床试验的生物统计学指导原则

一、概述

新药经临床前研究后,其有效性和安全性由人体临床试验进行最终验证。临床试验是根据研究目的,通过足够数量的目标受试者(样本)来研究药物对疾病进程、预后以及安全性等方面的影响。

临床试验除了遵循《药物临床试验质量管理规范》(GCP)以外,还必须事先应用统计学原理对试验相关的因素作出合理、有效的安排,最大限度地控制混杂与偏倚,减少试验误差,提高试验质量,并对试验结果进行科学的分析和合理的解释,在保证试验结果科学、可信的同时,尽可能做到高效、快速、经济。因此,统计学是临床试验设计、实施和分析的有力工具,在药物的临床研发过程中发挥不可或缺的重要作用。

本指导原则以临床试验的基本要求和统计学原理为核心,阐述统计学在临床试验中的作用和地位,以及在试验设计阶段、试验实施阶段和结果分析阶段的统计学考虑,旨在为药品注册申请人和临床试验的研究者针对临床研发中如何进行设计、实施、分析和评价提供技术指导,以保证药物临床试验的科学、严谨和规范。

本指导原则适用于以注册为目的的药物(化学药物、生物制

品、中药民族药和天然药物) 的确证性临床试验, 对探索性临床试验以及上市后临床试验也同样具有指导意义。

二、临床试验的总体考虑

(一) 临床研发规划

药物临床试验的主要目标是评价和确定受试药物的风险/获益比, 同时也要确定可能从该药获益的特定适应症人群及适宜的用法与用量。为此, 需要设计一系列的临床试验, 而每一个临床试验都有其特定的目的, 其设计、执行和拟采用的分析方法等细节均应在试验方案中予以明确。所以每个研究药物都应首先考虑其临床研发的总体规划。

创新药物的临床研发一般由 I 期临床试验开始, 进入 II 期概念验证试验 (Proof-Of-Concept, POC) 和剂量探索 (Dose Finding) 试验, 然后是 III 期确证试验, 每期试验由于研究目的的不同, 可能包含着多个试验项目。临床研发规划就是这些试验研究的总体规划。

在新药申请时, 应当清晰地描述该药临床研发规划的主要内容, 以及每个临床试验在其中的地位和作用。在解释和评价受试药物的总体证据时, 通常需要把几个试验的数据进行综合分析。因此, 同一临床研发规划中, 不同临床试验的多个方面应该尽量采用相同的标准, 如医学编码词典、主要指标的定义和测量时间点、对于方案违背的处理方式等等。在药物的临床研发规划中应预先阐明是否需要涉及共同医学问题的多个试验进行荟萃分

析 (Meta-Analysis), 并明确它们的设计共同点及关键统计问题。

(二) 探索性试验和确证性试验

临床试验的早期, 需要进行一系列的探索性试验, 这些试验也应有清晰和明确的目标。探索性试验有时需要更为灵活可变的方法进行设计并对数据进行分析, 以便根据逐渐积累的结果对后期的确证性试验设计提供相应的依据。虽然探索性试验对有效性的确证有参考价值, 但不能作为证明有效性的关键性证据。临床试验的后期, 需要经过确证性试验为评价药物的有效性和安全性提供有力证据。确证性试验是一种事先提出假设并对其进行统计检验的试验, 以说明所开发的药物对临床是有益的, 一般为随机对照的临床试验。因此, 对涉及药物有效性和安全性的每一个关键性的问题都需要通过确证性试验予以充分的回答。

在确证性试验中, 最关键的假设应根据试验主要目的产生。主要假设应于试验开始前在试验方案中预先设定并于试验结束后严格按照预先设定的分析计划完成假设检验。除此之外, 在试验方案中还应阐明试验设计方法、统计分析方法及相关理由。确证性试验对于试验方案和标准操作程序 (SOP) 的严格遵从是非常重要的。如果在试验过程中对方案有不可避免的修订, 应给予说明并记载。对方案修订可能对结果产生的影响应予以评估。

确证性试验还应对试验药物的疗效进行准确的估计。对于药物疗效的说明除了需要证明关键假设的统计学意义之外, 还需要评估试验药物疗效具有临床意义。

（三）观察指标

观察指标是指能反映临床试验中药物有效性和安全性的观察项目。统计学中常将观察指标称为变量。观察指标分为定量指标和定性指标。观察指标必须在研究方案中有明确的定义和可靠的依据，不允许随意修改。

对于观察指标，在研究的设计阶段，首先需要根据研究目的，严格定义与区分主要指标和次要指标，其次是根据主要指标的性质（定量或定性）和特征（一个或多个、单一指标或复合指标、临床获益或替代指标、客观/主观指标或全局评价指标等），调整研究的统计设计策略，以达到研究的预期目的。

1.主要指标和次要指标

主要指标又称主要终点，是与试验主要研究目的有本质联系的，能确切反映药物有效性或安全性的观察指标。主要指标应根据试验目的选择易于量化、客观性强、重复性高，并在相关研究领域已有公认标准的指标。

一般情况下，主要指标仅为一个，用于评价药物的疗效或安全性。若一个主要指标不足以说明药物效应时，可采用两个或多个主要指标。方案中应详细描述所关注的主要指标的设计参数及其假设、总 I 类错误率和 II 类错误率的控制策略。主要指标将用于临床试验的样本量估计，多个主要指标的情况下，将制定对总 I 类错误概率的控制策略并保证研究有足够的把握度。

主要指标，包括其详细定义、测量方法（若存在多种测量方

法时，应该选择临床相关性强、重要性高、客观并切实可行的测量方法)、统计分析模型等，都必须在试验设计阶段充分考虑，并在试验方案中明确规定。方案中主要指标在试验进行过程中不得修改，若须做修改则应在充分论证的基础上谨慎行事，并在揭盲前完成，不允许揭盲后对主要指标进行任何修改。

次要指标是与次要研究目的相关的效应指标，或与试验主要目的相关的支持性指标。在试验方案中，也需明确次要指标的定义，并对这些指标在解释试验结果时的作用以及相对重要性加以说明。一个临床试验，可以设计多个次要指标，但不宜过多，足以达到试验目的即可。

2.复合指标

当难以确定单一的主要指标时，可按预先确定的计算方法，将多个指标组合构成一个复合指标。临床上采用的量表(如神经、精神类、生活质量量表等)就是一种复合指标。将多个指标组合成单一复合指标的方法需在试验方案中详细说明。主要指标为复合指标时，可以对复合指标中有临床意义的单个指标进行单独的分析。

当采用量表进行疗效评价(如精神类药物、中药、民族药)，应该采用国际或领域内公认的量表。采用国外量表作为主要疗效指标时，由于可能存在语言、文化、生活习俗、宗教信仰等多方面的差异，需提供跨文化调适、翻译对等性的研究结果；采用自制量表时，需提供效度、信度和反应度(对疾病严重程度及其变

化的区分程度) 的研究结果。没有对效度、信度和反应度进行过研究, 或者效度、信度和反应度都很低的量表不建议作为临床试验的主要疗效指标。

3.全局评价指标

全局评价指标是将客观指标和研究者对受试者疗效的总印象有机结合的综合指标, 它通常是等级指标, 其判断等级的依据和理由应在试验方案中明确。全局评价指标可以评价某个治疗的总体有效性或安全性, 带有一定的主观成份, 因此, 其中的客观指标常被作为重要的指标进行单独分析。

以全局评价指标为主要指标时, 应该在方案中考虑: 该全局评价指标与主要研究目的临床相关性、信度和效度、等级评价标准和单项缺失时的估计方法。不建议将“综合疗效和安全性”的全局评价指标作为临床试验的主要指标, 因为这样会掩盖药物之间在疗效和安全性方面的重要差异, 从而导致决策失误。

4.替代指标

替代指标是指在直接评价临床获益不可行时, 用于间接反映临床获益的观察指标。例如降压药物的临床获益, 常被认为是降低或延迟“终点事件”(心脑血管事件) 的发生, 但若要评价“终点事件”发生率, 需要长时间的观察。在实际中, 降压药的临床试验, 采用替代指标“血压降低值/血压达标”来评价药物的疗效, 因为临床研究和流行病学已证实: 将“血压”控制在正常范围内, 可以降低“终点事件”的发生。

一个指标能否成为临床获益的替代指标，需要考察：1) 指标与临床获益的关联性和生物学合理性；2) 在流行病学研究中该指标对临床结局的预测价值；3) 临床试验的证据显示药物对该指标的影响程度与药物对临床结局的影响程度一致。

选择替代指标为主要指标，可以缩短临床试验期限，但也存在一定的风险，尤其是“新”替代指标。药物在替代指标上的优良表现并不一定代表药物对受试者具有长期的临床获益，药物在替代指标上的不良表现也不一定表示没有临床获益。例如，在抗肿瘤药物早期临床试验中，“无进展生存时间”等指标被作为“总生存时间”的替代指标被广泛使用，但其与总生存时间的关联性在不同的肿瘤临床试验中程度不一，因此仍需强调III期临床研究中，采用临床终点的重要性。

5. 定性指标

在某些临床试验中，有时需要将定量指标根据一定的标准转换为等级指标、或将等级指标转化为定性指标，如：用药后血压降低到“140/90mmHg”以下、糖化血红蛋白降低到7.0%以下的受试者比例（达标率）。定量或等级指标转换定性指标的标准，应该具有临床意义、为相关领域公认、并在试验方案中明确规定。由于将定量指标转换为定性指标会损失部分信息导致检验效能的降低，在样本量计算时需加以考虑。如方案定义主要指标为定量指标转化的定性指标时，则研究结论应主要依据该定性指标，而不是其所源于的定量指标。

（四）偏倚的控制

偏倚又称偏性，是临床试验在设计、执行、测量、分析过程中产生的、可干扰疗效和安全性评价的系统误差。在临床试验中，偏倚包括各种类型的对研究方案的违背与偏离。由于偏倚会影响疗效、安全性评价结果，甚至影响临床试验结论的正确性，因此在临床试验的全过程中均须控制偏倚的发生。随机化和盲法是控制偏倚的重要措施。

1. 随机化

随机化是临床试验的基本原则，也是疗效和安全性评价的统计学方法的基础。

临床试验中随机化原则是指临床试验中每位受试者均有同等的机会被分配到试验组或对照组中的实施过程或措施，随机化过程不受研究者和/或受试者主观意愿的影响。随机化的目的是使各种影响因素（包括已知和未知的因素）在处理组间的分布趋于相似。随机化与盲法相结合，可有效避免处理分组的可预测性，控制对受试者分组的选择偏倚。临床试验的随机化的方法，一般采用区组随机化法和/或分层随机化法。

如果受试者的入组时间较长，区组随机化是临床试验所必须的，这样有助于减少季节、疾病流行等客观因素对疗效评价的影响，也可减少因方案修订（如入选标准的修订）所造成的组间受试者的差异。区组的大小要适当，太大易造成组间不均衡，太小则易造成同一区组内受试者分组的可猜测性。研究者及其相关人

员,应该对区组长度保持盲态,这在开放的临床试验中尤为重要。也可设定2个或多个区组长度,或采用中央随机化系统以尽可能减少分组的可预测性。

如果药物的效应会受到一些预后因素(如受试者的病理诊断、年龄、性别、疾病的严重程度、生物标记物等)的影响时,可采用分层随机化,以保持层内的组间均衡性。

当需要考虑多个分层因素,如肿瘤类临床试验,需考虑年龄、病理类型、基线水平等因素,采用分层随机化,可能导致试验无法进行,此时可采用“动态随机”使被控制的预后因素组间有良好的均衡性。在动态随机化中,已入组的受试者特征将影响下一个受试者的分组,系统将根据各层面上的组间均衡性决定受试者的随机化组别。

尽管“动态随机”可以实现多分层因素下的随机化,但不建议设计过多的分层因素,因为过多的分层因素可能造成其他因素在处理组间的不均衡,建议分层因素一般不宜超过3个。临床试验中通常采用区组随机化的方法,如采用动态随机化,被控制的因素应包括在主要指标分析模型中,用以控制混杂因素对主要指标评价的影响。特别指出的是在III期临床试验中,应避免使用基于主要指标观察结果的动态随机化。

随机化的方法和过程包括随机分配表的产生方法、随机分配遮蔽的措施、随机分配执行的人员分工等,应在试验方案中阐明,但使人容易猜测分组的随机化的细节(如区组长度等)不应包含

在试验方案中。在临床试验中，随机分配表应该是一份独立的文件，以记录受试者的处理（或处理顺序）安排。随机分配表应具有重现性，即可以根据种子数、分层因素、区组长长度重新产生相同的随机分配表。试验用药物将根据随机分配表进行编码，在临床操作中，要求研究者严格按照入组受试者的随机分配结果及药物编码分配药物，任何偏离，都应该如实记录，以待数据分析前进行评估。值得注意的是动态随机化中的随机表仅仅起到遮蔽作用，真正的随机分配表是由动态随机化系统根据已入组的受试者信息采用最小随机化原理产生的，因此随机化系统中的随机分配表应作为独立文件在申报资料中提交。

2.盲法

临床试验的偏倚可能来自于临床试验的各个阶段、各方面人员。由于对随机化分组信息的知晓，研究者可能选择性入组受试者，受试者可能受到主观因素的影响，可能产生疗效与安全性的评价偏倚或选择性确定分析人群等。盲法是控制临床试验中因“知晓随机化分组信息”而产生的偏倚的重要措施之一，目的是达到临床试验中的各方人员对随机化处理分组的不可预测性。

根据设盲程度的不同，盲法分为双盲、单盲和非盲（开放）。在双盲临床试验中，受试者、研究者（对受试者进行筛选的人员、终点评价人员以及对方案依从性评价人员）、与临床有关的申办方人员对处理分组均应处于盲态；单盲临床试验中，仅受试者或研究者一方对处理分组处于盲态；开放性临床试验中，所有人员

都可能知道处理分组信息。临床试验的设盲程度，应综合考虑药物的应用领域、评价指标和可行性，应尽可能采用双盲试验。当双盲难度大、可行性较差，可考虑单盲临床试验，甚至开放性研究。一般情况下，神经、精神类药物的临床试验采用量表评价效应、用于缓解症状（过敏性鼻炎、疼痛等）的药物或以“受试者自我评价”等主观指标为主要指标的临床试验、以安慰剂为对照的临床试验，均应采用“双盲”；在一些以临床终点（如死亡）为主要评价指标的临床试验中（抗肿瘤药物），也可以接受开放性研究。

双盲的临床试验，要求试验药和对照药（包括安慰剂）在外观（剂型、形状、颜色、气味）上的一致性；如果试验药与对照药在用药方式有差异，还需要做到试验组与对照组在药物使用上的一致性。若要达到双盲的目的，可采用双模拟技术。在使用双模拟技术的临床试验中，受试者的用药次数与用药量将会增加，可能导致用药依从性的降低。

若双盲实施起来有相当的困难或根本不可行时（例如，手术治疗与药物治疗的对比研究；不同药物在剂型、外观或用法上存在很大的差异；因中药组方不同导致气味上的差异等），可以采用单盲或开放性临床试验，其理由必须在方案中详细说明，而且尤为重要是这种信息的知晓不得影响受试者分配入组的随机性，方案中还须有控制偏倚的具体措施，例如采用客观的主要指标，或采用中央随机化系统管理受试者的入组，或参与疗效与安

全性评价的研究者在试验过程中尽量处于盲态等。

无论是双盲、单盲临床试验，盲态的执行（随机化分配表的产生、保存以及释放）应该有标准操作程序进行规范，且在方案中明确规定破盲人员的范围。即使是开放性临床试验，研究相关人员也应尽可能保持盲态。方案中应该规定随机分配表的释放条件与流程。随机分配表释放的基本条件为：已完成数据库的锁定和分析人群及统计分析计划的确定工作。

三、试验设计的基本考虑

（一）试验设计的基本类型

1. 平行组设计

平行组设计是最常用的临床试验设计类型，可为试验药设置一个或多个对照组，试验药也可设多个剂量组。对照组可分为阳性或阴性对照。阳性对照一般采用按所选适应症的当前公认的有效药物，阴性对照一般采用安慰剂，但必须符合伦理学要求。试验药设一个或多个剂量组完全取决于试验的目的。

2. 交叉设计

交叉设计是按事先设计好的试验次序，在各个时期对受试者逐一实施各种处理，以比较各处理间的差异。交叉设计是将自身比较和组间比较设计思路综合应用的一种设计方法，它可以较好地控制个体间的差异，以减少受试者人数。

最简单的交叉设计是2种药物2个阶段的形式，又称 2×2 交叉设计，对每个受试者安排两个试验阶段，分别接受A、B两种试

验用药物，而第一阶段接受何种试验用药物是随机确定的，第二阶段必须接受与第一阶段不同的另一种试验用药物。因此，每个受试者接受的药物可能是先A后B（AB顺序），也可能是先B后A（BA顺序），故这种试验又简记为AB/BA交叉试验。两阶段交叉试验中，每个受试者需经历如下几个试验过程，即准备阶段、第一试验阶段、洗脱期和第二试验阶段。

每个试验阶段的用药对后一阶段的延滞作用称为延滞效应。前个试验阶段后需安排足够长的洗脱期或有效的洗脱手段，以消除其延滞效应。采用交叉设计时应考虑延滞效应对试验数据分析评价的影响。

2×2交叉设计难以区分延滞效应与时期-药物的交互作用。如需进一步分析和评价延滞效应，则可考虑采用2个处理多个阶段的交叉设计（例如：2×4的ABBA/BAAB交叉设计）。

多种药物多个阶段的交叉设计也是经常用到的，例如：3×3交叉设计，即3种处理（A、B、C）、3个阶段、6种顺序（ABC/BCA/CAB/ACB/CBA/BAC）的交叉设计。

由于每个受试者接受了所有处理组的治疗，提供了多个处理的效应，因此交叉试验中应尽量避免受试者的失访。

3.析因设计

析因设计是通过试验用药物剂量的不同组合，对两个或多个试验用药物同时进行评价，不仅可检验每个试验用药物各剂量间的差异，而且可以检验各试验用药物间是否存在交互作用，或探

索两种药物不同剂量的适当组合，常用于复方研究。析因设计时需考虑两种药物高剂量组合可能带来的毒副反应。

如果试验的样本量是基于检验主效应的目的而计算的，关于交互作用的假设检验，其检验效能往往是不足的。

（二）多中心试验

多中心试验系指由一个单位的主要研究者总负责，多个单位的研究者参与，按同一个试验方案同时进行的临床试验。多中心试验可以在较短的时间内入选所需的病例数，且入选的病例范围广，临床试验的结果更具代表性。但影响因素亦随之更趋复杂。

多中心试验必须遵循同一个试验方案在统一的组织领导下完成整个试验。各中心试验组和对照组病例数的比例应与总样本的比例大致相同。多中心试验要求试验前对人员统一培训，试验过程要有良好的质控措施。当主要指标易受主观影响时，需进行统一培训并进行一致性评估。当主要指标在各中心的实验室的检验结果有较大差异或参考值范围不同时，应采取相应的措施进行校正或标化以保证其可比性，如采用中心实验室检验等。如预期多中心间检验结果有较大差异，应在临床试验方案中预先规定可能采用的差异性的检验及校正方法。

在多中心临床试验中，可按中心分层随机；当中心数较多且每个中心的病例数较少时，可不按中心分层。

国际多中心试验可视为一种特殊形式的多中心试验，在不同国家或地区所观察的试验结果可能作为相应国家或地区药品注

册申请的重要依据。在这种特殊的需求下，国家或地区间的临床实践差异有可能对临床结果的解读产生较大的影响。在临床试验设计时应提前对这种差异进行预估，并在临床试验方案中对将采用的分析不同国家地区结果差异性/一致性的统计方法做预先规定。常用的一致性的评价方法有（但不限于）以国家或地区为预设亚组的亚组分析，或采用适当的统计分析模型等。当单独以某特定国家或地区试验数据作为主要注册申请依据时，应说明样本量能够合理的支持相对应的安全性及有效性的评价。

（三）比较的类型

临床试验中比较的类型，按统计学中的假设检验可分为优效性检验、等效性检验和非劣效性检验。在临床试验方案中，需要明确试验的目的和比较的类型。

优效性检验的目的是显示试验药的治疗效果优于对照药，包括：试验药是否优于安慰剂；试验药是否优于阳性对照药；或剂量间效应的比较。等效性检验的目的是确证两种或多种治疗的效果差别大小在临床上并无重要意义，即试验药与阳性对照药在疗效上相当。而非劣效性检验目的是确证试验药的疗效如果在临床上低于阳性对照药，其差异也是在临床可接受范围内。

在显示后两种目的试验设计中，阳性对照药的选择要慎重。所选阳性对照药需是已广泛应用的、对相应适应症的疗效和用量已被证实，使用它可以有把握地期望在目前试验中表现出相似的效果；阳性对照药原有的用法与用量不得任意改动。阳性药物选

择时应考虑以下两个方面：

1.阳性对照有效性的既有证据

阳性对照效应来源于文献报道的有良好试验设计的试验结果，这些历史试验已明确显示本次非劣效试验中采用的阳性对照或与其类似的药物优于安慰剂，且随时间迁移，阳性对照的疗效基本维持稳定。根据这些试验结果可以可靠地估计出阳性对照的效应大小。阳性对照的效应量是非劣效试验的关键设计参数（用以确定非劣效界值），既不能用历史研究中最好的疗效作为其效应量的估计，也不能仅用荟萃分析的点估计作为效应量的估计，效应量估计时要充分考虑历史研究间的变异。

2.阳性对照药物效应的稳定性

阳性对照效应的估计来源于历史研究，虽然考虑了历史研究间的变异，但仍有历史局限性，受到很多因素诸如当时的受试人群、合并用药、疗效指标的定义与判定、阳性对照的剂量、耐药性以及统计分析方法等的影响。因此，采用非劣效试验设计时要尽可能地确保本次临床试验在以上提及的诸多因素方面与历史研究一致。另外非劣效/等效性设计，良好的偏倚控制和质量控制是此类设计的关键。因此，在试验设计和实施阶段都应该提高试验质量要求，只有高质量的临床试验才能保证非劣效/等效临床试验的检定灵敏度。

进行等效性检验或非劣效性检验时，需预先确定一个等效界值（上限和下限）或非劣效界值（上限或下限），这个界值应不

超过临床上能接受的最大差别范围,并且应当小于阳性对照药与安慰剂的优效性试验所观察到的差异。非劣效界值确定一般采用两步法, M_1 是阳性对照扣去了安慰剂效应的绝对疗效的保守估计,一般借助荟萃分析法并考虑历史试验间的变异后确定; M_2 是非劣效界值,其确定要结合临床具体情况,在考虑保留阳性对照疗效的适当比例后,由统计专家和临床医学专家共同确定。在等效界值的确定中,可以用类似的方法确定下限和上限。从技术层面讲,等效性检验双侧置信区间等同于两个同时进行的单侧假设检验,而非劣效检验是单侧检验。非劣效/等效检验统计推断一般采用置信区间法。值得注意的是两组之间差别无统计学意义并不能得出两组等效或非劣的结论。

(四) 样本量

临床试验中所需的样本量应具有足够大的统计学检验把握度,以确保对所提出的问题给予一个可靠的回答,同时也应综合考虑监管部门对样本量的最低要求。样本的大小通常以试验的主要疗效指标来确定,如果需要同时考虑主要疗效指标外的其他指标时(如安全性指标或重要的次要指标),应明确说明其合理性。一般来说,在样本量的确定中应该说明以下相关因素,包括设计的类型、主要疗效指标的明确定义(如在降压药的临床试验中应明确说明主要指标是从基线到终点的血压改变值,或试验终点的血压达标率)、临床上认为有意义的差值、检验统计量、检验假设中的原假设和备择假设、I类和II类错误率以及处理脱落和方

案违背的比例等。在以事件发生时间为主要疗效指标的生存分析中，可以根据统计学检验把握度直接得到试验所需事件数。在此情况下需要根据事件发生率，入组速度以及随访时间推算试验所需样本量。

样本量的具体计算方法以及计算过程中所需用到的主要指标的统计参数（如均值、方差、事件发生率、疗效差值等）的估计值应在临床试验方案中列出，同时需要明确这些估计值的来源依据。在确证性临床试验中，一般只有一个主要疗效指标，参数的确定主要依据已发表的资料或探索性试验的结果来估算，其中所预期疗效差值还应大于或等于在医学实践中被认为是具有临床意义的差异。需要强调的是，计划中的试验应与前期试验或文献中的试验具有一致的试验设计和目标人群。如果不完全一致，需对相应统计量的估值进行调整。I类错误概率一般设定为双侧0.05。在非劣效检验等单侧检验中，I类错误概率一般设定为0.025。此外，如果试验设计中存在多重性的问题时，应考虑对I类错误概率进行必要的控制，以保证试验的总体I类错误概率不超过预设值。II类错误概率一般情况下设定为不大于0.2，在探索性试验中可适当放宽。

通过估计得到的试验所需样本量一般仅针对试验中指定的主要指标的主要分析（相对其他分析如敏感性分析或亚组分析而言）。在一个以“全分析集（Full Analysis Set，简称FAS）”为主要分析的试验中，应考虑统计参数估计值所依据的前期试验或资

料是否使用了相同的分析集或者具有相似的脱落率及方案违背率。考虑到脱落患者或违背方案者对疗效的稀释效应，全分析集的疗效往往小于符合方案集。此外，在全分析集中也常会观测到比符合方案集更大的变异。

另外，等效或非劣效试验中通常事先假设试验组与对照组疗效相同而进行样本量估算，当试验组的真实疗效差于阳性对照组时则试验的检验把握度将低于设定目标。

（五）适应性设计

适应性设计 (Adaptive Design) 是指事先在方案中计划的在临床试验进行过程中利用累积到的数据，在不影响试验的完整性和合理性的前提下，对试验的一个或多个方面进行修改的一种设计。

好的适应性设计可以加快药物研发的速度，或更有效地利用研发资源。但适应性设计要特别考虑：（1）试验的修改是否会引入 I 类错误增大；（2）试验的修改是否导致试验结果难于解释。因此，无论对试验进行何种修改，其修改计划和分析策略必须在试验数据揭盲之前在试验方案中进行明确严谨的表述。在适应性设计计划的期中分析中，保持申办者和研究者的盲态非常重要，通常需要一个独立的数据监查委员会 (Independent Data Monitoring Committee, IDMC) 来通知申办者是否按照事先拟定的方案修改进一步进行试验。

适应性设计有多种可能，包括：（1）试验组和对照组入组

分配方式的改变，如由固定区组分配变更为动态随机入组分配；

(2) 入组人数的改变，如样本量的重新计算；(3) 试验终止条件的改变，如根据期中分析结果提示有效或无效性而提前终止试验；(4) 或其他设计方法（如临床终点，统计方法）的改变。

目前应用的适应性设计中，成组序贯试验和盲态下样本量的重新计算被认为是在理论和实践中比较广泛被接受的。而其他的诸多设计对于深入认识试验结果的影响因素或提高研究效率（如富集设计）是有帮助的，但对于非盲态下改变临床终点或受试人群等适应性设计，由于可能引入偏倚而影响对结论的判断，故不宜应用于确证性试验中，可在早期探索性试验中使用。

1.成组序贯设计

成组序贯设计常用于有期中分析的临床试验中。适用于下列三种情况：(1) 怀疑试验药物有较高的不良反应发生率，采用成组序贯设计可以较早终止试验；(2) 试验药疗效较差，采用成组序贯设计可以因无效较早终止试验；(3) 试验药与对照药的疗效相差较大，但病例稀少，或临床观察时间过长。可见，成组序贯设计一般用于创新药物的临床试验，而不用于仿制药的临床试验。

成组序贯设计是把整个试验分成若干个连贯的分析段，每个分析段病例数可以相等也可以不等，但试验组与对照组的病例数比例与总样本中的比例相同。每完成一个分析段，即对主要指标（包括有效性和/或安全性）进行分析，一旦可以做出结论即停

止试验，否则继续进行。如果到最后一个分析段仍不拒绝无效假设，则作为差异无统计学意义而结束试验。

成组序贯设计的优点是当试验药与对照药间确实存在差异时，或试验药与对照药不可能达到统计学意义时，可较早地得到结论，从而缩短试验周期。

成组序贯设计的盲底要求一次产生，分批揭盲。由于多次重复进行假设检验会使I类错误增加，故需对每次检验的名义水准进行调整，以控制总的I类错误率不超过预先设定的水准（比如 $\alpha=0.05$ ）。试验设计中需明确 α 消耗函数的方法。

采用成组序贯设计，由于需要进行多次期中分析，需特别注意盲态的保持，以免引入新的偏倚。同时，在试验开始前应预先明确统计分析方法，规定提前终止试验的标准。期中分析的数据需由独立的第三方进行统计分析，并由审核，以便做出是否继续下一阶段临床试验的决策建议。

2.盲态下的样本量的重新计算

当原设计中样本量是在不确切信息的假设条件下估计的，对时间比较长的临床试验，可以在试验进行中对这些假设进行验证，以便对样本量进行重新估计。为了避免揭盲对试验的I类错误及试验的科学完整性的影响，这种估计应该是在不揭盲的状态下进行的，主要是对连续变量的变异度或事件发生率进行估计。此估计可用于计算新的样本量，新的样本量若和原样本量相似或比原样本量小，应保持试验样本量不变。若新的样本量比原样本量大

并且是切实可行，应通过修订方案写明新的样本量。此类样本量的调整也可由IDMC来完成，并事先应在试验方案中对样本量的重新估计进行计划，样本量再估计应不超过两次。

3.富集设计

适应性的富集设计是指当期中分析提示某一亚组人群的疗效优于另一亚组人群的疗效时而调整入组标准，对尚未入组的病例规定只入组疗效好的某一亚组人群的设计。富集设计能够减少研究人群的异质性，从而提高研究的效率。该设计通常根据研究对象与疾病或者预后相关的某些特征把目标人群分为亚组，例如，研究心血管疾病时，可以考虑按目标人群病人是否有糖尿病、高血压等分为高危人群和非高危人群；在肿瘤领域中，通常具有某些生物标记的人群对治疗的反应比没有生物标记的人群要好，这时可以考虑把目标人群分为生物标记阳性和阴性两个亚组。最常用的病人分组因素包括（但不限于）人口学、病理生理学、组织学、遗传学的等等特征；研究方案中一般必须预先明确指出期中分析是根据病人的哪些特征做的亚组分析。根据病人特征进行分组的方法必须经过验证。

期中分析时根据研究方案中预设的病人特征，估计疗效并决定是否需要调整入组标准。由于期中对方案的调整对后续试验在随机、双盲等方面都有一定的影响，方案中必须明确规定避免引入偏倚和调整I类错误概率的方法。最后的结果分析是根据期中分析（方案调整）之前搜集的所有研究对象的数据和之后某一亚

组人群的数据加权，而不仅仅是感兴趣的亚组人群的数据；而分析结果的解释也必须明确地说明各亚组人群的构成。值得指出的是，由于富集设计的复杂性，可能对试验的基本原则（如，随机化、双盲、I类错误概率等）有严重影响，在没有充分可靠方法处理和避免这些影响时，和在病人特征对疗效的影响预先不明确的情况下，须慎用富集设计。

四、试验进行中的基本考虑

（一）试验数据和受试者入组的监查

临床试验过程中，研究者应严格按照试验方案认真进行临床试验，其研究过程的质量对研究数据及结果的可靠性有着重要的影响，因此，认真进行试验过程的数据监查能及早地发现问题，并使问题的发生和再现达到最小。

按照 GCP 要求，临床试验的申办者应在临床试验过程中委派监查员，对整个临床试验的质量进行监查。为监控试验的质量，对试验管理的监查应包括研究是否按计划进行，是否遵从方案，收集的数据质量如何，是否达到了预期收集的样本数量目标，设计的假设是否合适，以及病人在试验中的安全等权益是否有保障等。

在病人入组时间较长的试验中，必须对病人入组的积累情况进行监查。如入组率远低于试验方案中预定的水平，则需查明原因，并采取相应措施缓解入组选择和质量的其他方面问题，确保试验的把握度。在多中心试验中，这些考虑适用于每一个

中心。

这类监查既不需要比较处理效应的信息,也不要对试验分组揭盲,所以对 I 类错误没有影响,是试验申办者应尽的职责,它可由试验申办者或由试验申办者委托的合同研究组织 (CRO) 来完成。这种监查一般从研究地点的选定开始,到完成最后一位病人数据的收集和清理而结束。

(二) 试验方案的修改

确定的试验方案经伦理委员会批准后,在试验进行过程中一般不得更改。对试验方案的任何修改都应在修订方案中写明,且修订方案一般需重新得到伦理委员会的批准。

在试验进行过程中,如发现按原入选/排除标准难以选到合格的病例时,需分析原因并采取相应措施,如监查中发现常有违反标准入选病例现象或入选病例的限制过度情况,则在不破盲的条件下可以考虑修改原入选/排除标准,但需注意入选/排除标准的修改可能导致目标受试人群的改变。修改后需调整相应的统计分析计划,如对方案修改前后进行分层分析及其结果一致性的考虑等应进行详细表述。

(三) 期中分析

期中分析是指正式完成临床试验前,按事先制订的分析计划,比较处理组间的有效性和/或安全性所作的分析。其分析目的是为后续试验是否能继续执行提供决策依据。基于期中分析结果中止试验无外乎两种情况,其一是可以预见即使试验继续执行至试

验结束也不可能得出试验药物有效的结论,或者是发现试验药物的安全性存在隐患;另一种是得出试验药物有效的结论。如果根据期中分析得出试验药物有效而提前中止试验,需要保证有足够的药物暴露时间和安全性数据,一般应继续随访以收集更多的安全性数据,以避免安全性评价不充分。

期中分析的时点(包括日历时点或信息时点)、具体实施方式和所采用的 α 消耗函数等应当事先制订计划并在试验方案中阐明。期中分析的结果可能会对后续试验产生影响,因此,一个临床试验的期中分析次数应严格控制。如果一个期中分析是为了决定是否终止试验而设计的,则常采用成组序贯设计。

期中分析包含了已揭盲的数据及结果,因此进行期中分析的人员应该是不直接参加临床试验的人员,比如可由IDMC执行,即使是开放的试验也应如此。期中分析结果对试验相关人员是保密的,试验相关人员仅仅会被告知是否继续试验或需要对试验方案进行修改。

对于确证性临床试验,原则上不得进行计划外期中分析。因为设计不良或计划外的期中分析可能引入偏倚,所得结论缺乏可靠性。如由于特殊情况进行了计划外的期中分析,则在研究报告中应解释其必要性、以及破盲的程度和必要性,并提供可能导致偏倚的严重程度以及对结果解释的影响。

(四) 独立数据监查委员会

独立数据监查委员会(IDMC),有时也称为数据和安全监

查委员会 (Data And Safety Monitoring Board, 简称DSMB) , 或数据监查委员会 (Data Monitoring Committee, 简称DMC) , 是由具备相关专业知识和经验的一组专业人员组成的独立委员会, 通过定期评估一项或多项正在进行的临床试验的累积数据, 评价试验药物的安全性和有效性。保证受试者安全和权益并确保试验的完整可靠性是IDMC的基本职责。

IDMC通常用于以延长生命或减少重大健康结局风险为目的的大规模多中心临床试验, 而大多数临床试验不要求或无需使用IDMC。可以考虑聘用IDMC的情况包括 (但不局限于) 下列一种或多种: (1) 对安全性或有效性的累积数据进行期中分析, 以决定是否提前终止试验; (2) 存在特殊安全问题的试验, 如治疗方式有明显侵害性; (3) 试验药物可能存在严重毒性; (4) 纳入潜在的弱势人群进行研究, 如儿童、孕妇、高龄者或其他特殊人群 (疾病终末期病人或智障的病人); (5) 受试者有死亡风险或其他严重结局风险的研究; (6) 大规模、长期、多中心临床研究。

IDMC具有以下三个特点: (1) 多学科性: IDMC成员应该包括有相应临床知识及掌握期中分析原则的临床专家、统计学家或医学伦理学家等; (2) 独立性: IDMC的成员需没有任何利益冲突; 其独立性可以防止试验的组织者、申办者的既得利益可能对数据评估带来的影响; (3) 保密性: 由于期中分析数据是非盲态的, 因此需要由与试验无利益冲突的人员来担任统

计分析的工作，并且注意保密性。一般需要选定独立统计师负责期中数据分析并完成期中报告。期中数据和报告在传递和提交过程中应采取一定的保密措施并妥善保存，避免被申办者或其他人员不慎或不当接触，直到试验结束申办者和研究者才能接触到期中报告。

IDMC的组织和实施过程需要在试验设计阶段或IDMC启动会之前制定详细的IDMC工作章程和程序，并经IDMC成员审阅、签字、存档，在整个IDMC运行过程中作为工作指南严格遵守。所有IDMC的讨论需有会议纪要，以便在试验结束后有案可查。同时，在确保保密性的基础上，IDMC成员、IDMC支持小组和申办者之间应进行及时有效地沟通和交流。

五、试验的数据管理

数据管理的目的是确保数据的可靠、完整和准确。临床试验中的数据管理相关方包括申办者、研究者、监查员、数据管理员和CRO等，各相关方应各司其职、各尽其责。数据管理全过程的实施，从数据采集到数据库的最终建立，都必须符合我国GCP的规定和监管部门的相应技术规范要求。

临床试验方案确定后，应根据病例报告表和统计分析计划书的要求制订数据管理计划，内容涵盖数据管理各过程，包括数据接收、录入、清理、编码、一致性核查、数据锁定和转换。数据的收集和传送，从研究者到申办者可通过多种媒体，包括纸质的病例报告表、电子数据采集系统以及用于临床试验数据管理的计

计算机系统。无论采用何种方式收集数据，资料的形式和内容必须与研究方案完全一致，且在临床试验前确定，包括确定对计划的依从性或确认违背试验方案的前后关系的信息（如有关服药的时间，缺失值需与“0 值”和空缺相区别）。数据管理各过程的执行中均应遵守全面和有效的标准操作程序。

无论是采用纸质化或电子化的数据管理，其各阶段均应在一个完整、可靠的临床试验数据质量管理体系（QMS）下运行，对可能影响数据质量结果的各种因素和环节进行全面控制和管理，使临床研究数据始终保持在可控和可靠的水平。临床试验的数据管理系统（CDMS）必须满足三个基本要求：经过基于风险考虑的系统验证，具备可靠性；具备数据可溯源性的性能；具备完善的权限管理功能。另外，在数据管理运行过程中应该建立和实施质量保证、质量控制和质量评估等措施。临床试验中用于数据管理和统计分析的计算机及其软件系统均应经过验证且有验证记录可查。

为达到试验数据共享和信息互通目的，临床试验过程中数据的采集、分析、交换、提交等环节，可考虑采用统一的标准化格式，如CDISC临床数据交换标准体系（Clinical Data Interchange Standards Consortium）。

临床试验完成后，应对试验的数据管理工作和过程进行总结并形成数据管理总结报告。数据管理计划和总结报告应作为药物注册上市的申请材料之一提交给监管部门。

六、统计分析和报告

（一）统计分析计划

统计分析计划（Statistical Analysis Plan, 简称SAP）是比试验方案中描述的分析要点更加技术性和有更多实际操作细节的一份独立文件, 包括对主要和次要评价指标及其他数据进行统计分析的详细过程。统计分析计划的内容包括设计的类型、比较的类型、随机化与盲法、主要指标和次要指标的定义与测量、检验假设、数据集的定义、疗效及安全性统计分析的详细细节。确证性试验要求提供详细分析原则及预期分析方法。探索性试验通常描述概括性的分析原则和方法。

统计分析计划由试验统计学专业人员起草, 并与主要研究者商定, 旨在全面而详细地陈述临床试验数据的分析方法和表达方式, 以及预期的统计分析结果的解释。

统计分析计划初稿应形成于试验方案和病例报告表确定之后, 在临床试验进行过程中以及数据盲态审核时, 可以进行修改、补充和完善, 不同时点的统计分析计划应标注版本及日期, 正式文件在数据锁定和揭盲之前完成并予以签署。如果试验过程中试验方案有修订, 则统计分析计划也应作相应的调整。如果涉及期中分析, 则相应的统计分析计划应在期中分析前确定。

（二）统计分析集

用于统计分析的数据集事先需要明确定义, 并在盲态审核时确认每位受试者所属的分析集。一般情况下, 临床试验的分析数

数据集包括全分析集 (FAS)、符合方案集 (Per Protocol Set, 简称PPS) 和安全集 (Safety Set, 简称SS)。根据不同的研究目的, 需要在统计分析计划中明确描述这三个数据集的定义, 同时明确对违背方案、脱落/缺失数据的处理方法。在定义分析数据集时, 需遵循以下两个原则: (1) 使偏倚减到最小; (2) 控制I类错误率的增加。

意向性治疗的原则 (Intention To Treat Principle, 简称ITT), 是指主要分析应包括所有随机化的受试者, 这种保持初始的随机化的做法对于防止偏倚是有益的, 并且为统计学检验提供了可靠的基础, 这一基于所有随机化受试者的分析集通常被称为ITT分析集。

理论上遵循ITT原则需要对所有随机化受试者的研究结局进行完整的随访, 但实际中这种理想很难实现, 因而也常采用全分析集 (FAS) 来描述尽可能的完整且尽可能的接近于包括所有随机化的受试者的分析集。

只有非常有限的情况才可以剔除已经随机化的受试者, 通常包括: 违反重要入组标准; 受试者未接受试验用药物的治疗; 随机化后无任何观测数据。值得注意的是, 这种剔除需要对其合理性进行充分的论证和说明。

符合方案集 (PPS), 亦称为“可评价病例”样本。它是全分析集的一个子集, 这些受试者对方案更具依从性。纳入符合方案集的受试者一般具有以下特征: (1) 完成事先设定的试验药物

的最小暴露量：方案中应规定受试者服用药物的依从性达到多少为治疗的最小量；(2) 试验中主要指标的数据均可以获得；(3) 未对试验方案有重大的违背。

受试者的排除标准需要在方案中明确，对于每一位从全分析集或符合方案集中排除的受试者，都应该在盲态审核时阐明理由，并在揭盲之前以文件形式写明。

安全集 (SS)，应在方案中对其明确定义，通常应包括所有随机化后至少接受一次治疗且有安全性评价的受试者。

对于确证性试验，宜同时采用全分析集和符合方案集进行统计分析。当两种数据集的分析结论一致时，可以增强试验结果的可信性。当不一致时，应对其差异进行讨论和解释。如果符合方案集被排除的受试者比例太大，则将影响整个试验的有效性。

ITT/全分析集和符合方案集在优效性试验和等效性或非劣效性试验中所起的作用不同。一般来说，在优效性试验中，应采用ITT/全分析集作为主要分析集，因为它包含了依从性差的受试者而可能低估了疗效，基于ITT/全分析集的分析结果是保守的。符合方案集显示试验药物按规定方案使用的效果，但与上市后的疗效比较，可能高估疗效。在等效性或非劣效性试验中，用ITT/全分析集所分析的结果并不一定保守，在统计分析时，可以用符合方案集和ITT/全分析集作为分析人群，两个分析集所得出的结论通常应一致，否则应分析并合理解释导致不一致的原因。

（三）缺失值及离群值

缺失值是临床试验中的一个潜在的偏倚来源，因此，病例报告表中原则上不应有缺失值，尤其是重要指标（如主要的疗效和安全性指标）必须填写清楚。对病例报告表中的基本数据，如性别、出生日期、入组日期和各种观察日期等不得缺失。试验中观察的阴性结果、测得的结果为零和未能测出者，均应有相应的符号表示，不能空缺，以便与缺失值相区分。

在临床试验中，数据缺失是难以避免的问题。在试验的计划、执行过程中应有必要的措施尽量避免缺失值的发生，在分析和报告中要正确处理缺失数据，否则会造成潜在的偏倚。缺失值的存在有可能导致试验结果无法解释。在分析中直接排除有数据缺失的受试者可能会（1）破坏随机性；（2）破坏研究样本对于目标人群的代表性。除此之外，对缺失值的直接排除还可能降低研究的把握度或减小变量的变异性引起 I 类错误率的膨胀。

如果在一些受试者中发生主要终点的缺失，在试验方案或统计计划书中应预先指定如何处理缺失值。

缺失机制可分为完全随机缺失（Missing Completely At Random, 简称 MCAR）、随机缺失（Missing At Random, 简称 MAR）和非随机缺失（Missing Not At Random, 简称 MNAR）。由于缺失机制无法通过已有数据进行判断，并且不同的处理方法可能会产生截然不同的结果，应当认识到任何缺失数据处理方法本身可能是潜在的偏倚来源。对完全随机缺失、随机缺失数据的

处理目前有末次观测值结转 (LOCF)、基线观测值结转 (BOCF)、均值填补、回归填补、重复测量的混合效应模型 (MMRM)、多重填补等多种不同的方法。

对于缺失值的处理方法，特别是主要疗效指标的缺失值，应事先在方案中根据以往的经验或既有相似试验的处理方法进行规定。然而如上所述，任何缺失数据处理方法本身都可能带来潜在的偏倚。所以缺失数据的处理方法应遵循保守的原则。即使同一种方法在不同情况下既有可能对试验药保守也有可能对试验药有利。然而，有时在对主要疗效指标的缺失值的处理方法进行预设时（如在盲态下）无法完全确定所用方法的保守性。必要时，也可以采用不同的处理缺失值的方法进行敏感性分析。

离群值问题的处理，应当从医学和统计学专业两方面去判断，尤其应当从医学专业知识判断。离群值的处理应在盲态检查时进行，如果试验方案未预先指定处理方法，则应在实际资料分析时，进行包括和不包括离群值的两种结果比较，评估其对结果的影响。

（四）数据变换

分析之前对关键变量是否要进行变换，最好根据以前的研究中类似资料的性质，在试验设计时即做出决定。拟采用的变换（如对数、平方根等）及其依据需在试验方案中说明，数据变换是为了确保资料满足统计分析方法所基于的假设，变换方法的选择原则应是公认常用的。一些特定变量的常用变换方法已在某些特定

的临床领域得到成功地应用。

（五）统计分析方法

统计分析应建立在真实、可靠、准确、完整的数据基础上，采用的统计方法应根据研究目的、试验方案和观察指标来选择，一般可概括为以下几个方面：

1.描述性统计分析

一般多用于人口学资料、基线资料 and 安全性资料，包括对主要指标和次要指标的统计描述。

2.参数估计、置信区间和假设检验

参数估计、置信区间和假设检验是对主要指标及次要指标进行评价和估计的必不可少的手段。假设检验应说明所采用的是单侧还是双侧检验，如果采用单侧检验，应说明理由。单侧检验的I类错误概率往往选择为双侧检验的一半，以保证单双侧检验的逻辑性。主要指标效应分析要说明采用的是固定效应模型还是随机效应模型。统计分析方法的选择要注意考虑指标的性质及数据分布的特性。无论采用参数方法或非参数方法，处理效应的估计应尽量给出效应大小、置信区间和假设检验结果。除主要指标和次要指标外，其他指标的分析以及安全性数据的分析也应简要说明所采用的方法。在确证性试验中，只有方案或统计分析计划中事先规定的统计分析才可以作为确证性证据的依据，而其他的分析只能视作探索性的。

3.基线与协变量分析

评价药物有效性的主要指标除受药物作用之外，常常还有其他因素的影响，如受试者的基线情况、不同治疗中心受试者之间差异等因素，这些因素在统计分析中可作为协变量处理。在试验前应认真考虑可能对主要指标有重要影响的协变量以及采用的可以提高估计精度的方法（如采用协方差分析方法），补偿处理组间由于协变量不均衡所产生的影响。对于确证性分析，应事先在方案中规定在统计模型中校正的协变量，以及校正的依据。当采用分层随机时，分层因素应作为协变量进行校正。对于事先没有规定校正的协变量，通常不应进行校正。也可以采用敏感性分析方法，将校正后的结果作为参考，而不应该取代事先规定的分析模型。

4.中心效应

多中心临床试验中，不同中心在受试者基线特征、临床实践等方面可能存在差异，导致不同中心间的效应不尽相同，这种中心之间的效应差异称为中心效应。常见三种情况：（1）无中心效应，即各中心试验组效应同质，对照组效应亦同质，此时各中心间效应是一致的；（2）有中心效应，但中心与处理组间不存在交互作用，即各中心试验组与对照组效应之差是同质的；（3）有中心效应，且中心与处理组间存在交互作用，此时，各中心试验组与对照组效应之差是异质的。中心与处理组间的交互作用，又分为定量的交互作用（各中心试验组与对照组效应之差方向一致）和定性的交互作用（至少一个中心的处理组与对照组的效应

之差与其他中心方向不一致)。

分析主效应时,对于情况(1),模型中应不包括中心效应;对于情况(2),模型中可包括中心项,但不包含中心与处理的交互项效应以提高检验效能;对于情况(3),若存在定量交互作用,则需要采用合适的统计学方法来估计处理效应,以保证结果的稳健性,结果解释时须非常谨慎,应努力从试验的管理、受试者的基线特征、临床实践等方面寻找原因;当存在定性的交互作用时,需找到合理的解释并重新进行的临床试验。

当中心数较多,或每个中心样本数均较少,一般无需考虑中心效应对主要变量及次要变量的影响,因为此时中心效应不会影响临床效果。

采用何种策略分析中心效应需事先在试验方案或统计分析计划中阐明。

5.亚组分析

临床试验中的亚组分析是对整体中根据某种因素分层的部分数据进行分析。

试验药物的疗效或安全性在不同的亚组中可能不同,而且这种差异往往具有特殊的临床意义。除非在方案设计时考虑到了计划的亚组分析,并且在样本量计算和多重性比较等方面事先给予了考虑,这样的亚组分析结果才能够被接受。由于亚组分析通常是小样本,且未按亚组随机化,故对于非确证性亚组分析的解释应当慎重,通常只能作为探索性研究的参考。

6.多重性问题

多重性问题是指在临床试验中，由于存在多个主要指标、多个比较组、多个时间点的比较、期中分析、亚组分析、多个分析集等情况，进行多次假设检验而导致 I 类错误概率增加的现象。如果试验将重要的次要指标结果也纳入关键性证据的情况下，即主要指标和重要次要指标共存时的假设检验亦需要考虑多重性问题。对于主要指标是复合指标的试验，如果宣称的疗效是基于复合指标中某个或某些成分时，需事先定义这些成分并纳入多重性考虑的确证性分析策略。

将假阳性率控制在事先设定的水平以内是非常重要的原则，在确证性临床试验结果的评价中具有重要的意义。在试验方案或统计分析计划中应预先说明对多重性问题的考虑、控制 I 类错误概率的原因及方法。处理多重性问题的方法有多种，如单步法、闭合检验程序、固定顺序的检验、序贯结构的策略等，在选择方法时可考虑将能够估计出疗效的可信区间作为选择的一个标准。

在对 I 类错误概率进行控制的同时可能会导致 II 类错误概率的增加，在估计样本量时应有所考虑。

（六）安全性与耐受性分析

安全性主要关注于药物对受试者的风险，在临床试验中，通常通过实验室检查结果（包括生化学和血液学指标）、生命体征、临床不良事件（疾病、体征、症状）及其他特殊的安全性检验（如心电图、眼科检查）等手段来评价。耐受性指受试者对于明显的

不良反应的耐受程度。

大多数试验中，对安全性与耐受性的分析，常采用描述性统计分析方法，必要时辅以置信区间进行说明。也可应用图表来描述治疗组间和个体间不良事件的发生模式（时间、空间、人群、性别分布）。不良事件的发生率通常以出现不良事件的病例数与暴露病例数之比来表示。此外，暴露强度（如人·年）也有可能作为分母。在各个阶段的临床研究过程中，应考虑对安全性评价指标定义的一致性，应考虑采用统一的不良事件编码词典（如MedDRA、WHOART和WHO-DD等）。

安全性和耐受性分析的数据集通常包括至少接受过一次治疗且有安全性评价的受试者。安全性的统计分析方法可以采用不同方式，可在方案及统计分析计划中结合临床判断，对不同的安全性指标按其重要性及与治疗的相关性划分为不同的类别：重要性较低且与治疗方法相关性较弱的安全性指标，可采用描述性分析方法；对于重要性适中且与治疗方法有一定相关性的安全性指标，建议加入置信区间分析；而对于重要性较高且与治疗方法相关性较强的安全性指标，可提供相应的统计检验p值以供参考。

（七） 统计分析报告

统计分析报告是临床试验统计师根据事先拟定的统计分析计划书，应用统计分析软件编写分析程序输出的统计分析表格和统计分析图形加以整理的重要文档，也是提供给临床主要研究者作为撰写临床试验总结报告的重要素材，并和统计分析计划一起

作为药物注册上市的申请材料之一提交给监管部门用于对临床试验结果的评价。

七、名词解释

安全集 (Safety Set, SS)：安全性与耐受性评价时，用于汇总的受试者集称为安全集。安全集应考虑包括所有随机化后至少接受一次治疗的且有安全性评价的受试者受试者。

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）等来判定。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。

处理效应 (Treatment Effect)：是指归因于临床试验中处理的效果。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间的比较（或对比）。

等效性试验 (Equivalence Trial)：是指主要目的为确认两种或多种治疗效果的差别大小在临床上并无重要意义的试验。通常以真正的治疗效果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。

独立数据监查委员会 (Independent Data Monitoring Committee, IDMC)：也称数据和安全监查委员会、监查委员会、数据监查委员会。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疗效指标，并可向申办者建议是否继续、修改或停止试验。

多中心试验（Multicentre Trial）：多中心试验系指由多个单位的研究者合作，按同一个试验方案同时进行的临床试验。

非劣效性试验（Non-Inferiority Trial, NI）：是指主要目的为显示试验药物的效应在临床上不劣于对照药的试验。

符合方案集（Per Protocol Set, PPS）：又称有效病例、疗效样本、可评价病例样本。是由充分依从于试验方案的受试者所产生的数据集，以确保这些数据可能会展现出治疗的效果。依从性包括以下一些考虑：如所接受的治疗、指标测量的可获得性以及试验方案没有大的违背等。

交互作用（Interaction）：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。定量的交互作用是指对比差异的大小在因素的不同水平时不同；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。

荟萃分析（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总结性的统计量进行统计合并，但此名词有时也用于对原始数据的合并。

盲态审核（Blind Review）：是指在试验结束（最后一位受试者最后一次观察）到揭盲之前对数据进行的核对和评估，以便最终确定统计分析计划。

偏倚（Bias）：是指与设计、实施、分析和评价临床试验有

关的任何因素导致的处理效应估计值与其真值的系统偏离。临床试验实施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“统计学”偏倚。

期中分析（Interim Analysis）：是指正式完成临床试验前，按事先制订的分析计划，比较处理组间的有效性或安全性所作的任何分析。

全分析集（Full Analysis Set, FAS）：是指尽可能接近符合意向性治疗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剔除受试者后得到。

全局评价指标（Global Assessment Variable）：为单一变量，是将客观指标和研究者对病人的病情及其改变总的印象综合起来所设定的指标，通常是一个有序分类等级指标。

随机分配表的释放（Randomization Code Release）：是指临床试验中对最后一例受试者的随访结束，且所计划的数据采集工作全部完成后，为进一步完成计划的统计分析工作而将一直保持盲态的受试者的随机分组信息对相关研究人员进行公开的解盲过程。

试验统计学专业人员（Trial Statistician）：是指接受过专门培训且有经验，可以执行本指导原则并负责临床试验统计方面的统计学专业人员。

双模拟（Double-Dummy）：是指在临床试验中当两种处理（如治疗）不能做到完全相同时，使试验处理（或治疗）仍能保

持盲态的一种技术。先准备处理 A (活性药和不能区分的安慰剂) 和处理 B (活性药和不能区分的安慰剂), 然后受试者接受两套处理: 活性药处理 A 和安慰剂处理 B, 或者安慰剂处理 A 和活性药处理 B。

替代指标 (Surrogate Variable): 是指在直接测定临床效果不可能或不实际时, 用于间接反映临床效果的指标。

统计分析计划 (Statistical Analysis Plan, SAP): 是比试验方案中描述的分析要点更加技术性和有更多实际操作细节的一份独立文件, 包括对主要和次要指标及其他数据进行统计分析的详细过程。

脱落 (Dropout): 是指受试者由于任何原因不能继续按试验方案进行到所要求的最后一次随访。

意向性治疗原则 (Intention-To-Treat Principle): 是指基于有治疗意向的受试者 (即计划好的治疗) 而不是实际给予治疗的受试者进行评价的处理策略。是可以对结果做出评定的最好原则。其结果是计划分配到每一个治疗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治疗过程。

优效性试验 (Superiority Trial): 是指主要目的为显示试验药物的效应优于对照药 (阳性药或安慰剂) 的试验。

置信区间 (Confidence Interval, CI): 是指按一定的概率或可信度(1- α)用一个区间来估计总体参数所在的范围, 该范围通常

称为参数的置信区间。

中央随机化系统 (Centralized Randomization System): 是指在多中心临床试验中为克服人为或其他未知因素对研究结果的偏倚影响, 由一个独立的组织或机构基于电话语音或网络方式实施药物随机分配的自动化计算机管理系统。常见有基于电话的交互式语音应答系统 (IVRS, Interactive Voice Response System) 和基于网络的交互式网络应答系统 (IWRS, Interactive Web Response System)。

八、参考文献

- 1.SFDA: 药物临床试验质量管理规范 (GCP) 2003
- 2.CDE: 临床试验数据管理工作技术指南 (征求意见稿) 2016
- 3.CCTS: 非劣效临床试验的统计学考虑。中国卫生统计, 2012, 29 (2): 270-274
- 4.CCTS: 临床试验中多重性问题的统计学考虑。中国卫生统计, 2012, 29 (3): 1-6
- 5.ICH E9: Statistical Principles for Clinical Trials 1998
- 6.ICH E8: General Considerations for Clinical Trials 1997
- 7.FDA: Adaptive Design Clinical Trials for Drugs and Biologics(Draft Guidance) 2010