

指导原则编号：【H】GCL4-1

化学药物和生物制品临床试验的生物统计学 技术指导原则

二〇〇五年三月

目 录

一、概述	1
二、整个临床试验的基本考虑	1
(一)探索性试验和确证性试验	1
(二)观察指标	2
(三)偏倚的控制	4
三、试验设计中的基本考虑	8
(一)试验设计的类型	8
(二)多中心试验	10
(三)比较的类型	10
(四)样本量	11
(五)资料的收集	12
四、试验进行中的基本考虑	12
(一)期中分析	12
(二)试验方案的修改	13
五、数据管理	13
六、统计分析	15
(一)统计分析计划书	15
(二)统计分析集	15
(三)缺失值及离群值	16
(四)数据变换	17
(五)统计分析方法	17

(六)安全性评价.....	18
七、统计分析报告.....	19
八、名词解释.....	20
九、参考文献.....	23
十、附录.....	23
十一、著者.....	23

化学药物和生物制品临床试验的生物统计学 技术指导原则

一、概述

新药经临床前研究后，其有效性和安全性的评价需要由临床试验加以确认。由于临床试验通常是根据研究的目的，通过足够数量的受试者（样本）来研究药物对疾病进程、预后等方面的作用，因此，临床试验设计必须应用统计学原理对试验相关的因素做出合理的、有效的安排，并最大限度地控制试验误差、提高试验质量以及对试验结果进行科学合理的分析，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，统计学在临床试验中有着不可缺少的重要作用。

本指导原则以临床试验的基本要求和统计学原理为重点，包含了对临床试验的总体考虑以及试验设计、试验过程和结果分析时的统计学问题，旨在为药品注册申请人和临床试验的研究者在整个临床试验中如何进行设计、实施、分析和评价提供指导，以期保证药物临床试验科学、严谨和规范。

本指导原则主要适用于临床试验的后期（确证性试验），但应尽可能用于临床试验的各个阶段。

二、整个临床试验的基本考虑

（一）探索性试验和确证性试验

药物临床试验的主要目标是寻找是否存在其风险/效益比可接受的安全有效的药物，同时也要确定可能由该药受益的特定对象、使用适应症及

适宜的用法与用量。为达到以上总体目标，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。

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信息。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正式依据。

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检查的随机对照试验，以说明所开发的药物对临床是有益的。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。

(二) 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。观察指标分为测量指标和分类指标。

观察指标必须在设计方案中有明确的定义和可靠的依据，不允许随意修改。

1、主要指标和次要指标

主要指标又称主要终点,是与试验目的有本质联系的，能确切反映药物有效性或安全性的观察指标。通常主要指标只有一个，如果存在多个主要指标时，应该在设计方案中，考虑控制 I 类错误的方法。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的标准。主要指标必须在临床试验前确定，并用于试验样本量的估计。

次要指标是指与试验目的相关的辅助性指标。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。次要指标数目也应当是有限的，并且能回答与试验目的相关的问题。

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。如临床上采用的量表就是一种复合指标。复合指标被用作主要指标时，组成这个复合指标的单个指标如果有临床意义，也可以同时单独进行分析。

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，它通常是有序等级指标。用全局评价指标来评价某个治疗的总体有效性或安全性，一般都有一定的主观成份。如果必须将其定义为主要指标时，应在试验方案中有明确判断等级的依据和理由。全局评价指标中的客观指标一般应该同时单独作为主要指标进行分析。

4、替代指标

替代指标是指在直接测定临床效果不可能时，用于间接反映临床效果的观察指标。替代指标所提供的用于临床效果评价的证据的强度取决于：1) 替代指标与试验目的在生物学上相关性的的大小；2) 在流行病学研究中替代

指标对临床试验结果的预测价值；3) 从临床试验中获得的药物对替代指标的影响程度与药物对临床试验结果的影响程度相一致的证据。

5、测量指标转换为分类指标

根据临床评价的需要，有时需将测量指标转换为二分类或多分类的分类指标，如：根据一个测量指标改变程度等于或超过某一数值时作为分类的定义。分类指标的定义应在试验方案中明确规定。由测量指标转换为分类指标通常会丧失部分信息，由此导致检验效能的降低应当在估计样本量时加以考虑。

(三) 偏倚的控制

偏倚又称偏性，是指在设计临床试验方案、执行临床试验、分析评价临床试验结果时，有关影响因素所致的系统误差，致使疗效或安全性评价偏离真值。偏倚会干扰临床试验得出正确的结论，在临床试验的全过程中均须防范其发生。随机化和盲法是控制偏倚的重要措施。

1、随机化

随机化是使临床试验中的受试者有同等的机会被分配到试验组或对照组中，而不受研究者和/或受试者主观意愿的影响，可以使各处理组的各种影响因素（包括已知和未知的因素）分布趋于相似。随机化包括分组随机和试验顺序随机，与盲法合用，有助于避免因处理分配的可预测，在受试者的选择和分组时可能导致的偏倚。

临床试验中可采用分层、区组随机化方法。分层随机化有助于保持层内的均衡性，特别在多中心临床试验中，中心就是一个分层因素。另外为了使各层趋于均衡，避免产生混杂偏倚，按照基线资料中的重要预后因素

(如病症的严重程度) 等进行分层, 对促使层内的均衡安排是很有价值的。区组随机化有助于减少季节、疾病流行等因素对疗效的影响。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。研究者及其有关人员应对区组的大小保持盲态。

当样本量、分层因素及区组大小决定后, 由试验统计学专业人员在计算机上使用统计软件产生随机分配表。临床试验的随机分配表就是用文件形式写出对受试者的处理安排, 即处理(或在交叉试验中的处理顺序)的序列表。随机分配表必须有可以重新产生的能力, 即当产生随机数的初值、分层、区组决定后能使这组随机数重新产生。

试验用药物应根据试验统计学专业人员产生的随机分配表进行编码, 以达到随机化的要求, 受试者应严格按照试验用药物编号的顺序入组, 不得随意变动, 否则会破坏随机化效果。随机化的方法和过程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。

2、盲法

盲法是为了控制临床试验过程中和解释结果时产生偏倚的措施之一。这些偏倚可能来自于多个方面, 如由于对治疗的了解而对受试者的分组进行选择、受试者对治疗的态度、研究者对安全有效性的评价、对脱落病例的处理以及在结果分析中剔除的数据等。

根据设盲程度的不同, 盲法分为双盲、单盲和非盲。如条件许可, 应尽可能采用双盲试验, 尤其在试验的主要变量易受主观因素干扰时。如果双盲不可行, 则应优先考虑单盲试验。在某些特殊情况下, 由于一些原因

而无法进行盲法试验时，可考虑进行非盲的临床试验。无论是采用单盲或非盲的临床试验，均应制订相应的控制试验偏倚的措施，使已知的偏倚来源达到最小。例如，主要指标应尽可能客观，采用信封随机法入选受试者，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。采用不同设盲方法的理由，以及通过其它方法使偏倚达到最小的措施，均应在试验方案中说明。

盲法的原则应自始至终地贯彻于整个试验之中。双盲临床试验中，从随机数的产生、试验用药物的编码、受试者入组用药、试验结果的记录和评价、试验过程的监查、数据管理直至统计分析，都必须保持盲态。监查员必须自始至终保持盲态。如果发生了任何非规定情况所致的盲底泄露，并影响了该试验结果的客观性，则该试验将被视作无效。

为使双盲临床试验得以顺利实施，还必须注意以下几个问题。

安慰剂：在双盲临床试验中，应保证所提供的安慰剂与所模拟的药物在剂型、外观、气味等方面完全一致，并不含有任何有效成份。

双模拟技术：临床试验中，当试验药和对照药外观不一致时，可为试验药和对照药各准备一种安慰剂，以达到试验组与对照组在用药的外观与给药方法上的一致。这一技术有时也会使用药计划较难实施，以至影响受试者的依从性。

胶囊技术：为达到双盲的目的，可将试验用药（包括试验药、对照药、安慰剂）分别装入外观相同的胶囊。但应首先证明药物在装入胶囊后与原剂型药物生物等效。

药物编盲与盲底保存：由不参与临床试验的人员根据已产生的随机分配表对试验用药物进行分配编码的过程称为药物编盲。随机数、产生随机数的参数及试验用药物编码统称为双盲临床试验的盲底。用于编盲的随机数产生时间应尽量接近于药物分配包装的时间，编盲过程应有相应的监督措施和详细的编盲记录，完成编盲后的盲底应一式二份密封，交临床试验负责单位和药品注册申请人分别保存。

应急信件与紧急揭盲：从医学伦理学方面考虑，双盲试验应为每一个编盲号设置一份应急信件，信件内容为该编号的受试者所分入的组别及用药情况。应急信件应密封，随相应编号的试验用药物发往各临床试验单位，由该单位负责保存，非必要不得拆阅。在发生紧急情况或病人需要抢救必须知道该病人接受的是何种处理时，由研究人员按试验方案规定的程序拆阅。一旦被拆阅，该编号病例将中止试验，研究者应将中止原因记录在病例报告表中。所有应急信件在试验结束后随病例报告表一起收回，以便试验结束后盲态审核。

试验方案中要对严重不良事件以及意外情况的处理作出规定，包括如何紧急揭盲、如何报告等。试验结束时应对破盲的原因、范围和时间作出分析，作为对疗效及安全性评价的参考。

揭盲规定：试验方案中，当试验组与对照组按 1: 1 设计时，一般采用两次揭盲法。两次揭盲都由保存盲底的有关人员执行。数据文件经过盲态审核并认定可靠无误后将被锁定，进行第一次揭盲。此次揭盲只列出每个病例所属的处理组别（如 A 组或 B 组）而并不标明哪一个为试验组或对照组。第一次揭盲的结果交由试验统计学专业人员输入计算机，与数据文件

进行联接后，进行统计分析。当统计分析结束后进行第二次揭盲，以明确各组所接受的治疗。

三、试验设计中的基本考虑

(一) 试验设计的类型

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方案。

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，以比较各处理组间的差异。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，它可以控制个体间的差异，同时减少受试者人数。

最简单的交叉设计是 2×2 形式，对每个受试者安排两个试验阶段，分别接受两种试验用药物，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。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。采用交叉设计时应避免延滞效应，资料分析时需检测是否有延滞效应存在。因此，

每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。

交叉设计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。

如果试验的样本量是基于检验主效应而计算的，则在估计交互作用时，检验效能将降低。

4、成组序贯设计

成组序贯设计常用于下列两种情况：（1）试验药与对照药的疗效相差较大，但病例稀少且临床观察时间较长。（2）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。

成组序贯设计是把整个试验分成若干个连贯的分析段，每个分析段病例数相等，且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（包括有效性和安全性）进行分析，一旦可以做出结论（拒绝无效假设，差异有统计学意义）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。其优点是当处理间确实存在差异时，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 类错误不超过预先设定的水准（比如 $\alpha=0.05$ ）。试验设计中需写明 α 消耗函数的计算方法。

（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，多个单位的研究者合作，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。

多中心试验必须在统一的组织领导下，遵循一个共同制定的试验方案完成整个试验。各中心试验组和对照组病例数的比例应与总样本的比例相同，以保证各中心齐同可比。多中心试验要求各中心的研究人员采用相同的试验方法，试验前对人员统一培训，试验过程要有监控措施。当主要指标可能受主观影响时，需进行统一培训和一致性检验。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，如：统一由中心实验室检验。

在双盲多中心临床试验中，盲底是一次产生的，应按中心分层随机；当中心数较多且每个中心的病例数较少时，可统一进行随机，不按中心分层。

（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。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认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显示试验药的治疗效果在临床上不劣于阳性对照药。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阳性对照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。

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（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对安慰剂的优效性试验所观察到的差异。等效界值或非劣效界值的确定需要由主要研究者从临床上认可，而不是依赖于试验统计学专业人员。试验中所选择的比较类型，应从临床角度考虑，并在制定试验方案时确定下来。通常以阳性为对照的临床试验中，如果要说明试验药物的效果不低于阳性对照药时，多倾向于进行非劣效性检验。

等效性或非劣效性的统计学检验常用可信区间法。等效性检验采用双侧可信区间，当可信区间完全落在等效界值之内，则推断为等效；非劣效性检验应采用单侧可信区间，如果可信区间的下界大于非劣效性检验的下限，则推断为非劣效。

（四）样本量

每个临床试验的样本量应符合统计学要求。

临床试验中所需的样本量应足够大，以确保对所提出的问题给予一个可靠的回答。样本的大小通常以试验的主要指标来确定。同时应考虑试验设计类型、比较类型等。

样本量的确定与以下因素有关，即设计的类型、主要指标的性质（测量指标或分类指标）、临床上认为有意义的差值、检验统计量、检验假设、I类和II类错误的概率等。样本量的具体计算方法以及计算过程中所需用到的统计量的估计值及其依据应在临床试验方案中列出，同时需要提供这些估计值的来源依据。在确证性试验中，样本量的确定主要依据已发表的资料或预试验的结果来估算。I类错误概率常用0.05，II类错误概率应不大于0.2。

（五）资料的收集

临床试验数据的收集和传送，可采用多种形式，目前较为常用的形式为病例报告表。

从试验数据的收集到数据库的完成，均应符合《药物临床试验质量管理规范》（GCP）的规定，尤其是及时的数据记录、错误更正、补遗等。这些步骤均是建立高质量数据库、完成试验计划并达到试验目的所必需的。

四、试验进行中的基本考虑

（一）期中分析

某些临床试验需要进行期中分析。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安全性所作的分析。期中分析的日程、安排、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，分析结果必须保持盲态。研究者仅仅会被告知是否继续试验或需要对试验方案进行修改。

设计不良的期中分析可能使结果有误，所得结论缺乏可靠性。如进行了计划外的期中分析，在研究报告中应解释其必要性、破盲的必要性，提供可能导致的偏倚的严重程度以及对结果解释的影响。

(二) 试验方案的修改

试验方案确定并经伦理委员会批准后，其研究设计一般情况下不宜更改。但在以下两种情况可以考虑修改：

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，在不破盲的条件下修改原入选/排除标准。

当原设计的样本含量是在不确切信息的假设条件下估计的，而期中分析结果表明指标的估计与期望值不符时，应修改假设条件，重新计算样本含量。

对试验方案的任何修改都应在修订方案中写明。修订方案需重新得到伦理委员会的批准。

五、数据管理

数据的正确性对保证临床试验的质量极为重要，因此必须十分重视。认真进行监查及数据管理能及早地发现问题，并可尽量避免问题的发生和再现。

研究者应根据受试者的原始观察记录，保证将数据正确、完整、清晰、及时地载入病例报告表。监查员须监查试验的进行是否遵循试验方案（如检查有无不符合入选/排除标准的病例等），确认所有病例报告表填写正确完整，与原始资料一致，如有错误和遗漏，及时要求研究者改正。修改时需保持原有记录清晰可见，改正处需经研究者签名并注明日期。

经过监查员检查后的病例报告表，需及时送交临床试验的数据管理员。对于完成的病例报告表在研究者、监查员、数据管理员之间的传送应有专门的记录并妥善保存。

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，并在第一份病例报告表送到以前，由数据管理员建立数据库，并保证其完整、正确和安全。数据管理员还应对每一份病例报告表进行初步审核，再交由两名操作人员独立地输入数据库中，并用软件对两份输入结果进行比较。如果有不一致，需查出原因，加以更正。数据管理员按病例报告中各指标数值的范围和相互关系拟定数据检查，如范围检查和逻辑检查等。所有错误内容及修改结果应有详细记录并妥善保存。如有必要，可再次对数据库中的指标（特别是主要指标）进行全部或抽样的人工检查，并与病例报告表进行核对。

数据管理中发现任何问题时，应及时通知监查员，要求研究者作出回答。他们之间的各种疑问及解答的交换应当使用疑问表，疑问表应保存备查。

上述工作完成后，由主要研究者、药品注册申请人、试验统计学专业人员和数据管理员进行盲态审核。盲态审核中确定每个病例所属分析集、

缺失值的处理及离群值的判断等。以上任何决定都需用文件形式记录下来。盲态审核下所作的决定不应该在揭盲后被修改。经盲态审核认为所建立的数据库正确无误后，对数据库进行锁定。此后，对数据库的任何改动只有在以上几方人员均同意（可以书面形式）的情况下才能进行。

数据库锁定后需妥善保存备查，并进行第一次揭盲，同时将盲底和数据库交试验统计学专业人员进行统计分析。

六、统计分析

（一）统计分析计划书

统计分析计划书由试验统计学专业人员起草，并与主要研究者商定，其内容应比试验方案中所规定的要求更为具体。

统计分析计划书上应列出统计分析集的选择、主要指标、次要指标、统计分析方法、疗效及安全性评价方法等，按预期的统计分析结果列出统计分析表备用。

统计分析计划书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，可以修改、补充和完善。但是在第一次揭盲之前必须以文件形式予以确认，此后不能再作变动。

（二）统计分析集

用于统计的分析集需在试验方案的统计部分中明确定义，并在盲态审核时确认每位受试者所属的分析集。在定义分析数据集时，需遵循以下两个原则：①使偏倚达到最小；②控制 I 类错误的增加。

根据意向性分析（简称 ITT）的基本原则，主要分析应包括所有随机化的受试者。即需要完整地随访所有随机化对象的研究结果，但实际操作中

往往难以达到。因此，常采用全分析集进行分析。全分析集（简称 FAS）是指尽可能接近符合意向性分析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出的。在选择全分析集进行统计分析时，对主要指标缺失值的估计，可以采用最接近的一次观察值进行结转。

受试者的“符合方案集”（简称 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，依从性包括以下一些考虑，如所接受的治疗、主要指标测量的可行性以及未对试验方案有大的违反等。将受试者排除在符合方案集之外的理由应在盲态审核时阐明，并在揭盲之前用文件写明。

在确证性试验的药物有效性评价时，宜同时用全分析集和符合方案集进行统计分析。当以上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，应对其差异进行清楚的讨论和解释。如果符合方案集中被排除的受试者比例太大，则会影响试验的有效性分析。

在很多的临床试验中，全分析集方法是保守的，但更接近药物上市后的疗效。应用符合方案集可以显示试验药物按规定的方案使用的效果，但可能较以后实践中的疗效偏大。

对安全性评价的数据集选择应在方案中明确定义，通常安全性数据集应包括所有随机化后至少接受一次治疗的受试者。

（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。

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，研究它们对结果是否不一致以及不一致的直接原因。

(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

临床试验中数据分析所采用的统计分析方法和统计分析软件应是国内外公认的，统计分析应建立在正确、完整的数据基础上，采用的统计模型应根据研究目的、试验方案和观察指标选择，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、可信区间和假设检验

参数估计、可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。试验方案中，应当说明要检验的假设和待估计的处理效应、统计分析方法以及所涉及的统计模型。处理效应的估计应同时给出可信区间，并说明估计方法。假设检验应明确说明所采用的是单侧还是双侧，如果采用单侧检验，应说明理由。

3、协变量分析

评价药物有效性的主要指标除药物作用以外，常常还有其它因素的影响，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，这些因素在统计学中可作为协变量处理。在试验前应认真识别可能对主要指标有重要影响的协变量及如何进行分析以提高估计的精度，补偿处理组间由于协变量不均衡所产生的影响。

在多中心临床试验中，如果中心间处理效应是齐性的，则在模型中常规地包含交互作用项将会降低主效应检验的效能。因此对主要指标的分析如采用一个考虑到中心间差异的统计模型来研究处理的主效应时，不应包含中心与处理的交互作用项。如中心间处理效应是非齐性的，则对处理效应的解释将很复杂。

（六）安全性评价

临床试验中，安全性评价是非常重要的一个方面。在临床试验的早期，这一评价主要是探索性的，且只能发现常见的不良反应；在后期，一般可通过较大的样本进一步了解药物的安全性。后期的对照试验是一个重要的以无偏倚的方式探索任何新的潜在的药物不良反应的方法。

为了说明在安全性和耐受性方面与其他药物或该药物的其他剂量比较的优效性或等效性，可设计某些试验。这种评价需要相应的确证性试验的

支持，这与相应的有效性的评价要求是相同的。

药物安全性评价的常用统计指标为不良事件发生率和不良反应发生率。当试验时间较长、有较大的退出治疗比例或死亡比例时，需用生存分析计算累计不良事件发生率。用于评价药物安全性和耐受性的方法以及度量准则依赖于非临床研究和早期临床研究的信息、该药物的药效学和药代动力学特性、服药方法、受试者类型以及试验的持续时间等。而构成安全性评价的资料则主要来源于不良事件的临床表现、实验室检查等。

从受试者中收集的安全性和耐受性变量应尽可能全面，包括受试者出现的所有不良事件的类型、发生时间、严重程度、处理措施、持续的时间、转归以及药物剂量与试验用药物的关系。

所有的安全性指标在评价中都需十分重视，其主要分析方法需在研究方案中指明。无论是否认为与处理有关，所有的不良事件均需列出。在安全性评价中，研究人群的所有可用资料均需考虑。实验室应提供检查指标的度量单位以及参考值范围，毒性等级也必须事先确定，并说明其正确性。

在大多数的试验中，对安全性与耐受性的评价常采用描述性统计方法对数据进行分析，必要时辅以可信区间以利于说明。

七、统计分析报告

试验统计学专业人员写出的统计分析报告是提供给主要研究者作为撰写临床试验总结报告的素材。

试验统计学专业人员根据确认的统计分析计划书完成统计分析工作，在统计分析报告中首先简单描述临床试验的目的、研究设计、随机化、盲法及盲态审核过程、主要指标和次要指标的定义、统计分析集的规定等。

其次对统计分析报告中所涉及的统计模型，应准确而完整地予以描述，如选用的统计分析软件、统计描述的内容、对检验水准的规定、以及进行假设检验和建立可信区间的统计学方法。如果统计分析过程中进行了数据变换，应同时提供数据变换的基本原理以及变换数据的理由和依据。统计分析结论应使用精确的统计学术语予以阐述。最后，应按照统计分析计划书设计的统计分析格式详细给出统计分析结果。

对药物有效性评价应给出每个观察时间点的统计描述结果。列出检验统计量、P 值。例如，两个样本的 t 检验的结果中应包括每个样本的数量、均值和标准差、中位数，最小和最大值、两样本比较的 t 值和 P 值；用方差分析进行主要指标有效性分析时，应考虑治疗、中心和分析指标基线值的影响，进行协方差分析。对于交叉设计资料的分析，应包括治疗顺序资料、每个阶段开始时的基线值、洗脱期及洗脱期长度、每个阶段中的脱落情况、还有用于分析治疗、阶段、治疗与阶段的交互作用方差分析表。

药物的安全性评价主要以统计描述为主，包括用药情况（用药持续时间、剂量、药物浓度）、不良事件发生率及不良事件的具体描述（包括不良事件的类型、严重程度、发生及持续时间、与试验药物的关系）；实验室检验结果在试验前后的变化情况；发生的异常改变及其与试验用药物的关系及随访结果。

八、名词解释

I 类错误 (The Type I error): 错误的拒绝无效假设，常用 α 表示。

II 类错误 (The Type II error): 错误的拒绝无效假设，常用 β 表示。

安全性数据集: 安全性与耐受性评价时，用于汇总的受试者集称为安

全性数据集。安全性数据集应包括所有随机化后至少接受一次治疗的受试者。

等效性试验 (Equivalence Trial): 是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。通常显示真正的处理差异是在临床上可以接受的等效性的上下限之间。

多中心试验 (Multicentre Trial): 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): 是显示试验药的治疗效果在临床上不劣于对照药的试验。

符合方案集 (Per Protocol Set): 又称有效病例、有效样本、可评价病例样本。是由充分依从于试验方案的病例子集所产生的数据集, 是全分析集的一个子集。依从性包括以下一些考虑, 如: 所接受的治疗、主要指标测量的可行性以及未对试验方案有大的违反等。

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。

结转 (last observation carried forward): 是对临床试验中有效性指标缺失值的一种估计方法, 即采用缺失值之前最接近一次的观察数据来代替缺失值。

盲态审核 (Blind Review): 在最后一份病例报告表输入数据库后, 第一

次揭盲之前对数据保持盲态的预分析审核，以便对统计分析计划作最后的决定。

偏倚 (Bias): 是指在设计临床试验方案、执行临床试验、分析评价临床试验结果时，有关影响因素所致的系统误差，致使疗效或安全性评价偏离真值。

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和安全性所作的分析。

全分析集 (Full Analysis Set): 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出的。

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，它通常是有序分类指标 (scale of ordered categorical ratings)。

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的生物统计学专业人员。

双模拟 (Double-Dummy): 是在临床试验中，当两种处理（如药物的剂型、给药方法等）不能做到相同时，使试验保持双盲的一种技术。即为试验药与对照药各准备一种安慰剂，以达到试验组与对照组在用药的外观与给药方法上的一致。

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观察指标。

统计分析计划 (Statistical Analysis Plan): 是包括比方案中描述的主要分

析特征更加技术性和更多详细细节的文件，并且包括了对主要和次要变量及其他数据进行统计分析的详细过程。

脱落 (Drop out): 是指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的受试者。

意向性分析原则 (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。其结果是随机到每一个处理组的病人即应作为该组的成员被随访、评价和分析，无论他们是否依从计划的处理过程。

优效性试验 (Superiority Trial): 是显示试验药的治疗效果优于对照药 (安慰剂或阳性对照药) 的试验。

九、参考文献

1. FDA: Guideline for the Format and Content of The Clinical and Statistical Sections of an Application 1988
2. EMEA: Biostatistical Methodology in Clinical Trials 1993
3. MHLW: Guideline for the Statistical Analysis of Clinical Trial 1992
4. ICH E9: STATISTICAL PRINCIPLES FOR CLINICAL TRIALS 1998

十、附录

十一、著者

《化学药物和生物制品临床试验的生物统计学技术指导原则》课题研究组