国际人用药品注册技术协调会

ICH 三方协调指导原则

临床试验中对照组的选择和相关问题

E10

现行第 4 阶段版本

2000 年 7 月 20 日

根据 ICH 进程，本指导原则由相应的 ICH 专家工作组制定，并已向监管机构征求意见。在进程的第 4 阶段，建议欧盟、日本和美国的监管机构采纳最终草案。

E10

文件历史

首次编码	历史	日期	重新编码 2005 年 11 月
E10	由指导委员会根据第二阶段程序批准，并公开征求意见。	1999 年 5 月 7 日	E10

目前第四阶段版本

E10	由大会根据第四阶段程序批准，并建议三个 ICH 监管实体采纳。	2000 年 7 月 20 日	E10

临床试验中对照组的选择和相关问题

ICH三方协调指导原则

本指导原则经2000年7月20日召开的ICH指导委员会会议上确认已达到ICH进程第四阶段，建议ICH三个监管机构采纳本指导原则。

目录

1.引言 1
1.1 指南的总体框架和目的 1
1.2 对照组的目的 3
1.2.1 随机化 4
1.2.2 盲法 4
1.3 对照的类型 5
1.3.1 安慰剂平行对照 6
1.3.2 无治疗平行对照 7
1.3.3 剂量-效应平行对照 7
1.3.4 活性(阳性)平行对照 8
1.3.5 外部对照(包括历史对照) 8
1.3.6 多个对照组 9
1.4 临床试验的目的及相关问题 9
1.4.1 有效性的证据 9
1.4.2 可比的有效性和安全性 10
1.4.3 比较的公平性 11
1.4.3.1 剂量 11
1.4.3.2 病人人群 11
1.4.3.3 终点选择和时间确定 12
1.5 检测灵敏度 13
1.5.1 非劣效/等效性试验的检测灵敏度 14
1.5.1.1 药效灵敏度的历史性证据和非劣效界值的选择 15
1.5.1.2 试验良好的实施 19
1.5.2 优效性临床试验的检测灵敏度 21
2.对照类型的详细考虑 21

2.1 安慰剂对照 21
2.1.1 描述(参阅 1.3.1) 21
2.1.2 减少偏倚的能力 23
2.1.3 伦理学问题 23
2.1.4 安慰剂对照研究的用途和在特定情况下推断的正确性 24
2.1.5 设计方案的修改以及结合其他对照药解决伦理的、实践的或推断问题 25
2.1.5.1 其他对照组 26
2.1.5.1.1 三臂研究；安慰剂和活性对照 .. 26
2.1.5.1.2 其他剂量 26
2.1.5.1.3 析因/联合研究 26
2.1.5.2 研究设计的其他改变 27
2.1.6 安慰剂对照试验的优点 31
2.1.6.1 能可靠地证明药物的有效性 31
2.1.6.3 高效率 32
2.1.6.4 最小化受试者和研究者期望值的影响 . 32
2.1.7 安慰剂对照试验的缺点 32
2.1.7.1 伦理方面的问题(见 2.1.3 和 2.1.4) .. 32
2.1.7.2 病人和医师的顾虑 33
2.1.7.3 通用性 33
2.1.7.4 无比较信息 34
2.2 无治疗平行对照(见 1.3.2) 34
2.3 剂量-效应平行对照(见 1.3.3) 34
2.3.1 描述 34
2.3.2 减少偏倚的能力 36
2.3.3 伦理学问题 36
2.3.4 剂量-效应研究的用途和特定情况下推断的正确性 36
2.3.5 修改设计和结合其他对照解决伦理的、实践的或推断问题 36
2.3.6 剂量-效应试验的优点 37
2.3.6.1 高效率 37

2.3.6.2 可能的伦理方面优点 37
2.3.7 剂量-效应研究的缺点 37
2.4 活性药物对照(见 1.3.4) 38
2.4.1 描述 38
2.4.2 减少偏倚的能力 39
2.4.3 伦理学问题 39
2.4.4 活性对照试验的应用及特定情况下推断的正确性 40
2.4.5 修改设计和结合其他对照解决伦理的、实践的问题或推断问题 40
2.4.6 活性对照试验的优点 41
2.4.6.1 伦理和实践的优点 41
2.4.6.2 信息含量 41
2.4.7 活性对照试验的缺点 42
2.4.7.1 信息含量 42
2.4.7.2 样本量大 42
2.5 外部对照(包括历史对照,见 1.3.5 部分) 42
2.5.1 描述 42
2.5.2 减少偏倚的能力 43
2.5.3 伦理学问题 45
2.5.4 外部对照试验的用途以及在特定情况下推断的正确性 46
2.5.5 修改设计和结合其他对照解决伦理的、实践的问题或推断问题 47
2.5.6 外部对照试验的优点 47
2.5.7 外部对照试验的缺点 48
3.选择平行对照组 48
表 1 不同情况下各种平行对照类型的用途 50
图 1 为描述有效性选择平行对照 51

临床试验中对照组的选择和相关问题

1.引言

在设计临床试验时，选择对照组一直是一个关键性的决定。这一选择对试验的许多方面都会产生影响，包括：试验所能得出的结论，试验伦理的可接受性，研究实施和分析过程中把偏倚减至最小的程度，受试者类型和招募速度，研究终点的选择，研究结果的科学性和公信度，监管当局对结果的接受度，以及研究实施和解读的其他方面。

1.1 指南的总体框架和目的

本指南的目的是描述临床试验中为证明治疗有效性选择对照组的一般原则，并讨论与此有关的试验设计和实施中的一些问题。本指南并非任何地区的管理要求，而仅描述采用各种不同设计的试验所能说明的问题。。该指南中所描述的一般原则与所有对照试验有关，而对照组的选择对于药物在研发过程中用以证明其有效性而开展的临床试验来说尤为重要。在选择对照组时，应当考虑现行的标准治疗，有足够的证据支持所选方案，同时要考虑伦理方面的因素。

指南首先描述了设置对照组的目的和为证明有效性通常采用的对照类型。之后探讨了使用活性对照试验（通过表明相对于对照组非劣效或等效来证实有效性）的关键设计和问题阐述 (参阅 1.5)。在某些情况下，非劣效的结果不能被解读为有效性的证据。特别是，非劣效的结果如果被解释为

有效的证据，该试验必须能够区分有效与低效或无效治疗。

指南比较详细地描述每一种对照 (参阅 2.~2.5.7)，并就下列问题加以讨论:

· 减少偏倚的能力;

· 与之应用有关的伦理学问题和实践问题;

· 在特定情况下的用途和推断质量;

· 研究设计的修改或与其他对照结合以解决伦理、可操作性或推断问题;

· 总的优点和缺点。

下列 ICH 指南与本指南密切相关:

E3: 临床研究报告的结构与内容

E4: 支持药物注册的量效信息

E5: 伦理因素

E6: 临床试验管理规范

E8: 临床试验的一般考虑

E9: 临床试验统计学指南

尽管采用该指南中描述和讨论的任何一种对照组的试验，至少在某些情况下可作为注册用临床试验，并被认可，但并不同时适用于所有情况。在第 3 部分、图 1 和表 1 中将列出选择对照类型的一般方法。

尽管指南主要针对的是有效性评价的临床试验，所探讨的许多方面也适用于评估特定的安全性假设，以及两种治疗

安全性或有效性的比较。

1.2 对照组的目的

对照组有一个主要目的，就是可以将试验治疗给病人带来的结果(如：症状、体征或其他发病情况的改变)与其他因素（如疾病的自然进展、观察者或者病人的期望或其他治疗措施）造成的结果区分开来。对照组的结果可以告诉我们，假如病人没有接受试验治疗或者接受另外一种已知有效治疗会发生什么情况。

假如在特定的病人人群中疾病的进程是相同的，或者是根据病人的特征就可以可靠地预测任何特定受试者或受试组的结果，那么，治疗的结果就可以简单地与未接受治疗的已知结果相比较。比如，可以假定疼痛会持续一定时间，血压不会变化。抑郁会持续一定的时间，肿瘤会进展，急性梗死后的死亡率与以前所观察到的一样，等等。事实上，仅在少数情况下，疾病的病程在特定人群中是可以预测的，并且可以使用以往研究过的类似的病人组作为“历史对照” (参阅 1.3.5)。然而，在多数情况下，需要有一个平行的对照组，因为不可能预测出足够精确的研究结果。

平行的对照组是从与试验组相同的人群中选出的，并且作为研究试验治疗的同一试验的一部分，在相同的时间段，按规定的方法接受治疗。试验组和对照组的所有基线值，以及除了试验治疗外的其他可能影响结果的有关变量都应当

相似。达不到这样的相似性就可能在研究中引入偏倚。这里所说的偏倚(以及在 E9 使用的偏倚)，是指在临床试验的设计、实施、分析和结果的解释方面出现的系统倾向，从而使对治疗结果的估计偏离它的真实值。随机化和盲法是常用的两种技术，用以减少这种偏倚的发生，并保证试验组和对照组在开始研究时彼此相似、在研究过程中得到的处置也相似(参阅 ICH E9)。试验设计是否包括这些特性是决定其质量和说服力的关键因素。

1.2.1 随机化

将单一样本人群随机分入试验组或对照组能最好地确保两组的受试人群相似。随机化可以避免那些可能影响结果的组间已知或未知基线变量的系统差异。不能消除组间系统差异是没有平行随机对照研究的一个主要问题(参阅 1.3.5，外部对照试验)。随机化也为统计推断提供了坚实的基础。

1.2.2 盲法

试验过程中，除了接受试验药物和对照药物之外，各组不仅基线应当相似，处理和观察也应相似。临床试验通常是“双盲”的，是指受试者与研究者，以及申办者或参与治疗或临床评估的其他临床试验人员，都不清楚每个受试者分配的治疗。盲法可以最大限度地减少由于受试者或研究人员了解分配的治疗后引起的在管理、治疗、对病人的评估或解释结果时出现的偏倚。例如:

·活性药物组的受试者，假如知道他们服用是活性药物，可能因为他们期望得益于治疗而报告有利的结果，或者更愿意留在研究中。

· 观察者可能不太会判断和报告非治疗组的治疗反应，或者对接受活性药物治疗病人的有利结果或不良事件更加敏感。

·了解治疗的分配情况可能会影响对研究或随访数据的收集。

·了解治疗的分配情况可能会影响对病人是否应继续接受治疗、接受伴随用药或其他辅助治疗的决定。

·了解治疗的分配情况可能会影响是否将某一受试者的结果纳入分析的决定。

· 了解治疗的分配情况可能会影响统计分析方法的选择。

盲法的目的是为了确保主观评价和决定不会因了解治疗分配而受到影响。

1.3 对照的类型

根据以下两个关键的特性可以对临床试验中的对照组进行分类：(1)所接受治疗的类型；(2)决定谁将进入对照组的方法。对照治疗类型可以是下列四种中的任何一种：(1)安慰剂、(2)无治疗、(3)不同剂量或方案的治疗、(4)不同的活性药物。决定谁将进入对照组的主要方法是随机化或选择

一个有别于治疗人群的对照人群(外部的或历史对照)。本文将对照组划分为五类。前四类为平行对照(对照组与试验组从同一人群中挑选并同时进行治疗)，通常是将受试者随机分配到各组，以不同对照治疗类型（前文所列）进行分类。不论接受何种比较治疗，外部(历史的)对照都归为第五类，出于对保证试验组与对照组可比性及减少重要偏倚的认真考虑，这种设计只能在特殊条件下使用。

研究中选用一种以上对照组的情况越来越普遍。每一类对照适合于某些情况，但没有一种可以用于或适于所有的情况。这五种类型分别是:

1.3.1 安慰剂平行对照

在安慰剂对照试验中，受试者被随机分配到试验治疗组或外表完全相同的不含试验药物组。药物可被滴定到有效水平或耐受水平，也可以给予一个或多个固定剂量。这类试验几乎总是双盲试验。该对照的名称表明其目的是控制“安慰剂”效应 (受试者的改善是因为他或她认为自己是在服药)，但这不是它唯一或主要的优势。更准确地说，安慰剂平行对照设计，通过采用盲法、随机化以及设立一个无治疗组，控制了除试验药物药理作用之外的所有对疾病实质和表现的潜在影响。这些影响包括自发性的改变(疾病的自然病程和回归均数现象)、受试者或研究者的期望、临床试验效应、使用其他治疗以及诊断或评估中的主观因素。在研究有效性

时，安慰剂对照试验试图发现治疗间的差异，但在评价安全性指标时，可能试图证明没有(指定大小的)差异。在这种情况下，如果有差异，试验是否能够显示差异就成为关键性的问题（见 1.5）。

采用安慰剂对照组并不是说对照组不接受治疗。在许多安慰剂对照试验中，新的治疗方法和安慰剂是加在通常的标准治疗之上的（所谓的“加载”研究，见 2.1.5.2.1 部分）。

1.3.2 无治疗平行对照

在无治疗的对照试验中，受试者被随机分配到试验治疗组或无治疗组。这种设计和安慰剂对照试验的主要区别是，治疗分配对受试者和研究者都是公开的。因为双盲设计的优点，这种设计很可能仅仅是在下列情况下才需要和适用：难以或不可能实行双盲(例如，容易识别毒性的治疗)；以及有理由确信研究终点是客观的，而且研究结果不大可能受 1.2.2 所列因素的影响。值得注意的是，即使整个试验不是双盲的，但通常仍可能运用盲态的终点评价。这是一种有价值的方法，在不能进行盲法的研究中应该考虑它，但它不解决知道治疗分配所带来的其他问题(参阅 1.2.2)。

1.3.3 剂量-效应平行对照

随机化的、固定剂量的剂量-效应试验中，受试者被随机分配到几个固定剂量组中。受试者可能一开始就接受其固定剂量，或逐渐增加到这一剂量，而要进行的组间比较是最

终剂量基础上的。剂量-效应试验通常是双盲的。试验可能包括一个安慰剂(零剂量)和/或活性对照。在浓度控制试验中，治疗组的剂量被滴定到数种固定的浓度窗；这种类型的试验在概念上类似于固定剂量的剂量-效应试验。在给药方案对照试验中，受试者被随机分配到两个或多个给药方案组中（如，每天 1 次与每天 2 次的比较，3 天与 7 天的比较）。

1.3.4 活性(阳性)平行对照

在活性对照(或阳性对照)试验中，受试者被随机分配到试验治疗组或活性对照组。这种试验通常是双盲的，不过有时不可能做到双盲；例如，许多肿瘤学研究，由于存在不同的给药方案、不同的给药途径(参阅 1.3.2)以及不同的毒性，双盲被认为是困难或不可能的。活性对照试验在证明有效性方面有两个不同的目标: (1)显示试验治疗的有效性与某种已知有效的治疗一样好; 或者(2)显示试验治疗的有效性优于活性对照。这种活性对照也可用于以比较两种治疗的有效性和/或安全性为主要目的的研究(参阅 1.4)。不管试验目的是证实新治疗的有效性还是对两种治疗的比较，试验是否能够区分有效、低效和无效治疗是关键问题。(参阅 1.5)。

1.3.5 外部对照(包括历史对照)

外部对照试验是将接受试验治疗的一组对象与本研究以外的一组病人进行比较，而不是与分配到不同治疗组的相同人群病人组成的内部对照组进行比较。外部对照可以是早

些时候接过治疗的一组病人（历史对照）或是在同一时间而在不同环境下治疗的一组病人。外部对照可能是限定的(特定的病人组)或非限定的（根据对结果的一般医学认知确定的对比组）。利用后一种比较组是十分不可靠的(这种试验通常被认为是非对照的)，因为一般印象中这种试验常常不准确。所谓的基线对照研究是指将受试者接受治疗后的状况与治疗前的状况相比较(例如血压、肿瘤大小)，没有内部对照并且是非对照或是外部对照（见 2 .5 部分）。

1.3.6 多个对照组

正如以下的进一步介绍(参阅 1.5.1)，在一项研究中采用一种以上的对照，通常是可行的而且有利。例如，采用活性对照和安慰剂。同样，试验可采用多种剂量的受试药物以及多种剂量的活性对照药物，可以有或没有安慰剂。这种设计也可用于两种相对强度尚未确定的活性药物的比较，或者是以确定相对强度为目的的试验。

1.4 临床试验的目的及相关问题

应该区分临床试验的两个目的: (1)评价一种治疗的有效性和/或安全性；(2)评价两种治疗的相对有效性、安全性、获益/风险关系或实用性。

1.4.1 有效性的证据

采用任何类型对照的研究，可能通过显示试验治疗优于对照(安慰剂、无治疗对照、低剂量试验药物、活性药物)来

证明试验治疗的有效性。此外，在某些情况下，活性对照试验可能通过显示新治疗与已知有效的治疗在有效性方面相似而证明其有效。然而，通过相似性来证明试验治疗有效仅在假设活性对照在试验条件下有效的前提下成立，因为如果两种药物在试验中都无效，彼此也可以有相似的表现（见 1.5 部分）。

旨在通过显示一个新药同一个标准治疗药物的有效性相似来证明其有效性的临床试验称为“等效性试验”。这些试验多数实际上是非劣效性试验，即试图通过超过一个规定量，即通常所称的界值，来显示新药的有效性不差于对照药物。

1.4.2 可比的有效性和安全性

在某些情况下，试验的重点是一种治疗与另一种治疗的比较，而不是试验药物本身的有效性。根据治疗的领域，可以将这些试验看成是为评价相对获益/风险提供所需的重要信息。所用的活性比较药必须能被研究数据将应用的区域所接受。并不需要证明其优于活性对照药物， 同时，根据情况，也可能并不需要证明其非劣效。例如，一种有效性较差的治疗可能在安全性方面有优势，因而被认为是有用的。

即使这种试验的重点在于对治疗进行比较，而不是证明有效性，但也需将实施和解释非劣效性试验的注意事项加以考虑(参阅 1.5)。特别是，比较性试验需要具备将两种治疗间

存在的差异检测出来的能力，因为一项试验若不能将事实上存在差异的两种治疗加以区别的话，是不能提供有用的比较性信息的。

1.4.3 比较的公平性

此外，对于可获取相对安全性和/或有效性资料的比较性试验需要公平，也就是说，试验条件不应不恰当的偏向于某一种治疗。实际上，为提供有效性证据而进行的活性对照等效或非劣效性试验，也几乎总是需要与对照作公平的比较，因为任何对于研究中的对照是否具有其通常作用的怀疑，都会削弱对试验分析灵敏度的信心(参阅 1.5)。在研究设计中，可能会偏向性地有利于其中一个治疗组的因素包括：剂量或病人人群的选择以及终点的选择和时间的确定。

1.4.3.1 剂量

在比较试验药物与活性对照药物时，为对照药和试验药物选择一个适当的剂量和给药方案是重要的。在检查两种治疗的比较结果时，对以下的问题予以考虑是重要的：是否使用了有效性显然较差而且剂量又太低的治疗药物，或者是否使用了耐受性显然较差而且剂量又太高的治疗药物。在某些情况下，为了能显示令人信服的较好的有效性或安全性，需要对对照药物和/或试验药物进行多种剂量的研究。

1.4.3.2 病人人群

活性对照试验受试者的选择会影响结果；在评价试验说

明什么问题时要仔细考虑所研究的人群。例如，假如受试者是来自既往对照治疗无反应的人群，那么就会产生有利于新治疗的偏倚。这样的试验结果不能推广到过去未经治疗的病人人群。不过，优效性结果仍然是证明新治疗在研究人群中有效性的证据。事实上，在对其他治疗无反应的病人中进行新治疗研究，将无反应者随机分配到新治疗组或无效治疗组(只要这样不会使病人出现风险)，能够证明新治疗在这些无反应者中的价值，也是临床上有价值的观察。

同样，有时可将对某一特定药物有或多或少的有利反应或不良反应的病人分为亚群。例如，黑人对β阻断剂和血管紧张素转化酶抑制剂的降压作用反应不好，所以在这些病人中，用这些药物与新的降压药进行比较，就会倾向于显示新药的优越性。这样就不宜得出新药一定优越的结论。不过，在认识到其局限性及可能得出合适的结论的前提下，在设计好的亚组中进行的研究仍可以提供丰富的信息。

1.4.3.3 终点选择和时间确定

当两种治疗用于同一种疾病或病情的时候，特别当它们是属于不同种类或不同方式的两种治疗时，它们可能对这种疾病的结果产生不同的影响。因此，假如在一项临床试验中对它们进行比较，不同的终点选择和时间确定就可能对其中某种治疗有利。例如，对于急性心肌梗死的病人，溶栓药可以降低死亡率，但会增加出血性卒中的危险。假如将一种新

的更加有效的溶栓药与一种比较老的溶栓药作比较，如果终点是死亡率，则新的溶栓药会显得更有效；但是，如果终点是由死亡率和造成失能的卒中组合时，新药将差于老药。类似的情况，在评价两种镇痛药对牙痛止痛效果时，若将终点观察时间侧重于早期止痛效果，起效更快的药物就会更加有利，而评价时间点侧重于较晚时的止痛效果时，则对镇痛作用更持久的药物有利。

1.5 检测灵敏度

检测灵敏度是指临床试验区分有效治疗与低效或无效治疗的能力。检测灵敏度对于任何临床试验都是重要的，但是对于旨在表明治疗之间差异的试验（优效性试验）与旨在表明非劣效性的试验来说，其潜在意义是不一样的。如果试验想要通过显示试验治疗与对照相比的优效性来证明其有效性，而缺少检测灵敏度，则将不能表明试验治疗的优效性，也就不能得出有效性的结论。反之，若试验想要通过显示试验治疗与活性对照相比的非劣效性来证明其有效性，而缺少检测灵敏度，则可能会得到一个显示非劣效性的无效治疗，由此导致错误的有效性结论。

如果同一试验中的两个治疗组显示出不同的有效性（即，一个治疗组显示优效时），则这一结果本身已说明该试验有检测灵敏度。反之，一个成功的非劣效性试验（即，一个治疗组显示出了非劣效性），或者是一个不成功的优效

性试验，一般不包含检测灵敏度的如此直接的证据。

1.5.1 非劣效/等效性试验的检测灵敏度

非劣效或等效试验的检测灵敏度可从两个因素推导出来:

1）药效灵敏度的历史性证据，即，既往可区分有效治疗与低效或无效治疗的、相似设计的试验，和

2）试验良好的实施，即，试验的实施不会削弱它区分有效治疗与低效或无效治疗的能力。

在开始非劣效性试验前，能够并应当评估药效灵敏度的历史性证据。特别是应当确定，在进行研究的特定治疗领域里，使用特定的活性对照药、或其他有相似作用的治疗的良好设计和进行的试验可以可靠地显示有效性。最理想的情况是，准备用作活性对照的活性治疗已经被有力地证实优于安慰剂。如果情况如此，就有了类似设计的临床试验药效灵敏度的历史性证据（见 1.5.1.1）。

临床研究的良好的实施只有在活性对照非劣效性试验完成后才能进行充分的评估。非劣效性试验不仅在设计上应当类似于决定药效灵敏度历史证据的临床试验（如，入选标准、允许的伴随治疗）；而且，实际入选的受试者、伴随治疗等都应当进行评价，以保证试验的实施在事实上类似于以往的试验。试验还应以高质量来实施（如，良好的依从性，很少的失访）。药效灵敏度的历史性证据和良好的试验实施

（1.5.1.2）共同保证新活性对照试验的检测灵敏度。

因此，非劣效性试验的设计和实施包含了一下四个关键步骤:

*1）*确定存在药效灵敏度的历史性证据。如果这一点不确定，通过非劣效性证明药物有效性是不可能的，且不可尝试使用。

*2）*设计试验。试验的重要设计细节应当类似于决定药效灵敏度历史证据存在的临床试验，如，研究人群，伴随用药，研究终点，导入期。

*3）*选择界值。应当确定一个可接受的非劣界值，并将历史性的数据和相关的临床的和统计学的因素考虑进去。

*4）*进行试验。试验应当严格按照历史性的试验进行，并且应当是高质量的。

1.5.1.1 药效灵敏度的历史性证据和非劣效界值的选择

如前所述，大多数活性对照药的“等效”试验，其实是旨在确定新治疗有效性的非劣效性试验。非劣效性试验结果的分析，在 ICH 文件 E9 和 E3 中加以讨论。简而言之，这种试验是将新疗法同已经确定的疗法作比较。在试验开始以前，要选定一个等效或非劣效的界值，有时称为“δ”。这一界值是划分受试药与对照药的差别是否达到统计学差异的程度。如果试验治疗与对照治疗之间差值的置信区间排除了大于等于界值的劣效等级，就可以宣布试验治疗非劣效于对

照治疗；假如置信区间包含了界值范围，则不能得到试验治疗非劣于对照治疗的结论。

给非劣效性试验选择的界值不能大于在计划进行试验的条件下与安慰剂相比活性对照药有把握取得的最小有效性。假如活性对照药与新药之间有效性的差异等于或超过界值，而且有利于对照药，则新药可能完全无效。所期望的活性药物可靠的最小作用，只有在存在药效灵敏度的历史性证据时才可能确定。界值往往是根据既往在与计划的新试验相似的条件下、经过恰当设计的安慰剂对照试验的经验来确定的，但也可以通过剂量反应或活性对照的有效研究获得。不管这些早期研究采用的何种对照组，在确定界值时，使用的是活性治疗相对于对照的优效测量值，而不是非对照的测量值，例如自基线的变化。必须注意，本文件并没有叙述如何准确地计算界值，而且有较少的计算界值的经验。

界值的确定要根据统计学推理和临床判断，要反映出使用证据的不确定性，而且应当适当保守。假如能恰当地做到这一点，当选定的界值不在活性对照药差异的可信限范围之内时，就可以保证新药的有效性大于零。实际上，选定的界值往往比表示活性对照药最小期望有效性要小，因为这样可以使保留部分临床可以接受的有效性(或活性对照药的部分有效性)。例如，在死亡率的非劣效研究中，一般认为证明试验治疗效果大于零并不足够；通常还会寻求保留影响死亡

率的一些主要部分的对照组。于主要目的是证明受试药与活性对照药治疗上等效(参阅 1.4.2)的试验，通常为了证明受试药物与对照药物十分相似，而不是简单地证明新药有任何有效性。

界值的选定只有根据以往的经验这一事实，说明非劣效性试验与历史性对照(外部对照)试验有一个共同的要素。这种研究设计只有在对预期药物有效性的历史性估计，能够得到对照药的早先研究结果文献的充分支持时才是适用和可靠的。这些研究应当可以得出结论，在设计相似的恰当规模的试验中，活性对照药总是能够与安慰剂相区别，同时应当可以测出对照药的预期最小有效性的大小。假如类似设计的安慰剂对照试验不是偶尔显示活性对照药与安慰剂无差异，而且不能用研究中的某些特点来加以解释，那么唯一可能的解释就是受试药物的优效。

根据历史经验得出的药效灵敏度是否适用于任何情况，从某种程度上看是一个判断的问题。在某些情况下，由于之前安慰剂对照试验结果的一致性，或是疾病治疗与不治疗的结局差别很大，药效的灵敏度十分清楚或显而易见。例如在很多感染性疾病的短期研究中，有效治疗的治愈率远远超过自发治愈率。不过，在许多情况下，在有良好对照的研究中，被认为有效的药物并不是都能显得优于安慰剂，因此，不能可靠地确定在某一试验中药物可能具有的最小有效性。这种

情况往往包括安慰剂组出现实质性的改善和改变,和/或治疗作用小或易变，例如抑郁、焦虑、痴呆、心绞痛、症状性充血性心力衰竭、季节性过敏以及症状性胃食管反流疾病。

对所有这些情况，标准治疗无疑是有效的，因为对这些已经证明有效的药物中的每一种都进行过许多有良好对照的研究。然而，根据已有经验难以描述药物肯定会显示最小有效性的研究条件(即有药效灵敏度的条件)，以及能用来确定适当界值的条件。在某些情况下，赖以预测药效灵敏度的经验可能并不可靠，例如随着时间的推移，诊断标准和治疗标准发生了重大改变（例如，详见 2.1.7.1）。如果使用活性对照或非劣效设计，而不能为选定非劣效界值的研究的药效灵敏度历史证据提供充分支持，那么非劣结果并不能为有效性提供有用的信息。

值得注意的是，有效性灵敏度历史证据的确定只适用于一种特别设计的试验。对于一个计划的要获得相似药效灵敏度的非劣效试验，该试验具有与历史试验相似的关键设计特点是非常重要的。这些设计特性包括，如，入选标准（病情的严重程度、伴随的疾病状况、诊断方法），对照药物的剂量和剂型，采用的伴随治疗，终点的测量和评价时间，为排除特定病人群所采用的清洗期。当试验设计特征上不可避免差别时（如，由于技术或治疗的进步），应当仔细考虑采用不同的药效灵敏度历史证据和选择界值。

1.5.1.2 试验良好的实施

即使有药效灵敏度的历史证据，新的试验也与既往研究设计类似，实际试验执行过程也可能削弱对灵敏度的评价。为确保试验的分析灵敏性，应当保证高质量地实施试验，且实际入组的病人、实际给予的治疗（除试验治疗之外的）、实际的评价指标都应当与所依据的药效敏感性的历史试验类似。

在试验进行中，有许多因素能减低在有效治疗、有效性差和无效治疗中被观察到的差别，从而降低试验的检测灵敏度，例如:

1.对治疗的依从性差。

2.研究人群对药物作用的反应差。

3.同时应用的其他药物，或其他治疗干扰了研究药物，或减低了研究药物可能反应的程度。

4. 一部分受试人群有疾病自发性改善的倾向，因此没有进一步的药物引起的改善空间。

5.诊断标准不明确(病人没有患要研究的疾病)。

6.由于知道所有病人都服用了可能有效的药物而引起的对终点评价的偏倚，例如，在读血压反应时倾向于正常血压，这样就减少了受试药与对照药之间的差异。

临床研究人员和申办者希望进行高质量的研究，临床试验管理规范指导原则(ICH E6)的发布将会有助于研究质量的

提高。不论怎样，必须明确，在要证明不同治疗的组间差异的试验中，很有必要采用良好的研究设计并尽量减少研究的误差，因为不完善的试验会增加不同治疗之间的差异存在时不能显示的可能性。。在安慰剂对照试验中，应当尽可能提高病人的依从性和病人人群对药物作用反应的可能性，以保证能将有效的治疗区别于安慰剂。然而，在许多临床研究中，虽然作了很多努力以确保研究的优异性和检测灵敏度，但临床研究仍常常不能可靠地区别有效药物和安慰剂。

相反，在试图说明两种治疗之间差异不超过特定范围的试验(非劣效性试验)中，对确保差异的检测方面就不会太关注，因为不显示差异大于界值才是研究所期望的结果。那些会缩小观察到的治疗组间差异的试验误差（如，依从性差、高安慰剂效应、某些伴随治疗、结果的错误分级），对于保存检测灵敏度来说特别要关注。。然而，当相信新药实际上优于对照药，就会有进行高质量试验的一个很强的促进因素，所以更可能排除非劣效界值。同样应当注意到，某些研究误差会加大变异，这可能会通过增宽可信区间而降低显示非劣效的可能性，结果导致不能排除治疗组与对照组之间的差异大于界值。因此，在这些研究中应采取一个有效的措施以减少变异，如可能由于测定技术差而引起的变异。

如所注意到的，确定一项非劣效性试验良好进行，不单要观注那些减低组间差异的因素，也要关注那些可能使试验

区别于确定非劣效界值基准的试验的因素。具体地说，应当确定是否存在入选人群、伴随治疗、治疗依从性、脱落的程度和原因的差异，影响到检测灵敏度。即便试验的设计和实施似乎与确定非劣效界值的基准试验十分类似，活性对照组结果却明显异常（如，抗生素试验的治愈率异常的高或低）则可提示存在重要的差异。

1.5.2 优效性临床试验的检测灵敏度

检测灵敏度的问题，尽管对非劣效性试验尤其的重要，事实上在任何不能检测出治疗组间区别的试验中都存在，包括安慰剂对照试验和剂量反应试验。如果一个治疗不能显示优于安慰剂，如，就意味着要么治疗无效，要么试验的设计和实施无法将有效治疗同安慰剂区分开。

在活性对照和安慰剂对照的临床试验中，一个有用的评估敏感度的方法是三臂试验，包括安慰剂和一个已知的活性对照治疗，一种有许多优点的试验设计。这样的试验既检测有效性大小（试验药物与安慰剂相比），还容许在检验的灵敏度已由活性对照与安慰剂的比较得到确立的情况下，进行研究药物与活性对照的比较（见 2.1.5.1.1）。

2.对照类型的详细考虑

2.1 安慰剂对照

2.1.1 描述(参阅 1.3.1)

在安慰剂对照研究中，基本上都是通过随机化的方法将

受试者分配到试验组或安慰剂组。安慰剂是一种"假"治疗，其物理特性例如颜色、重量、味道和气味都要尽可能地与试验药物相同，但不含有试验药物。某些试验可能研究一个以上剂量的试验药，或者同时包括一个活性对照药和一个安慰剂对照。在这些情况下，研究者使用一个以上的安慰剂（“双模似”）可能要比做到使全部药物看上去都相同更容易。安慰剂对照的应用促使并且也总是伴随着双盲试验。试验组与安慰剂组之间结果的差异用来评价在该试验条件下所研究药物的有效性。在本节描述中，提出了可被成功采用的多种设计：平行或交叉设计(参阅 ICH E9)、试验药物单一固定剂量或剂量递增、以及数个固定剂量。特别值得注意的几种设计将在后面介绍。需要指出的是，并非每一项含有安慰剂的研究都是安慰剂对照研究。例如，为了便于试验中盲法的实施，一项采用活性对照的研究可以分别为试验药和对照药各设置一种安慰剂(双模拟)，这仍然是一项活性对照研究，而不是安慰剂对照研究。安慰剂对照研究是比较安慰剂治疗与试验药治疗的研究。

应当注意并非所有的安慰剂都完全没有活性。例如，用于局部皮肤制剂研究所使用的一些辅料对照可能有一定的治疗效果，这不会对评价该制剂特定的有效性造成影响。当选择的辅料可能存在不良的作用时会导致特殊的问题。在这种情形下，可以使用“无治疗”组来检测试验制剂和辅料的

共同作用。

2.1.2 减少偏倚的能力

采用随机化和盲法的安慰剂对照试验，一般能最大限度地减少受试者和研究者的偏倚，但是这种试验会受到由于通过识别出一种药物的药理作用而破盲的影响(也许在交叉设计中关系较大)；在这种情况下，盲态下评价结果有利于减少偏倚。

2.1.3 伦理学问题

在尚无有效治疗的情况下检验一种新的治疗时，对于新的治疗与安慰剂对比的研究通常并无伦理方面的问题。但是，如果在计划进行的试验中，被研究的疾病已有有效的治疗时，采用安慰剂对照可能会出现伦理学、可接受性以及可行性的问题。当已知现有治疗可以防止受试人群发生如死亡或不可逆转的发病等严重损害时，一般不宜用安慰剂对照。但是偶尔也有例外，例如，标准治疗有严重的毒性以致许多病人会拒绝治疗。

在其他情况下，当没有严重损害时，即使可能会导致病人感到不适，但只要他们的参加是非强迫性的，而且他们对可能接受的治疗以及延迟治疗的后果完全知情，要求病人参加安慰剂对照试验可以认为是合乎伦理的。但是，这种试验即使符合伦理，也可能会出现重要的实际问题。例如，病人或医生可能不接受对疼痛或其他症状的延迟治疗，他们可能

不愿参加这样的研究。在有有效治疗的情况下，对新药进行安慰剂对照试验是否能被受试者和研究者所接受是一个由研究者、病人和机构审查委员会（IRB）/独立伦理委员会（IEC）共同决定的问题，而且，可接受性在不同的 ICH 区域会有所不同。可接受性可能取决于研究的特别设计以及所选择的患者人群，这将在下面讨论(参阅 2.1.5)。

在某些情况下，一项具体的安慰剂对照试验是否合乎伦理规范，可能取决于临床上要证明的内容和试验的特定环境。例如，在轻度原发性高血压且不伴有终末器官疾病的病人中，进行新降压药的短期安慰剂对照研究，一般可以考虑接受，如果是长期的或要有较严重的病人参加的研究，则很可能不会被接受。

必须注意，采用安慰剂对照或无治疗对照并不意味着病人完全不接受任何治疗。例如，在肿瘤研究中，如果没有被批准的治疗药物，在安慰剂/无处理组以及试验药物组的病人都会接受必要的姑息治疗，例如镇痛剂，和更好的支持护理。许多安慰剂对照试验被作为加载（add-on）试验进行，所有的病人接受一个标准的特定治疗或由研究医生/或机构选择的治疗（见 2.1.5.2.1）。

2.1.4 安慰剂对照研究的用途和在特定情况下推断的正确性

当使用安慰剂对照试验来显示治疗的有效性时，它尽可

能不受假设和外部(额外研究)信息的影响。在试验设计或实施方面出现的大多数问题增加了未能证明治疗差异(从而确定有效性)的可能性，因此，一项良好研究的试验质量很高。即使研究的主要目的是比较两种活性药物，或者是评价量效关系，加入安慰剂就提供了一种内部标准，加强了从其他比较中得出的推论。

安慰剂对照试验也能最大程度的区别由药物导致的不良反应和由潜在疾病或并发疾病造成的不良反应。应注意，在用于证明两个治疗间的相似性时，例如，通过显示在药物治疗的患者与安慰剂治疗的患者中某个特定的副作用事件的发生率相似，来证明该药物没有这个特定的副作用，安慰剂对照试验与任何等效或非劣效试验一样具有相同的检验灵敏度问题(参阅 1.5.1)；为了结果解释，研究人员必须知道如果研究药物引起不良事件，则该事件应被观察到。 通常，这样的研究应该包括确实引起有关不良事件的活性对照组，但在某些情况下，对于特定的研究设计，有可能通过记录“药物不良反应的历史敏感性”来得出研究对这种效应具有敏感性的结论。

2.1.5 设计方案的修改以及结合其他对照药解决伦理的、实践的或推断问题

通过对研究设计进行修改而仍保留其推断的优势，常常可以解决安慰剂对照试验中伦理或实践的局限。此外，通过

加入其他治疗组，例如多个剂量的研究药物或已知的活性对照药物，可以从安慰剂对照试验中获得更多的信息。

2.1.5.1 其他对照组

2.1.5.1.1 三臂研究；安慰剂和活性对照

正如第 1.5.1 节中提到的，包括活性对照以及安慰剂对照组的三臂试验，可以较容易地评定未能区分试验治疗与安慰剂的情况是意味着试验治疗的无效，还是仅仅是试验缺乏检验效能的结果。这类试验中安慰剂-标准药物的比较可以提供有关检测灵敏度的内部证据。如果认为重要的话，可以设定活性药物组的样本量大于安慰剂组，以改善活性药物与安慰剂比较的精确度。这一方法也会因为减少了随机分配到安慰剂组的机会，使试验更容易被病人和研究者接受。

2.1.5.1.2 其他剂量

除了安慰剂之外，几个固定剂量的试验药物组之间的随机化分配有助于评估量效关系，并特别有利于在比较试验中确保治疗间的公平比较。(参阅 ICH E4：支持药物注册的量效信息)。

2.1.5.1.3 析因/联合研究

析因设计可用于探索试验药单独效应以及与另一个药物几种剂量合用时的效应。一项这种类型的研究可以阐明各种组合治疗的效果。这类研究常用于对新的抗高血压药物进行评价，但也可用于同时应用一种以上治疗的其他场合，例

如，这类研究证明了阿司匹林与链激酶在预防心脏病发作后发生死亡的独立相加作用。

2.1.5.2 研究设计的其他改变

2.1.5.2.1 基于安慰剂对照的加载研究、替代研究

加载研究是对接受标准治疗的人进行的新药的安慰剂对照试验。当已知标准治疗可以降低死亡率或不可逆转的发病时以及当采用标准治疗作为活性对照的非劣效性试验无法进行或难以对结果进行解释(参阅 1.5)时，这类研究尤其重要。通常在抗癌药物、抗癫痫药物以及抗心力衰竭药物的研究中采用这类方法。这一设计仅在标准药不能完全治愈时(然而，几乎总是如此)适用，其优点是可以证明临床结果有所改善(强于“只是”非劣效)。当然，这一研究中确立的有效性仅仅是联合用药时产生的有效性，而且单独用药时的剂量可能与联合用药时的有效剂量有所不同。一般来说，这种方法仅在新药与标准药物的药理学机制不同时适用，当然也有例外，例如，由于可以延缓耐药的发生，药理作用相关药物的联合治疗用于艾滋病可以显示益处。

这一设计稍作修改后就成为替代研究，有时能获得单独用药的信息，特别适用于对慢性疾病治疗的研究；在这一设计中，将已接受有效剂量常规治疗的病人随机分组后分别加用新药或安慰剂，再通过逐渐减量撤除常规治疗，然后采用事先认定的有效标准观察试验药物和安慰剂维持受试者基

线状态的能力。这一方法已用于在类固醇依赖性病人中采用替代物减少类固醇使用的研究，这样可以避免在停药洗脱期间初始的类固醇撤除以及症状复发。这一方法也已经用于研究抗癫痫药物的单独治疗。

2.1.5.2.2 “及早脱离”；补救治疗

一项研究设计计划从无效治疗中及早脱离是有可能。及早脱离是指即时撤出那些临床状态恶化或改善未达到预定的水平(血压在预定时间内未得到控制，癫痫发作频率大于规定的水平，血压升高到一定的水平，心绞痛发生的频率超过预设的水平，肝炎病人的肝脏酶未能在预定的时间恢复正常)的受试者、发生了试验药物要预防的某一事件的受试者(不稳定心绞痛的首次复发，癫痫大发作，阵发性室上性心律不齐)、以及需要补救治疗的受试者。在这种情况下，研究终点变为需要改变治疗。必须详细设定这些终点是否出现的判断标准，而且测定时间的选择应当保证病人在疾病不能得到良好的控制时不至于没有有效药物进行治疗。这种试验设计的主要问题是它只能获得短期的有效性信息。而同样具有“及早脱离”特点的随机撤药试验(参阅 2.1.5.2.4)，却可以获得长期的有效性信息。

2.1.5.2.3 限制性安慰剂治疗时间

在长期使用安慰剂不能接受的情况下，在活性对照试验开始时，短时间加入安慰剂对照能够确定检测灵敏度（至少

对短期有效性而言）。随后，试验在没有安慰剂对照的情况下继续进行。

2.1.5.2.4 随机撤药

在随机撤药研究中，接受研究药物一段时间治疗的对象被随机分配到继续使用受试药物治疗或使用安慰剂(即停用有效药物)治疗。这种试验的受试者可以来自下列研究：有组织地开放单组研究、一组现有的临床队列（但通常都要经过一个方案规定的“洗脱”期以确定开始治疗时的基线）、对照试验中的活性对照组，或者活性对照试验中的一组或两组。继续治疗的试验组与随机接受安慰剂的试验组之间出现的任何差异都可以证明活性治疗的效应。随机化前对药物的观察期没有限制，从而当无法接受长期安慰剂治疗时，这种方法可以用来观察长期的持续有效性。停药后的观察期可以是固定的时间，或采用及早撤药，或到发生事件的时间(例如抑郁症复发)。当采用及早脱离设计时，应特别注意对患者的监测和对研究终点评价的程序，以确保迅速地发现指定治疗失败的病人。

随机撤药方法适合于以下几种情况。首先，它适用于研究复发性疾病发作的药物(例如抗抑郁药)，在这种情况下，撤药研究实际上是一个防止复发的研究。其次，它可用于能够抑制症状或体征(慢性疼痛、高血压、心绞痛)的，而难以进行长期安慰剂对照试验的药物；在这种情况下该研究可以

确立长期的有效性。再次，这一设计可以用于决定治疗的持续时间(例如，梗死形成后使用β阻断剂进行治疗)。

随机撤药设计的优点是：当其采用诸如症状重新出现等“及早脱离”为终点时，病人必须使用安慰剂的阶段比较短。

这一类型的设计可以解决有关剂量的问题。当所有病人接受了一个初始的固定剂量后，在“停药期”可以将他们随机地分配到几个不同的剂量组(也可以是安慰剂组)。当由于根据药效动力学资料，或由于母药或活性代谢产物的半衰期长而造成的活性药物大量蓄积，有理由相信初始剂量和维持剂量可能不同时，这一方法特别有用。值得注意的是，在初始的安慰剂对照剂量滴定研究后，随机撤药设计可以用于评价剂量-反应。剂量滴定研究对确定有效性是一种有效的设计，但不能提供良好的剂量-反应关系信息。在随机撤药期，将有反应者随机地分配到几个固定的剂量或安慰剂组，就可以精确地研究剂量-效应关系，且确保滴定设计的有效性。

在采用随机撤药设计时，充分认识停药现象的可能性是重要的，明智的方法是相对缓慢地逐渐减少用药。病人可能对某一种药物产生耐药而导致药物无效，但药物的撤除可能会引起疾病加剧，从而对药物的持续有效性作出错误的结论。由于随机撤药研究中的受试者大多为药物应答者且排除了对药物不耐受的人群，因此在这些研究中观察到的有效性可能会比在一般人群中观察到的有效性要大，认识到这一点

也是重要的。当研究中明确地仅包括对试验药物应答的受试者或仅包括完成了前一阶段研究的受试者 (这通常代表着一种良好的应答并且总是表明受试者对药物是耐受的) 时，就会产生这一现象。就旨在确定治疗应持续多长时间的研究来说，这样的纳入标准则提供了研究人群以及要做的比较。

2.1.5.2.5 其他设计考虑

在任何一种安慰剂对照研究中，不平衡的随机(例如受试药物与安慰剂的比例为 2: 1)能够提高研究药物的安全性数据，并能使研究对病人和/或研究者更具吸引力。

2.1.6 安慰剂对照试验的优点

2.1.6.1 能可靠地证明药物的有效性

就像其他优效性试验，安慰剂对照试验包含检测灵敏度的内部证据。当证明存在差异时，不参考外部发现，该差异也是可以解释的。

2.1.6.2 检测“绝对”有效性和安全性

安慰剂对照试验可以检测总的药物介导效果。相比之下，活性对照试验或剂量比较试验检测的是相对于其他治疗的效果。安慰剂对照试验还可以区分不良事件是由药物引起还是由于某些潜在的疾病或“背景噪音”所造成。对于一个三组试验(试验组、安慰剂组、活性对照组)，即便其主要目的是比较试验药物和活性对照药物，药物绝对效应大小的信息也是有用的。

2.1.6.3 高效率

安慰剂对照试验具有较高的效率，与其他任何平行对照研究相比，它需要较小的样本量就可以检测出治疗的有效性。

2.1.6.4 最小化受试者和研究者期望值的影响

采用盲法安慰剂对照可能减少因受试者或研究者的期望值而产生的结果改善量，因为他们都知道，有些受试者会接受非活性药物。这样可以增加研究检测药物真实有效性的能力。

2.1.7 安慰剂对照试验的缺点

2.1.7.1 伦理方面的问题(见 2.1.3 和 2.1.4)

对于某一特定人群来说，如果已知有某种有效的治疗方法能预防死亡或不可逆的发病，那么，从伦理上说，通常不能在安慰剂对照试验中对这一部分人群进行研究；但具体哪些疾病和人群适用于上述情况可能还存在争议。对伦理方面的关注也可以促使研究尽可能以病情较轻的患者作为受试者，或者尽管长期结局更为有意义，也会促使研究考察短期的终点。当安慰剂对照试验不合乎伦理，并且活性对照试验不可信时，新药的研究就有可能非常困难。例如，在急性心肌梗死的病人中进行一个溶栓剂的安慰剂对照试验，会被认为是不伦理的。然而，在这种情况下，基于历史数据建立一个有价值的非劣效界值也是困难的。因为急性再血管化过程

的出现可能会改变使用血栓溶解剂获益的大小。在 2.1.5 中介绍的设计可能适用于其中的某些情况。

2.1.7.2 病人和医师的顾虑

即使一般的共识认为停止或延迟治疗不会导致伤害，医师和(或)病人可能仍不愿意接受病人有被分配到安慰剂组的可能性。自认为疾病没有得到改善的受试者可能会退出试验，因为他们将缺乏有效性归因于接受了安慰剂治疗，这使研究的分析复杂化。但应注意到，有时因缺乏有效性而退出试验可以作为研究的终点。虽然，这样也能获得一些药物有效性方面的信息，但是，与受试者在临床状态下接受指定的治疗相比，这种信息的精确度要低一些。

2.1.7.3 通用性

有争议认为，任何对照试验，特别是安慰剂对照试验，代表了一种人为的环境，得出的结果不同于真实的“真实世界”有效性。如果由于伦理或实践的考虑，在安慰剂对照试验中的研究人群无代表性，研究结果的通用性就会有问题。例如，患有更严重疾病的病人可能因为试验方案，研究者或病人的选择等因素而被排除于安慰剂对照试验之外。在有些情况下，可能只有有限的病人或中心愿意参加研究。这些考虑是否在实践上(而不是在理论上)限制了研究结果的通用性还没有被确定。

2.1.7.4 无比较信息

缺乏活性对照组的安慰剂对照试验几乎不能提供有关有效性比较方面的有用信息，这些信息在许多情况下都是倍受关注且重要的。这种信息也不能可靠地从交叉研究的比较中获得，因为研究的条件可能是很不一致的。

2.2 无治疗平行对照(见 1.3.2)

随机的无治疗对照的一般特性及其优缺点均与安慰剂对照试验相似。而与安慰剂对照试验不同的是，它不可能是完全盲法，这就会影响试验的各个方面，包括受试者的保留，病人管理，以及观察的所有项目(见 1.2.2)。只有盲法不可行或不切实际时，这种设计在进行安慰剂对照试验的情况下是合适的。当使用这种设计时，一些关键性决策如受试者资格和终点的确定或者管理中的变更最好由对治疗分配处于盲态的观察者作出。与资料分析有关的决定，例如分析数据集中所包含的病人，也应该由不知道治疗分配的人作出(进一步讨论见 ICH E9)。

2.3 剂量-效应平行对照(见 1.3.3)

2.3.1 描述

剂量-效应研究是指受试者被随机分配到两个或多个剂量组中的一组，可以设或不设安慰剂对照组。剂量-效应研究可以确定剂量和有效性与不良反应之间的关系，和（或）证明药物的有效性。前者将在 E4 中予以讨论；证明药物有

效性是本指导原则的主题。药物有效性的证明可以基于配对比较研究中各剂量组之间或各剂量组与安慰剂之间的显著性差异，或者即使两组间无显著差异，但有效性有随剂量增加而显著增加的趋势。但在后一种情况下，可能需要进一步的研究来评价低剂量的有效性。正如 E9 中所述，应当预先指定用于分析主要有效性的特定方法。

治疗组间给药方案变化的试验，有同量效反应试验类似的问题。由于采用给药方案对照的试验来证明有效性很少见，因此以下的讨论集中在量效反应试验。

在剂量-效应研究中设立安慰剂组(零剂量组)有几个优点。首先，可以避免无法解释的研究，这类研究由于各剂量产生相似的效应而无法评价是否所有剂量同样有效或同样无效。其次，设立安慰剂组后可以估计总的药物介导效果，尽管当各剂量组都相对较小时，这种估计可能不太精确。其三，由于药物与安慰剂组间的差异通常大于各剂量之间的差异，使用安慰剂可以减小样本量。不同剂量组的样本量不需要相同，例如，较大的样本量可以得到比较精确的低剂量有效性的信息，也可以增加研究的把握度，证明预期最佳剂量的明确有效性。量效关系研究可以包括一个或多个剂量的活性对照组。随机撤药设计也可以将受试者分配到多个剂量水平。

2.3.2 减少偏倚的能力

如果剂量-效应研究采用盲法，则和其他随机盲法设计一样，能够减少受试者和研究者的偏倚。如果说药物的药理学作用可能打破某些病人或研究者的盲态，那么，剂量-效应研究可能比安慰剂对照试验更容易保持盲态。盲法治疗可能需要多重模拟或一些外观相似但不同剂量的药物制剂。

2.3.3 伦理学问题

剂量-效应研究的伦理问题或实践问题方面的考虑与影响安慰剂对照试验的问题相似。如果有一种治疗方法可以有效地防止死亡或防止不可逆性的发病，那么，从伦理上说，有意让患者随机分到有效性差的治疗并不比随机分到安慰剂组更易被接受。如果某种治疗措施是被用于不太严重的情况，或者药物的毒性相对于治疗获益相当大，那么，采用较低的、效果略差但毒性较低的剂量或安慰剂的剂量-效应研究对于患者和研究者来说可能是可接受的。

2.3.4 剂量-效应研究的用途和特定情况下推断的正确性

通常，当安慰剂对照研究适用并具有相似的可信度时，盲法的量效研究对于确定有效性和安全性是有用的(见2.1.4)。

2.3.5 修改设计和结合其他对照解决伦理的、实践的或推断问题

通常，为减轻诸如伦理的、实践的或推断方面的问题，

对安慰剂对照设计进行的修改也可以用于剂量-效应研究(见2.1.5)。

2.3.6 剂量-效应试验的优点

2.3.6.1 高效率

虽然，以完全有效的剂量组与安慰剂比较来证明药物的有效性是效率最高的，但这种设计可能产生难以接受的毒性，并且不能够获得任何量效关系的信息。如果量效关系是一致的，量效试验就能有效地证明药物有效性，并且也能获得量效关系的信息。如果药物的最佳有效剂量未知，那么更为谨慎的做法是选择一个剂量范围，而不是选择一个可能被证明是次优或有不可接受毒性的单一剂量。

2.3.6.2 可能的伦理方面优点

在有些情况下，特别是当药物的有效性和重要毒性与剂量相关时，剂量-效应研究可能代表一项在伦理上和实际上都可行的显示差异的试验 而此时甚至安慰剂对照试验也是不可行的，因为病人和研究者有理由接受以较小的有效性换取更大的安全性。

2.3.7 剂量-效应研究的缺点

需要认识到的一个潜在的问题是:（药物的剂量与有效性结果之间有显著的相关性），当各剂量组间两两比较无显著差异时，正向的剂量-效应趋势可以确定药物的有效性，但是可能也无法确定哪个剂量(除最大剂量之外)是真正有效

的。当然，单剂量研究也存在类似的问题，对于低于所研究剂量的其他剂量无法获得任何信息。

另一个必须意识到问题是，在剂量-效应研究中，各剂量组之间没有差异是常见的；如果没有安慰剂组为有效性提供清楚证明，这种研究就是没有任何结果的试验。

如果药物的治疗剂量范围完全未知，那么这种设计可能是无效的，因为许多病人可有会使用低于或超出治疗剂量范围的药物剂量。

在证明药物有效性方面，剂量-效应研究设计可能比安慰剂对照增量设计效率低；但多数情况下，它们能获得更好的量效信息。（见 E4）

2.4 活性药物对照(见 1.3.4)

2.4.1 描述

活性对照(阳性对照)试验是将研究药物与已知的活性药物进行比较。这类试验通常是随机和双盲的。试验设计最关键的问题是，该试验是用于证明两种治疗之间的差异，还是证明非劣效或等效。申办者如果想利用一个相对于标准治疗药物的非劣效性试验来说明试验药物的有效性，就应如 1.5 中所讨论的那样，解决试验检测灵敏度问题。在药物非劣效性或等效性试验中，活性对照治疗在所使用的剂量和研究条件下应该有确定的有效性(见 ICH E9 部分；临床试验中的统计原则)。通常，这就意味着该活性对照药在申请研究的地

区已被接受，且剂量和适应证与受试研究中的相同。另一方面，即使活性对照药物的剂量太低或者活性对照药的益处不明确(但不能有害)，获得有利于试验药物结果的优效性研究比较容易解读为有效性的证据。然而，试验中试验药物所表现的相对于对照药物的优效性，只有当活性对照药物在最佳的给药剂量和给药方案下、用于合适的病人时才能判断为真正的优效(见 1.4.3)。因为不能确保研究的检测灵敏度，如果试验药物未能显示优效，那么缺少恰当的对照治疗还会让该研究作为一个非劣效研究不可用。

2.4.2 减少偏倚的能力

随机、盲法的活性对照试验通常可将受试者和研究者方面的偏倚减至最小，但仍有必要提醒一点。在非劣效性试验中，研究者和受试者都清楚所有受试者会接受活性药物，但不知道具体用哪个药。这可能会导致对部分对象的评价产生偏倚的解释，也就是倾向于将效果处于临界状态的对象归入有效一类，如在抗抑郁药的研究中，这种偏倚可能减少治疗间的差异，并增加非劣效结果不能作为有效性证据的可能性。

2.4.3 伦理学问题

因为活性对照试验中所有对象都接受活性药物治疗，所以一般认为这类试验所面临的伦理学问题及实践问题要比安慰剂对照试验少。但应该意识到，其中接受新治疗的受试

者，并不是接受标准治疗(就好像安慰剂组一样)，也有可能接受的是一个无效的或有害的药物。如果已知活性对照治疗能改善存活或降低不可逆性损害发生率，在此情况下安慰剂或无治疗对照将会是不可接受的，这一点非常重要。因此，试验治疗应有非常可靠的理论基础。如果没有充足的理由来预期新药至少与标准对照药物一样好，在条件允许时，采用加载研究可能更合适(见 2.1.5.2.1)。

2.4.4 活性对照试验的应用及特定情况下推断的正确性

当一个新治疗显示其优于活性对照，且假设活性对照确实无害时，该研究就可以像其他优效性试验一样，推断出新治疗的有效性。当活性对照试验是用于说明试验药物的非劣效性时，那么就应该特别考虑在 1.5 讨论过的检测灵敏度。如果已经确立了检测灵敏度，这种活性对照试验还可用于评价相对有效性。

2.4.5 修改设计和结合其他对照解决伦理的、实践的问题或推断问题

正如前面所讨论的(见 2.1.5)，活性对照试验可以包括一个安慰剂组，多个试验药物剂量组，和/或其他活性对照药物的剂量组。在止痛药试验中，试验和对照药物都同时包括数个剂量组的比较性剂量-效应研究是非常典型的。在活性对照药物研究中，剂量可以是固定的，也可以呈滴定形式，既可以使用交叉设计，也可以使用平行设计。非劣效性试验的

检测灵敏度有时可以从试验末的随机安慰剂对照撤药相(见 2.1.5.2.4)或是试验初短期与安慰剂的对比(见 2.1.5.2.3)获得支持。在所选择的人群中(对其他治疗或活性对照无响应者)的活性对照优效性研究是非常有用的，通常也易于解释，但其结果可能不通用。

2.4.6 活性对照试验的优点

2.4.6.1 伦理和实践的优点

活性对照的试验设计，无论是旨在表明非劣效性、等效性或优效性，都可减少由于未使用已证明有重要健康获益的药物而引起的伦理方面的顾虑。也可以打消病人和内科医生因未能使用已证明有效的治疗方法而担忧。活性对照试验有助于受试者的招募和 IRB/IEC 对试验的批准，而且有可能进行较大样本的研究；并能减少由于缺乏有效性而退出试验的受试者数量。

2.4.6.2 信息含量

当活性对照研究显示出试验治疗优效于活性治疗时，更易于解读为有效性的证据。所需要的较大样本量有时在活性对照试验中更易达到和被接受，而且能提供更多的安全性信息。如果设计恰当，活性对照研究也能提供关于相对有效性的信息。

2.4.7 活性对照试验的缺点

2.4.7.1 信息含量

1.5 中讨论了非劣效性或等效性试验中检测灵敏度的问题以及试验对得出有效性结论的能力。即使当检测灵敏度是可靠的、研究也适于检测有效性时，活性对照研究仍不能直接评价绝对作用的大小，难以定量描述安全性结果。

2.4.7.2 样本量大

在非劣效性试验中，为了合理的确保界值不高于活性对照的预计最小作用，通常较为保守地选择拒绝非劣效界值。其次，由于通常有意排除多于对照药物作用的合理部分的丢失 (见 1.5.1)，经常使用比较小的非劣效界值。由于界值选择的保守，样本量可能非常大。在活性对照优效性试验中，两种药物之间的预期差异总是比药物与安慰剂间的预期差异小，这也导致了大样本量。

2.5 外部对照(包括历史对照,见 1.3.5 部分)

2.5.1 描述

外部对照试验是指对照组的病人并非属于受试组所在的同一随机试验，即不存在平行随机对照组。因此，对照组与接受治疗者并不完全来自同一人群。通常，对照组是先前(历史对照)所观察的且有完善记录的患者群，但可以是在另一机构同期观察的一组人群，或是同一机构研究之外的人群。外部对照研究可以是优效性研究（例如对于非治疗组）

或非劣效性研究。有时来源于一个较大规模研究的某些病人也可根据其与治疗组相似的某些特征被选择作为对照组；甚至可能为接受治疗的病人匹配特别的对照。

在所谓“基线对照研究”中，病人随着时间的状况变化与基线情况比较。虽然这些研究有时会被认为是使用“病人自身作对照”，但实际上不存在内部对照。更确切的说，是自基线的变化与无试验药物治疗条件下可能状况的估计之间进行对比。基线对照研究和那些使用更复杂的开-关-开(药物-安慰剂-药物)设计的研究，都属于这一类型，但都没有包括平行随机对照组。值得注意的是，在这些试验中，总是将观察到的相对于基线的变化或研究周期间的变化与所估计的无治疗干预时可能发生的状况相比较(至少是隐式的)。通常是根据“总体认识”进行这种估计，且不参照特定的对照人群。尽管对于某些情况，例如当治疗作用很显著、接受治疗后立即发生、且不可能是自发作用时(例如：全身麻醉、心脏复率、可测量的肿瘤缩小)，这是十分合理的；但在多数情况下并不是很明显，这就需要寻找一个特定的历史经验。此类试验的设计者或分析者需要意识到这种研究的局限性，并且准备为它的使用辩护。

2.5.2 减少偏倚的能力

不能控制偏倚是外部对照试验的主要和公认的缺陷，这一缺陷足以使该研究设计在许多场合不适用。要确立治疗组

与对照组之间的可比性从而充分发挥对照组的主要作用(见1.2)总是比较困难，在许多情况下是不可能的。组间在很多方面可能不同，除了所采用的研究治疗不同可影响结局外，还包括人口学特征、诊断标准、疾病分期或严重程度、伴随治疗以及观测条件(如，评价结果的方法、研究者的预期)。这些不同可能包括重要的、但不能识别的预测因素，这些因素尚未被检测。当使用外部对照时，不能用盲法和随机化来减少偏倚。已经证明无治疗的历史对照组的结果往往比随机化研究中显然相似的对照组的结果更差，可能反映了一种选择偏倚。在随机化研究中，对照组必须符合某些标准才能进入研究，这些标准比外部对照组的入选标准更为严格，而且能辨别比外部对照组患更轻度疾病的人群。确认外部对照组通常是回顾性的，这会导致选择上的潜在偏倚。无法控制偏倚的一个后果是，外部对照试验结果的说服力有赖于获得比平行对照试验所需的更为极端的统计学显著性水平和治疗组间更大的估计差异。。

无法控制偏倚使外部对照设计限于在治疗作用十分显著以及疾病的一般过程是可以准确预测的条件下使用。另外，外部对照的使用应被限于终点是客观的，并且基线和治疗变量对终点的影响可准确描述的试验。

值得注意的是，缺少随机化和盲法以及试验组和对照组之间缺乏可比性保障而导致的问题，可能使该设计本身存在

实质性偏倚并且不能定量。然而，一些设计和实施外部对照试验的具体方法也可能使试验更具说服力并减少偏倚。应当选择有详细信息的对照组，必要时包括各病例的人口统计学数据、基线状况、伴随治疗以及研究进程。对照病例应尽可能地与研究中预期接受试验药物治疗的人群相似，并且应在除研究治疗外相同背景下以相同的方式接受治疗。研究观查在时间选择和方法学上应与在对照病例中使用的相似。为了减少选择偏倚，应在比较分析前选择对照组；这并非总是可行的，因为对照组的结果可能已经发表。任何选择标准的匹配或者由于人群差异所做的调整都应在选择对照和进行研究之前说明。当不存在一个明显的“最适宜”外部对照时，建议采用多个外部对照，只要在分析计划中适当地说明在作出推断时如何使用每一个对照(例如：研究组必须充分优于作用最强的对照组才能作出有效结论)。在某些情况下，让一组独立的评价人员根据通用标准以盲态的方式重新评价对照组和试验组中的终点是十分有用的。

2.5.3 伦理学问题

当一种药物预期用于治疗尚无令人满意的治疗方法的严重疾病时，尤其是根据理论基础、动物数据或早期人体用药经验，该新药被认为非常有前途时，不愿意设立不接受新治疗的平行对照组进行对比研究是可以理解的。同时，实施没有实际机会可靠地展示治疗有效性的研究是不负责任的

或不符合伦理的。应该预料到许多有前途的治疗其有效性没有预期的那么好，或者在对照试验中显示无效，在这些情况下研究者可能会面临极难以作出判断的局面。在一些特殊情况下，启动外部对照试验可能很诱人，希望获得令人信服的显著作用，如果未能实现，立即转向随机化试验。

相应的，通常也更为可取，，在处理没有满意治疗方法的严重疾病，但疾病的进程也不能很好地预测时，即使是最早期的研究也应当是随机的。当研究实施先于“治疗是有效的”这一印象时，这通常是可行的。研究应在独立的数据监察委员会的监控下进行，这样可尽早地检测到治疗的显著利益。平行对照试验能快速检出极端效应，另外，可以检出外部对照试验不能准确反应的较轻微、但仍有价值的效应。

2.5.4 外部对照试验的用途以及在特定情况下推断的正确性

只有试验进行前就非常确信试验治疗优于所有现有的治疗，以至于其他设计不可接受，而且被治疗的疾病或状态的进程有良好的记录且具有高度的可预测性，通常才会考虑外部对照试验。。即使在这些情况下，也很有可能使用替代的、随机、平行对照设计(见 2.1.5)。

如果研究终点是客观的，治疗组结果与外部对照结果有显著差异，而且治疗-对照之间的比较有高水平的统计显著性，能很好地描述影响疾病结果的协变量的特征，对照组和研究组在所有已知相关的基线水平、治疗(研究药物除外)和

观测变量都十分相似，这时外部对照设计是极可能有说服力的。然而，即使在这种条件下，也有外部对照试验导致错误结论的实例记录。

当考虑使用外部对照试验时，适当注意设计和操作有助于减少偏倚（见 2.5.2）。

2.5.5 修改设计和结合其他对照解决伦理的、实践的问题或推断问题

如上文所述(见 2.1.5.2.4)，外部对照设计可以通过使用随机安慰剂对照的撤药相（通常与及早脱离规则相关），使其结合随机化和盲法的一些元素。因此，早期治疗的结果(识别对治疗有反应的受试者并继续接受治疗)可通过一个严格的基本无假设、无偏倚的研究来证实。

2.5.6 外部对照试验的优点

外部对照试验的主要优点是所有病人都能接受有前途药物的治疗，这样可使该研究对病人和医生有更强的吸引力。

因为所有病人都暴露于试验药物，试验设计具有潜在的高效性，这对一些罕见疾病尤其重要。但是，尽管在外部对照试验中采用单治疗组，但对外部对照组结果的估计通常会保守，可能会导致需要比安慰剂对照研究大的样本量。由于治疗组和对照组间的差异可能确定，也可能无法测定，需要特别小心（如，采用一个更加严格的显著性区间），结果通

常有利于治疗组。

2.5.7 外部对照试验的缺点

外部对照试验不能采用盲法设计，因此存在由病人、观测人员和分析人员引起的偏倚，这是主要缺点。这些问题可能在某种程度上能够缓和，但即使是 2.5.2 所推荐的方法也不能彻底解决这些问题，因为治疗分配不是随机化的，开始治疗时对照组与治疗组的可比性以及试验期间受试者的可比性得不到保证或者不能很好地进行评估；外部对照试验有过高估计治疗有效性的趋势，这一点已经得到很好的证明。需要认识到这种研究统计显著性的检测可信度不如随机试验高。

3.选择平行对照组

表 1 描述了各种对照类型的用途，图 1 提供了在不同类型对照组中进行选择的决定程序。虽然图表旨在说明如何选择用于有效性分析的对照，但一些设计也允许试验药物和对照药物的比较。特定区域内疗法可及性和医疗实践情况会影响对照的选择。

特定条件和特定目的下主要对照类型(安慰剂、活性对照和剂量-效应对照)的可能用途见表 1；该表应在有特定背景细节描述的情况下使用，这样才能起到作用。在所有情况，应该假设研究是经过合理设计的。外部对照是特殊情况因此没有列入表中。

通常，能够表明优于平行对照治疗是最有说服力的有效性证据。如果由于伦理学或实践原因优效性试验不可行或不适宜，而且能够规律地观察到活性对照确定的治疗效果(例如在大多数条件下抗生素的作用)，则可以使用非劣效性或等效性试验，而且是有说服力的。

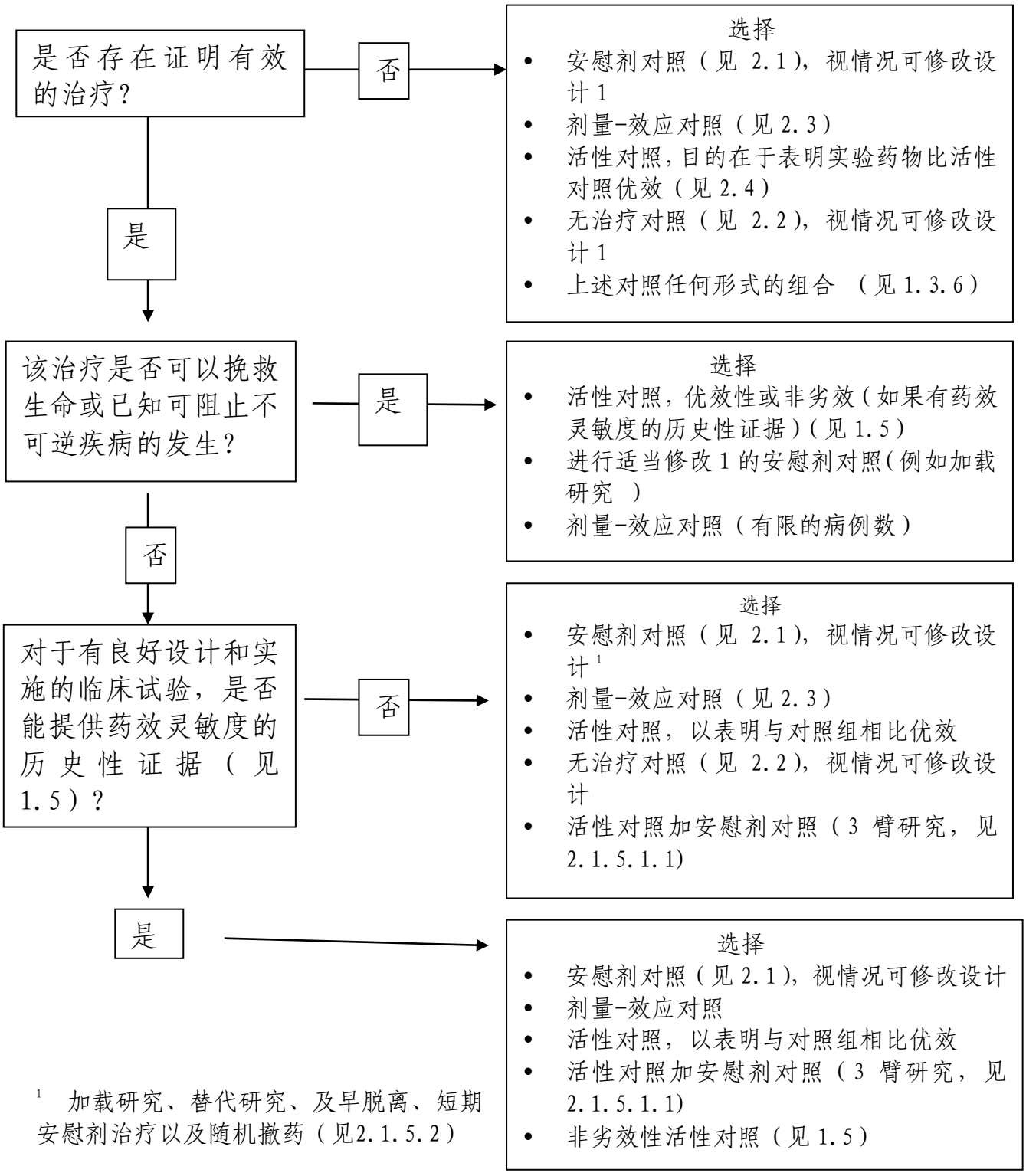
表 1　不同情况下各种平行对照类型的用途

	对　照　类　型							
试验目的	安慰剂	活性对照非劣效性	活性对照优效性	剂量效应(D/R)	安慰剂+活性对照	安慰剂+D/R	活性对照+D/R	安慰剂+活性对照+D/R
衡量“绝对”作用大小	Y	N	N	N	Y	Y	N	Y
观察是否有作用	Y	P	Y	Y	Y	Y	Y	Y
观察剂量-效应关系	N	N	N	Y	N	Y	Y	Y
治疗间的比较	N	P	Y	N	Y	N	P	Y

Y=是，N=否，P=可能，根据是否有药效灵敏度的历史性证据

图 1　为描述有效性选择平行对照

该图表明选择对照组的基本逻辑；可以根据特定区域中可获得的药物或医疗实践作出决定。



[1] 加载研究、替代研究、及早脱离、短期安慰剂治疗以及随机撤药（见2.1.5.2）