

药物临床试验适应性设计指导原则

（征求意见稿）

国家药品监督管理局药品审评中心

2020 年 5 月

目 录

一、 引言.....	1
二、 适应性设计中需要考虑的因素.....	2
(一) 合理性.....	3
(二) 完整性.....	4
(三) 可行性.....	5
三、 常用的适应性设计.....	5
(一) 成组序贯设计.....	6
(二) 样本量重新估计.....	7
(三) 两阶段无缝适应性设计.....	8
(四) 适应性富集设计.....	10
(五) 适应性主方案试验设计.....	12
(六) 多重适应性设计.....	13
四、 适应性设计的其他考虑.....	14
(一) 贝叶斯方法的应用.....	14
(二) 模拟方法的应用.....	16
五、 监管的考虑.....	17
(一) 与监管部门的沟通.....	17
(二) 对申报资料的要求.....	18
(三) 其他的考虑.....	18
参考文献.....	20
附录 1、词汇表.....	22
附录 2、适应性设计案例.....	25

药物临床试验适应性设计指导原则

一、引言

本指导原则侧重于讨论临床试验中采用适应性设计的重要概念和原则，主要从监管角度来讨论一些常用的适应性设计，提出建议供申办方在计划使用适应性设计时予以考虑，同时也提出了一些具体要求，目的是帮助申办方评估是否采用以及如何进行适应性设计，以支持新药临床试验申请和新药上市申请。对于申办方提交的适应性设计方案，监管部门会依据提交的材料及临床试验中所涉及的各种因素，在综合评估后给出具体的咨询意见。

适应性设计可能会涉及多种统计方法，因此，在具体设计适应性临床试验方案时，申办方应同时参考其它相关的 ICH 指导原则和国内指导原则。

临床试验最为常用的设计是固定样本量的设计，也常被称为传统设计，被广泛地用于确证性试验，它是指对试验的数据仅在试验终止时做一次有效性分析并在试验进行中没有对试验进行任何的分析和修改。很多确证性试验是基于前期有限的样本量而设计，有可能存在较大的偏差。如何根据试验期间累积的数据对试验做出相应的修改从而修正设计的偏差就成为试验设计时需要考虑的一个重要问题。适应性设计是对固定样本量的设计进行某些调整的设计，本指导原则中适应性设计被定义为：按照预先设定的计划，在期中分析时根据试验期间累积的数据对试验做出相应修改的临床试验设计，这种修改又称为适应性修改。适应性修改计划必须在临床试验开始前的试验方案和统计分析计划中预先设定。

成组序贯设计是最早应用于临床试验的适应性设计，其后，适应性设计较广泛地用于样本量的重新估计。在 2007 年欧盟 EMA 和 2010 年美国 FDA

发布适应性设计监管指南后，适应性设计的研究逐步推广和发展到了多种类型的试验设计，例如从多个剂量中筛选出具有潜在获益剂量的设计，到目前的多个目标人群、多个假设、多个终点和多重检验等更为复杂的设计。随着更多新的理论方法不断地推出以及应用经验不断地增多，越来越多的适应性设计在临床试验中得以应用，几乎涵盖了药物研发的各个阶段和所有的疾病领域。

许多临床试验的失败源于前期信息有限导致的设计偏差，而适应性设计因为在试验期间可以根据试验期间累积的数据对方案进行修改，以修正初始设计的偏差，从而显著增加了试验的成功率。适应性设计还可以提高试验的效率，如减少试验所需要的样本量、缩短不同研究阶段之间的时间间隔、选择更合适的终点、选择更合适的目标人群、利用相同数量的受试者获得更多的数据等。此外，复杂的适应性设计还可以在一个试验中同时处理多个试验目标、多种试验药物和多种疾病。

尽管适应性设计具有上述诸多优势，但由于其复杂性，同时也给试验设计、数据分析和结果解释，特别是试验的实施，带来了诸多挑战和问题，例如，怎样控制整体 I 类错误率、怎样选择分析模型和怎样估计疗效，以及因试验实施困难可能导致的操作偏倚等，这些都限制了适应性设计的应用。

本指导原则主要适用于化学药、生物制品和中药的确证性临床试验，对于探索性研究也具有参考意义。

二、适应性设计中需要考虑的因素

在决定是否采用适应性设计之前，应全面深入地权衡适应性设计和传统设计之间的优劣，尤其是适应性设计在设计、实施和统计分析方面的复杂性，以及由此而带来的在试验实施中可能会引入的、不可避免的操作偏倚以及其

他各种挑战。是否采用适应性设计需综合考虑诸多因素，特别是合理性（validity）、完整性（integrity）和可行性（feasibility）。

（一）合理性

适应性设计的合理性是指所采用的统计分析方法不会导致疗效的估计产生偏倚。试验的合理性是关于试验结果的可信度、可解释性和说服力的综合描述。保持试验的合理性意味着应该有正确的统计推断方法，例如怎样计算调整后的 p 值、怎样估计效应量及置信区间，以及怎样衡量不同阶段治疗效果的一致性。

由于本指导原则的关注重点是能够支持注册的确证性试验，因此要求试验的整体 I 类错误率严格控制在双侧 0.05（或单侧 0.025）的水平。判断适应性设计是否合理，最重要的标准是所使用的统计方法能否控制整体 I 类错误率。对有些适应性设计来说，如采用双侧检验，由于不同阶段的 p 值不能反映组间比较的方向，有可能使得最终的整体 p 值难以解释。为避免这种情况，可在试验方案中选择单侧检验。但对另一些适应性设计，例如不对称的双边假设，双侧检验会是更合适的选择。除了某些特殊情形外，适应性修改都需调整试验的 I 类错误率。

适应性设计可能同时涉及多个目标人群、多个假设、多个终点或多重检验，故对统计分析方法的合理性有着很高的要求。如果对适应性修改没有相应的合理有效的统计方法，则不宜采用该设计。此外，由于适应性设计的复杂性，在某些情况下没有适用的统计推断的理论公式，统计方法的合理性在一定程度上需要基于模拟方法进行验证，这有可能又增加了额外的不确定性。

适应性设计的分析需要结合来自多个阶段的数据，各个阶段疗效估计的不一致性不仅会使基于合并数据的统计推断变得困难，还会导致试验结果难

以解释。另外许多适应性修改是以达到统计学阳性结果为目的，如果最终统计检验结果虽为阳性但临床获益太小，也不足以支持所验证的药物疗效。

（二）完整性

适应性设计的完整性是指对试验操作所引入偏倚的良好控制。保持试验的完整性意味着按照预先设定的方案进行修改和保持期中分析结果的盲态，以求最大限度地减少操作偏倚。

避免引入操作偏倚是所有临床试验的最基本要求。适应性设计由于涉及临床试验许多方面的修改，有可能影响后续试验的执行，对保持试验的完整性增加了额外的难度。因此，适应性设计的所有期中分析都应该由独立于申办方的适应性设计委员会或第三方专家及其独立统计支持团队完成，并保证期中分析的结果不为申办方、研究者和受试者所知悉，以免影响后续试验的执行和引入操作偏倚。在大多数情形下，如果适应性设计不是特别复杂，可由独立数据监查委员会同时负责适应性调整的操作和执行；如果适应性设计较为复杂，申办方可以考虑选择具备适应性修改专业知识的专家组成独立数据监查委员会；如果适应性设计极为复杂，而数据监查委员又无足够的相关经验，则有必要设置一个独立的适应性设计委员会。根据复杂的程度，适应性设计委员会中可以包含多名统计师。因适应性修改涉及多个环节，设立一个有效的防火墙以防止期中分析结果外泄而造成可能的操作偏倚是执行中最为重要的任务。为此，适应性设计的方案应包含一个完善的操作流程，特别是关于如何设置相关信息的访问权限。同时，为避免不可控制因素对试验结果的影响，还要考虑怎样避免根据试验所做的修改而被间接地推出期中分析的结果。应该注意的是，适应性设计委员会在所有提供给申办方的任何修改的建议中，均不应涉及具体的期中分析结果。申办方也要准备好所有需要的标准操作流程，并将所有涉及适应性修改的相关程序融入其中，同时记录好

94 所有实际操作的过程以备在申报新药时审查。以上这些因素都应该在试验的
95 设计阶段仔细考虑，并特别需要在试验进行中严格地执行，以免影响试验的
96 完整性和试验结论的可靠性。

97 (三) 可行性

98 适应性设计的可行性是指试验的适应性修改能否在实际中实施。由于适
99 应性设计比传统设计更为复杂，并且实施和分析更加困难，在规划适应性设
100 计之前，需要考虑以下因素：适应性调整策略应该能够保障试验的合理性和
101 完整性；相对于试验周期，应该有充裕的时间根据试验累积数据的分析结果
102 进行适应性修改和后续试验的执行；期中数据收集和数据清理应该可以快速
103 完成，以便按预定计划完成期中分析而无需中途暂停招募受试者；应该能够
104 快速修改随机化程序/药物供应系统；应该具备足够的药物供应管理的能力以
105 及能够负担增加的药物供应；应该提前准备好适应性设计的数据采集系统；
106 应该保证与各相关方的沟通顺畅有效；应该能够配备专业软件来完成复杂设
107 计和相关分析的计算等，以满足试验运行中适应性修改和执行的需求。同时，
108 在试验设计阶段，申办方可以与研究者沟通，根据临床考虑提出一个试验的
109 目标列表，评估所考虑的适应性设计在实际中能够顺利进行的可行性。如果
110 相关适应性修改难以实施，则应该考虑其它设计。

111 综上所述，若计划采用适应性设计，需要仔细地评估其是否确有优势。
112 若无法决策，可以采用模拟方法与传统设计比较，以评估适应性设计的效率，
113 并选出较优的设计。如果评估后适应性设计没有体现出太多的优势，建议谨
114 慎考虑适应性设计。

115 三、常用的适应性设计

116 适应性设计是按照预先计划，根据累积的临床试验数据的分析结果，在
117 保证试验的合理性和完整性的前提下，对临床试验方案进行修改。一方面，

适应性修改是“按预先设定的计划”进行的，而不是临时提出的修改方案；另一方面，适应性修改是一个自我学习的过程，即通过对累积数据的不断学习，相应地修改试验方案，以适应不断变化的研究环境。因此，适应性设计旨在更好地改进进行中的临床试验，而不是因设计本身缺陷而有极大可能导致临床试验失败所做的临时补救。

适应性设计在应用中涉及范围很广，限于篇幅，本指导原则仅讨论几种常用的适应性设计，包括成组序贯设计、样本量重估、两阶段无缝适应性设计、适应性富集设计、适应性主方案试验设计、多重适应性设计。这些设计的原理和方法也适用于大多数其他适应性设计。此外，本章节中所讨论的方法将会通过几个假想的案例予以说明（见附录 2）。

（一）成组序贯设计

成组序贯设计是指方案中预先计划在试验过程中进行一次或多次期中分析，依据每一次期中分析的结果做出后续试验的决策，决策通常有四种可能：①依据优效性终止试验；②依据无效性终止试验；③依据安全性终止试验；④继续试验。期中分析的时间可以基于日历时间，也可基于累积数据的占比，如受试者入组比例或发生目标事件数的比例。如果期中分析至少有一次疗效分析，且均有以无效性或优效性提前终止试验的可能，则应调整每次分析的 I 类错误率，并将整体 I 类错误率控制在双侧 0.05（或单侧 0.025）水平。调整 I 类错误率的常用方法包括 Pocock 方法、O'Brien & Fleming 方法和 Lan & DeMets 方法。由于期中分析仅使用了部分数据，结果仍有较大的不确定性，评估有效性的方法应较为保守以增加结论的可靠性。无效性终止试验的设计分为绑定边界和非绑定边界。绑定的无效性边界会降低拒绝无效假设的概率，因此可以在控制整体 I 类错误的前提下，适当放宽优效性边界，提高得到阳性结果的概率，但在期中分析结果一旦跨越无效性边界时必须终止试验。非

绑定边界即使在试验结果跨越该边界时，独立数据监查委员会基于全面考虑后仍然可以建议试验继续进行。

选择期中分析的时间点也要仔细考虑。如果成组序贯调整计划中存在以优效性提前终止试验的可能，时间点的选择应该考虑期中的数据是否充分以便能够提供可靠的疗效估计和安全性评价的结果，也包括重要的次要终点以及一些重要的亚组结果的估计。若期中分析是要验证药物的安全性和无效性，时间点则应该侧重于如何最大程度地保护受试者。

（二）样本量重新估计

样本量重新估计是依据预先设定的期中分析计划，利用累积的试验数据重新计算样本量，以保证最终的统计检验能达到预先设定的目标或修改后的目标，并同时能够控制整体 I 类错误率。

初始样本量的估计通常取决于效应量、主要终点的变异度、试验随访时间、受试者脱落率等诸多因素，而这些常常基于以往的研究数据。多数情况下，试验设计阶段样本量的估计所需要的参数信息往往不够充分，可能会导致错估样本量。适应性设计中的样本量重新估计为此类问题提供了有效的解决方案。

样本量重新估计的方法可以分为盲态方法和非盲态方法。

盲态方法，也称为非比较分析方法（**non-comparative analysis**），是指期中分析时不使用实际试验分组的信息，或者虽然使用了实际试验分组的信息，但未做任何涉及组间比较的分析，如在期中分析时对两个治疗组的数据合并后做的汇总分析。

盲态方法的样本量重新估计是指根据累积的数据，计算样本量的重要参数（如合并方差或标准差），然后对样本量进行重新估计，因期中分析时不涉及组间的疗效比较，故一般不需要调整 I 类错误率。该方法比较容易实施，

166 一般不会引入操作偏倚，而且相关的统计方法也较为完善，只需要在试验设
167 计的阶段预先做好规划。

168 非盲态方法，也称比较分析方法（comparative analysis），是指期中分析
169 时使用了试验分组信息（包括各组的真实名称或可区分的分组代码）的分析，
170 分析内容涉及组间的比较。

171 非盲态方法的样本量重新估计是指根据累积数据以及分组信息，计算样
172 本量的重要参数（如每组的效应量），然后对样本量进行重新估计，因期中
173 分析涉及组间的疗效比较，通常需要对 I 类错误率进行相应调整。

174 非盲态分析的样本量重新估计需要预先在研究方案中阐明，包括何时进
175 行重新估计、决策时使用什么标准、重新估计时使用什么方法、如何调整检
176 验水准 α 以便控制整体 I 类错误率、由谁执行非盲态分析，以及最后由谁执行
177 整个操作过程。应该特别注意，一个试验中一般建议只做一次样本量重新估
178 计。当重新估计的样本量少于初次设计的样本量时，除非有非常特别的理由，
179 通常不接受样本量减少的调整。

180 适应性设计中是否采用非盲态样本量重新估计需要考虑多种因素。例如，
181 若有比较可靠的前期数据，非盲态下样本量重新估计是否必要？采用非盲态
182 下样本量重估所付的代价（如检验水准调整）与初始设计时略微放大样本量
183 相比，哪种策略更为有利？期中分析能否很快完成，是否可能因为入组较快
184 完成而导致没有充足时间用来调整试验？在什么时间点进行期中分析？因此，
185 设计应根据试验本身的特点，仔细考虑各种因素，然后选择出最合适的方法。

186 样本量重新估计的方法有很多文献可供参考，可根据具体试验选择一种
187 合适的方法。

188 （三）两阶段无缝适应性设计

两阶段无缝适应性设计，是指将一个试验分为两个阶段，在第 1 阶段结束时进行期中分析，依据预先设定的判断标准，对第 2 阶段的试验进行适应性修改。无缝设计通常分为操作无缝设计和推断无缝设计。操作无缝设计可将第一阶段试验受试者排除在主要分析之外，不需要对 I 类错误进行调整。推断无缝设计在主要分析中包含第一阶段试验受试者，并根据自适应的性质和假设检验策略作出相应的调整。

根据试验目标和试验终点，两阶段设计一般可分为四种类型，即同目标/同终点、同目标/不同终点、不同目标/同终点和不同目标/不同终点。任何一个两阶段无缝适应性设计都可以被适当地归类，可根据具体试验选择合适的方法。

如果试验在两个阶段的试验组数也相同，成组序贯设计可视为同目标/同终点试验中的一个特例。两阶段无缝试验中，常见的是 I/II 期无缝适应性设计和 II/III 期无缝适应性设计。前者多用于探索性试验，在第 1 阶段可能会探索一个生物标记，并在第 2 阶段探索其早期有效性；后者常用于确证性试验，在第 1 阶段可做剂量选择，并在第 2 阶段验证其有效性。

独立的 II 期试验通常包括多个试验组，如相同药物的多个剂量水平，目的是选出合适的剂量并决定是否进行 III 期试验。III 期试验是一个独立于 II 期的试验，其最终分析并不包含 II 期试验的数据，这种设计没有充分地利用 II 期试验的数据。推断无缝适应性设计的最终分析则包含了试验的两个阶段入组的所有受试者的数据。这种设计具有很多优点，例如可以缩短通常由 II 期试验结束时到 III 期试验开始时的时间间隔、减少试验的总样本量、缩短试验的时长、减少试验的费用、增加最终分析的样本量等。同时，因第 1 阶段入组的受试者有更长的随访时间，或能更早地观察到药物的长期安全性。

采用无缝 II/III 期适应性设计需要考虑多种因素。由于期中分析时对第 1 阶段结果可能无法进行全面深入地分析，加之选择一个 III 期试验的剂量还会涉及很多其他因素，以及适应性设计可能带来的设计、操作和执行方面的困难，因此，如果对试验药物了解甚少，一般不宜选择两阶段无缝适应性设计。还有一些情形，采用两阶段无缝适应性设计可能会有较大的风险。例如，III 期试验的主要终点需要较长的随访时间，第 1 阶段或许只能够用替代终点进行判断，当替代终点与主要终点关联性不高甚至较差时，用替代终点选择 III 期试验的剂量会带来很大的不确定性。又如，在达到主要终点所需要的时间较长而入组时间又较短的情况下，为避免过多的受试者进入未选中的剂量组，需要暂停入组等待期中分析的结果，由此可能会引发出其他的问题。

以上讨论的两阶段适应性设计的方法也可以直接应用于其他类似的试验，例如第 1 阶段涉及不同药物的选择，或者联合用药和单药的选择。

（四）适应性富集设计

以两阶段无缝适应性设计为背景，适应性富集设计(**adaptive enrichment design**)是指试验第 1 阶段结束后，根据期中分析的结果，依据预先设定的标准对目标人群进行适应性调整，以决定第 2 阶段的目标人群。试验的第 2 阶段可能继续在全人群中进行，或者仅入组亚群并有可能需要做一些对应的适应性调整，或者加大样本量继续入组全人群，这同时也自然地加大了亚群的入组率。试验的最终分析目标可能仅是全人群，或可能仅是亚群，或可能是全人群和亚群，其侧重点则由 α 的分配比例来决定。试验的最终分析将包含试验的两个阶段入组的所有受试者的数据，并有相应的调整方法以控制整体 I 类错误率。

如果已知试验药物只对某特定亚群有效，那么，临床试验应该只在该亚群中招募受试者。但实际中更为常见的情形是试验药物有可能对某亚群有较

大的疗效，但不清楚对全人群是否也有足够大的疗效。在这种情况下，如果试验药物对全人群有足够大的疗效，只入组亚群受试者就会失去显示对全人群有效的机会；如果试验药物对全人群疗效较小但对某亚群有效，入组全人群受试者极有可能得不到预期的阳性结果，同时也失去了显示对亚群有效的机会。采用两阶段无缝适应性设计来选择目标人群，利用试验本身的结果以便可以更科学地选择出目标人群，增加药物研发的成功率。

因为适应性设计中目标人群的选择涉及全人群和亚群，如果第 1 阶段的期中分析采用非盲态下的组间比较，应明确定义两个人群的统计假设和相应的统计方法，以控制整体 I 类错误率。

对于目标人群的选择标准，可以基于疾病特征、预后生物标志物或预测生物标志物等各种标准。一般而言，采用公认的疾病相关特征或预后相关生物标志物来选定目标人群，试验的设计和操作简单。目前，采用预测生物标志物来选择目标人群的研究日趋增多，但许多预测生物标志物的临床价值尚不明确。如果试验要用一个全新的预测生物标志物来选择目标人群，必须要有对应的诊断方法。所用诊断方法必须已经被监管部门批准上市，如果没有，可能需要同时研发。如果试验数据最终不支持所研发的体外诊断试剂（未能获得器械监管部门批准上市申请），会直接导致试验中关于试验药物的结论也不可靠。此外，如果需要同时研发伴随治疗的体外诊断试剂尚未确定亚群界定的阈值，若使用部分前期入组的受试者来确定阈值，这些受试者应该从最终的分析数据集中剔除，以免最终的阳性结果难以解释。为更好地理解生物标志物预测的准确性和全面评估试验的结果，通常鼓励将非目标人群的信息也纳入研究。

在未能充分了解亚群中药物效应的情况下，决定是否采用适应性设计选择目标人群较为困难，同时也有较大的风险。如果无视有疗效的目标亚群，

在全人群中可能得不到阳性结果，或者即使在全人群中得到了阳性结果，但在非目标亚群患者中实际无效，从而导致伦理问题。另一方面，如果实际有疗效的目标人群为全人群和若干亚群，而通过试验仅仅选择出了其中的一个亚群，导致有效的药物得不到充分地利用。

（五）适应性主方案试验设计

主方案试验设计是指一个整体临床试验方案含有多个子方案，不同的子方案可同时检验一种药物对于多种疾病的临床效果，也可同时检验多种药物对于一种疾病的临床效果，或者同时检验多种药物对于多种疾病的临床效果。每一个子方案可以是单臂试验，也可以是随机对照试验。如果子方案都是随机对照试验，各个子方案有可能共用一个对照组，也可能有各自的对照组。主方案试验也用来泛指由患者特定特征（如疾病、组织学类型、分子标记物）为标志的临床试验。主方案试验具有很多优点，例如能够为患者提供最大的入组机会并选择最合适的受试药物的机会。常见的主方案设计包括篮式试验、伞式试验和平台试验。

适应性主方案设计是指在主方案设计中包含了一种或多种适应性调整的设计，它可以灵活地采用多种适应性调整，例如添加一个或多个新的子方案，提前结束一个或多个子方案，重新估计样本量，调整检验的假设、主要终点和主要统计方法，或对不同的子方案设计做不同的适应性调整。

主方案试验需要统一管理，设有统一管理机构，如机构审查委员会、独立数据监查委员会、独立审查委员会。实施时需要采用中央随机化系统、电子数据采集系统、中央标本库和实验室，使用统一设计的病例报告表、知情同意书和临床监查人员等。

篮式设计旨在评估一种药物治疗具有同一种生物学特征的不同疾病类型的临床效果，该设计在一个整体临床试验方案中含有多个子方案，通常每一

个子方案都是一个单臂试验，而每一个子方案都针对一种或多种疾病类型。在肿瘤学领域，有时使用肿瘤组织学进行传统的试验非常困难或很难实现，而基因测序和全基因组分析已经实现了详细的分子生物学分类，篮式试验就是为了适应这一新的分类变化提出的。

伞式设计旨在评估多种药物针对同一种疾病或生物标记物类型的靶向治疗的临床效果，该设计在一个整体临床试验方案中含有多个子方案，每一个子方案针对一种或多种药物，可能是单臂试验或随机对照试验。伞式试验常用来为确证性研究选择候选药物，也可用来作为确证性研究。

平台设计旨在评估多种治疗针对多种疾病的临床疗效，该设计在一个整体临床试验方案中包含有多个子方案，每一个子方案都是一个随机对照试验，一般会采用同一个对照组。平台试验通常会维持试验长期进行，并允许新的试验药物随时加入试验平台，同时，对照药物随着时间推移也可能变更。

主方案试验虽然具有很多优点，但由于其复杂性，在计划、执行、统一管理结构的建立、尤其是统计分析等方面都面临着较大地挑战。如果计划采用主方案试验，应在试验各个方面可能涉及的各种问题做全面、深入和细致的研究后，再慎重选择。

（六）多重适应性设计

多重适应性设计是指一个试验中采用两种或两种以上适应性调整方法的试验设计。以上所讨论的适应性设计方法都可以同时用于同一个临床试验。例如，一个临床试验在第 1 阶段结束时确定了下阶段的用药剂量，其后需要做样本量重新估计，再其后需要选择目标人群。

原则上讲，如果一个临床试验设计包含了多种适应性调整，只要符合合理性、完整性和可行性的要求，多重适应性设计都可以考虑。由于多重适应

性设计的复杂性，建议慎重地考虑一个试验中是否的确有必要引入过多的适应性调整。

四、适应性设计的其他考虑

（一）贝叶斯方法的应用

贝叶斯适应性设计是指一个使用了贝叶斯方法并同时含有适应性调整的试验设计。贝叶斯方法是一类统计方法，它根据贝叶斯原理将一个分布函数（先验分布）总结的前期试验的信息/数据和目前试验中得到的数据结合在一起，从而得到一个用来总结这些信息/数据的新的分布函数（后验分布），并基于此后验分布函数做出统计推断的方法。前期试验的信息/数据可以基于目前试验中将要检验的药物，也可以基于其他相关的药物。

在临床试验中，首要任务是要得到一个准确和可靠的药物疗效估计。有时可以用一个先验分布来总结前期试验的信息/数据得到一个药物疗效的初始估计。因为前期试验的信息/数据不够充分或有其他的不确定性，其本身不能够得到一个准确和可靠的疗效估计，需要在目前的试验中收集更多的数据。根据新收集的数据，对疗效的初始估计（先验分布）进行更新并得到一个新的估计（后验分布）。用贝叶斯方法得出的疗效估计常可视为前期试验的信息/数据和目前试验的数据以某种特定方式而做出的加权平均，即如果没有目前试验的数据，疗效估计就会全部基于前期试验的信息/数据；如果有目前试验的数据，疗效估计就是一个加权平均。目前试验的数据的权重会随着其数据量的增加而加大，并向 1 趋近。

本指导原则中讨论的使用基于频率理论的适应性设计大都可用于贝叶斯适应性设计。由于贝叶斯方法在统计推断中使用了前期或相关的信息/数据，在某些方面它自然有其优越性。贝叶斯方法的灵活性是可用一些统计模型来借用相关的数据。在很多情形下，独自进行一个达到合适样本量的临床试验

会比较困难，若用贝叶斯方法来借用相关的数据从而得到更为可信的结论或许有其必要性。例如，儿童的临床试验中借用成人临床试验中的数据；在罕见病中由于无法入组足够多的病人而借用类似的疾病适应症的数据；在某一区域没有入组足够的病人而借用相邻区域的数据；在一个非劣性临床试验中借用过去试验中的数据从而减少对照组的病例数。贝叶斯方法对这些借用都会给出定量的分析和解释。

尽管贝叶斯方法在某些方面具有优越性，其最大的问题是统计推断的结果有不确定性。使用同样的前期试验的信息/数据和目前试验的数据，若选择不同的先验分布或者即便选择相同的先验分布而使用不同的参数值，贝叶斯推断也可能得出不同的结论。另外贝叶斯方法在最终的统计推断时也无公认的方法来选择决策标准。因为这些问题，目前贝叶斯方法更多地用于 I 期临床试验中药物剂量的探索，II 期临床试验中用于选择后续研发策略，III 期临床试验中的期中分析时的无效性推断和一些预测分析，以及其他很多不以注册为目的的分析。

由于适应性设计的复杂性以及基于频率理论的统计方法的局限性，尽管贝叶斯方法有其不足，在一些设计中使用贝叶斯方法或许是一种更为合适的选择。如果使用贝叶斯方法，需要有足够的前期信息/数据、文献和研究以支持所使用的统计模型的合理性，包括所选择的先验分布及各个参数值。另外由于贝叶斯推断基于先验分布和参数值的选择而导致的不确定性，还需要通过大量模拟结果来说明在各种假想的，有可能在实际中发生的情形下方案的实施特性，特别是要通过模拟显示试验中基于后验概率所定义的决策标准是否合理，例如可用对应于基于频率理论的统计方法的整体 I 类错误率来衡量所选择的决策标准。再者，还要考虑使用贝叶斯方法在实际中的可行性，例如如何向研究者解释各种统计模型的意义，基于后验概率所定义的决策标准

的意义，药物疗效估计的解释，基于不等应答适应性概率的随机分组是否给受试者带来额外的安全性风险，以及更新应答适应性概率所带来的延迟是否会使入组的实际操作变得极其困难等问题。这里的应答适应性概率是指根据已入组的病人的数据计算出的各个试验组的疗效，并以此概率来重新更新未来病人的随机入组的比例。

鉴于贝叶斯方法中面临的许多挑战，若采用贝叶斯方法，需要仔细和慎重考虑以上所讨论的各种问题，事先做好研究和规划。

（二）模拟方法的应用

基于模拟方法的适应性设计是指在适应性试验中，基于模拟方法考察所做统计推断的合理性。在临床试验中，统计检验是在统计假设下基于某个分布理论或近似的正态分布理论而做出的统计推断。这些分布理论或近似的正态分布理论所要求的条件在传统的临床试验中一般都会得到满足。为了适应药物研发的需求，许多新颖、复杂的试验设计不断出现，例如主方案试验同时涉及多个目标人群、多个假设、多个终点或多重检验，这对推导统计检验的分布理论提出了新的挑战。在很多极其复杂的试验中，基于分布理论的条件有可能不再满足，因此借助模拟方法来建立统计推断所需要的依据经常是唯一的選擇。

统计模拟试验的最大优点是在假设的临床试验情形下能更好地了解试验特性。具体到临床试验的模拟，重要的是，怎样选择模拟的模型和参数使其尽可能合理地描述试验在实际中发生的情形，并能够控制整体 I 类错误率。因为有很多关于统计模拟的文献可供参考，其涉及的模拟方法、计算机语言、模拟软件、模拟误差的控制等诸多计算细节就不在此一一讨论。

若无明确的分布理论依据，通过模拟方法来证明临床试验中统计检验的整体 I 类错误率能在零假设下完全得到控制从理论上来说是不可能的。整体 I

类错误率涉及整个零假设空间，即假设试验和对照药物疗效相同，这从理论上来说有无穷的可能性，故任何一个模拟都无法穷举所有的情形以便验证。怎样在模拟中排除一些明显不合理的情形使其更符合实际，则需要从疾病特征和历史数据来考虑，并使基于缩小的零假设空间的模拟结果从统计角度仍具有可靠性。另外在模拟时，除了考虑怎样选择主要参数外，还要考虑干扰参数、入组率、脱落率/删失率、随访时间和模拟的准确率等诸多其他因素。在选择这些参数后，再加入适应性设计中涉及的各种修改，以及有可能涉及的多个目标人群、多个终点和多重检验，以便显示所提出的统计方法在临床试验中经过多重调整后的 I 类错误率仍可得到控制。

鉴于基于模拟方法的统计推断具有不确定性，除非应用的适应性设计非常有必要，并且比传统的临床试验确有很大的优势，否则需要综合各个方面的因素再进行慎重地选择。如果有充足的医学文献、前期的数据等证据显示应用适应性设计具有必要性，并且可靠的模拟方法、模拟结果显示适应性设计确实具有很大的优越性，则可以考虑一个基于模拟方法的适应性设计。

五、监管的考虑

鉴于适应性设计的复杂性，申办方应在设计试验时及早与监管部门沟通，以便争取更多的时间用来完善方案。

（一）与监管部门的沟通

对于以探索性研究为目的的适应性设计，不一定需要与监管部门沟通交流。但是，若试验有可能影响很多受试者的安全性，如病例数较大的主方案试验；或试验的早期以探索性研究为目的，但晚期有可能演化成以确证性研究为目的，则有必要与监管部门进行沟通交流。通常以确证性研究为目的的适应性设计都应监管部门提前进行沟通交流，以便在设计早期有充分的时间

间考虑监管部门的建议、顾虑或意见，尤其是对于复杂或采用了全新方法的设计。凡经备案的与监管部门达成的一致意见应反映在修订的试验方案中。

（二）对申报资料的要求

申办方准备递交的资料应包含所有用来支持采用适应性设计的理论、文献和数据以便监管部门审评。资料的准备应主要围绕预先设定的调整计划，充分论述医学意义、合理性、完整性和可行性。

医学意义是判断采用适应性设计是否合适的重要因素。申报资料中应包含用来支持调整后试验结果仍然具有医学意义的充分证据。比如，在经过一次或多次调整后，对试验结果的解释有可能变得非常困难，或者试验结果最终达到统计学阳性但无医学意义。

合理性主要针对的是统计方法，最重要的判断标准是所使用的统计方法能否将整体 I 类错误率控制在双侧 0.05（或单侧 0.025）的水平。申报资料中应包括预先设定的调整计划，所有调整的程序和细节，以及所引用的全部文献。如果适应性设计极为复杂且无具体的理论公式，可能需要通过模拟方法来说明。申办方在计划时需要考虑模拟结果能否经过第三方独立的验证。

完整性考虑的是试验的操作和执行，判断标准是所使用的设计不会因操作或执行引入偏倚。申报资料中应包含所有操作流程，特别是怎样设定防火墙以保证分析结果不被泄露。对于数据监查委员会负责的适应性调整，可参考其他相关的指导原则。

可行性旨在评估试验可能要做的适应性修改能否在实际中实施，这一点需要申办方做出综合判断。

以上只是申报资料应该包含的基本内容。如果申办方认为还有其它资料有助于与监管部门的沟通交流，也可一并申报递交。

（三）其他的考虑

426 原则上，适应性设计中的修改计划必须在临床试验开始前的试验方案和
427 统计分析计划中预先设定。一般来说不建议对试验进行非预先设定的修改。
428 但在临床试验的操作实践中，基于本试验中的数据提示需要对试验进行非预
429 先设定的修改，此时应经过谨慎考虑后在保证不破坏试验的合理性、完整性
430 及可行性的前提下对试验做出合理的修改，并需提前与监管机构进行沟通确
431 认。此外，基于外部数据对于一个进行中的临床试验做出某些修改，并不属
432 于适应性修改，而是通过试验方案的修正案来体现，通常需与监管部门进行
433 沟通。根据外部数据对试验方案进行修改的情形有很多，例如，某试验中对
434 标志物为阴性的病人药物毒性太大或者同一类药物刚完成的试验显示仅对
435 标志物为阳性的病人有效，需要修改目标人群仅为阳性病人；同一类药物刚完
436 成的试验显示某主要终点的选择不合适，或新发布的相应指南对主要终点另
437 有定义，需要修改主要终点；新的标准治疗的改变，需要修改标准治疗的
438 对照组，或因无法继续入组病人而需要提前终止试验。申办方尤其要注意这些
439 修改是仅基于外部数据，而非基于进行中的试验本身的结果。

1. CDE. 临床试验数据监查委员会的建立和实施指导原则. 2020
2. Bauer P, Köhne K. Evaluation of experiments with adaptive interim analyses. *Biometrics*1994; 50:1029–1041, correction in *Biometrics*1996; 52:380.
3. Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine* 1999; 18: 1833-1848.
4. Bauer P, Posch M. Letter to the Editor. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections by H. Schaefer and H.-H. Mueller, *Statistics in Medicine* 2001; 20: 3741–3751. *Statistics in Medicine* 2004; 23: 1333–1335.
5. Chang M. Adaptive design method based on sum of p-values. *Statistics in Medicine* 2007; 26:2772–2784.
6. Chen C, Li X, Li W, Beckman RA. Adaptive Expansion of Biomarker Populations in Phase 3 Clinical Trials. *Contemporary Clinical Trials*2018;71:18-85.
7. Chen JYH, DeMets DL, Lan GKK. Increasing the sample size when the unblinded interim results is promising. *Statistics in Medicine* 2004; 23:1023-1038.
8. Chow SC, Chang M. Adaptive Design Methods in Clinical Trials. *CRC Press*;2nd edition 2011.
9. Chow SC, Shao J, Wang H, Locknygina Y. Sample size calculations in clinical research. *Chapman & Hall/CRC, 3rd edition*2018.
10. Cui L, Hung HMJ, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics* 1999; 55:853-857.
11. EMA Scientific Guidelines: Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. <https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines;2007>.
12. FDA Draft Guidance for Industry. Adaptive Designs for clinical trials of drugs and biologics. <https://www.fda.gov/drugs/guidances-drugs/all-guidances-drugs2018>.
13. Friede T, Kieser M. Sample size recalculation in internal pilot study designs: a review. *Biometrical Journal*2006;48:537–555.
14. Friede T, Parsons N, Stallard N. A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in Medicine*2012; 31:4309–4320.
15. Friede T, Stallard Nigel. A comparison of methods for adaptive treatment selection. *Biometrical Journal* 2008;50:767–781.
16. Gould AL. Interim analyses for monitoring clinical trials that do not materially affect the Type I error rate. *Statistics in Medicine* 1992; 14:1039-1051.
17. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*1988; 75: 800–802.
18. Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical Statistics* 2011; 10:347–356.
19. Lan KG, DeMets DL. Discrete Sequential Boundaries for Clinical Trials. *Biometrika*1983; 70:659–663.
20. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics* 1999; 55: 1286–1290.
21. Maca J, Bhattacharya S, Dragalin S. et al. Adaptive Seamless Phase II/III Designs Background, Operational Aspects, and Examples. *Drug Information Journal* 2006; 40: 463-474.
22. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*1976; 63:655–660.
23. Müller HH, Schäfer H. Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*2001; 57:886–891.
24. Müller HH, Schäfer H. A general statistical principle for changing a design any time during the course of a trial. *Statistics in Medicine*2004; 23:2497–2508.
25. O'Brien PC, Fleming TR. A Multiple Testing Procedure for Clinical Trials. *Biometrics* 1979; 549–556.

26. Pocock SJ, Group Sequential Methods in the Design and Analysis of Clinical Trials. *Biometrika* 1977; 64:191–199.
27. Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics* 1995; 51:1315–1324.
28. Rosenblum M, Van Der Laan MJ. Optimizing randomized trial designs to distinguish which subpopulations benefit from treatment. *Biometrika* 2011; 98: 845-860.
29. Shih WJ. Sample size re-estimation – a journey for a decade. *Statistics in Medicine* 2001; 20:515-518.
30. Shih WJ, Li G, Wang Y. Methods for flexible sample-size design in clinical trials: Likelihood, weighted, dual test, and promising zone approaches. *Contemporary Clinical Trials* 2016; 47: 40-48.
31. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 1986; 73:751–754.
32. Stallard N, Hamborg T, Parsons N, Friede T. Adaptive designs for confirmatory clinical trials with subgroup selection. *Journal of Biopharmaceutical Statistics* 2014; 24:168–187.
33. Wang SJ, Hung HMJ, O'Neill RT. Adaptive patient enrichment designs in therapeutic trials. *Biometrical Journal* 2009; 51:358–374.
34. Wassmer G, Brannath W. Group sequential and confirmatory adaptive designs in clinical trials. *Springer*; 2016.
35. Wu PS, Lin M, Chow SC. On sample size estimation and re-estimation adjusting for variability in confirmatory trials. *Journal of Biopharmaceutical Statistics* 2016; 26:44-54.
36. Zhang JJ, Blumenthal G, He K, Tang S, Cortazar P, Sridhara R. Overestimation of the effect size in group sequential trials. *Clinical Cancer Research* 2012; 18: 4872-4876.

附录 1、词汇表

期中分析 (Interim analysis)：是指在试验期间使用试验累积数据进行的任何分析。同一试验中，可以设计有一次或多次期中分析。

适应性设计 (Adaptive design)：按照预先设定的计划，在期中分析时根据试验期间累积的数据对试验做出相应修改的临床试验设计，这种修改又称为适应性修改。适应性修改计划必须在临床试验开始前的试验方案和统计分析计划中预先设定。

合理性 (Validity)：适应性设计的合理性是指所采用的统计分析方法不会导致疗效的估计产生偏倚。试验的合理性是关于试验结果的可信度、可解释性和说服力。

完整性 (Integrity)：适应性试验的完整性是指对试验操作所引入偏倚的良好控制。保持试验的完整性意味着按照预先设定的方案进行修改和保持期中分析结果的盲态，以求最大限度地减少操作偏倚。

可行性 (Feasibility)：适应性试验的可行性是指试验的适应性修改能否在实际中实施。

成组序贯设计 (Group sequential design)：是指方案中预先计划在试验过程中进行一次或多次期中分析，依据每一次期中分析的结果做出后续试验的决策。

盲态/非比较分析方法 (Blinded/Non-comparative analysis)：是指期中分析时不使用实际试验分组信息，或者虽然已知实际试验分组信息，但未做任何涉及组间比较的分析，如在期中分析时对两个治疗组的数据合并后做的汇总分析。

534 **非盲态分析/比较分析方法 (Non-blinded/Comparative analysis)**：是指期中
535 分析时使用试验分组信息（包括各组的真实名称或可区分的分组代码）的分
536 析，分析内容涉及组间比较。

537 **两阶段无缝适应性设计 (Two-stage seamless adaptive design)**：是指将一个
538 试验分为两个阶段，在第 1 阶段结束时进行期中分析，依据预先设定的判断
539 标准，对第 2 阶段的试验进行适应性修改。

540 **适应性富集设计 (Adaptive enrichment design)**：是指试验第 1 阶段结束后，
541 根据期中分析结果，依据预先设定的标准对目标人群进行适应性调整，以决
542 定第 2 阶段的目标人群。

543 **主方案试验设计 (Master protocol design)**：是指一个整体临床试验方案含
544 有多个子方案，不同的子方案可同时检验一种药物对于多种疾病的临床效果，
545 也可同时检验多种药物对于一种疾病的临床效果，或者同时检验多种药物对
546 于多种疾病的临床效果。

547 **适应性主方案设计 (Master protocol with adaptive designs)**：是指在主方案
548 设计中包含了一种或多种适应性调整的设计。

549 **多重适应性设计 (Multiple adaptive design)**：是指一个试验中采用两种或
550 两种以上适应性调整方法的试验设计。

551 **贝叶斯方法 (Bayesian method)**：是一类统计方法，它根据贝叶斯原理将一
552 个分布函数（先验分布）总结的前期试验的信息/数据和目前试验中得到的数
553 据结合在一起，从而得到一个用来总结这些信息/数据的新的分布函数（后验
554 分布），并基于此后验分布函数做出统计推断的方法。

555 **贝叶斯适应性设计 (Bayesian adaptive design)**：是指一个使用了贝叶斯方
556 法并同时含有适应性调整的试验设计。

557 **基于模拟方法的适应性设计 (Simulation-based adaptive design)**：是指在适应

558 性试验中，基于模拟方法考察所做统计推断的合理性。

附录 2、适应性设计案例

示例中涉及的各种试验设计内容、统计方法选择，以及具体参数设置仅为便于简洁阐述设计要点，并不具有普遍适用性和全面性。

示例 1：成组序贯设计试验

假想一个多中心、随机、双盲、阳性药平行对照的 III 期优效性临床试验，目的是验证试验药对某适应症的症状改善效果，主要终点是治疗 52 周时某连续型变量较基线的变化值，并可假设该变量近似服从正态分布。

试验采用成组序贯设计，期中分析计划为：在入组 n_1 例受试者时进行第一次期中分析（安全性分析）， n_2 例时进行第二次期中分析（无效性分析）， n_3 例时进行第三次期中分析（有效性或无效性分析）。

基于上述设计，根据预期的效应参数（如两组变化值的均数及共同标准差）、检验效能（如 90%）、各组分配比例、设定的整体 I 类错误率水平（如单侧 0.025）及相应的控制方法等，可计算出试验所需的总样本量，并制定三次期中分析的决策策略如下：

1) 第一次期中分析时，由独立数据监查委员会（DMC）综合分析安全性数据，如果 DMC 认为试验药物存在较严重的安全性问题，可建议终止试验。此次期中分析仅以安全性为目的，因此无需进行 α 调整。

2) 第二次期中分析时，考察对照组的主要终点变化值相较于试验组是否更优，若对照组较优，则因试验药物无效性提前终止试验。此次期中分析仅考察无效性，不涉及有效性判断，因此无需进行 α 调整。

3) 第三次期中分析时，采用非绑定边界方法，将针对试验药物疗效的统计检验所得 P 值与两条边界进行比较，边界值的确定基于某一 α 和 β 的消耗函数方法，如果 P 值小于下边界，则以有效性提前终止试验；如果 P 值大于上

边界,则由 DMC 经综合考虑后提出试验终止或继续进行的建议;如果 P 值在上下界值之间,则试验继续进行。

示例 2: 盲态下样本量重估

假想一个随机、双盲、安慰剂平行对照的 III 期优效性临床试验,主要终点为访视第 4 周时某量表评分相对于基线的变化值,并假设其近似服从正态分布。预期试验组与对照组主要终点指标之差的总体均数为 6.0,标准差为 10.0,设定名义检验水准为单侧 0.025,检验效能为 90%,采用两组平衡设计,求得总样本量为 120 例(每组 60 例)。假设 20%的脱落率,则初始设计的总样本量为 150 例。考虑到将参数标准差预设 10.0 可能不准确,故计划利用期中分析时进行样本量重估。

期中分析计划为:在完成或中止试验的受试者达到约 50%初始样本量时进行盲态下期中分析,计算累积数据的合并标准差,若求得的合并标准差大于原假定的 10.0,则根据此参数重新估算最终的总样本量,估算时其它参数与初始假设保持一致;若求得的合并标准差小于 10.0,则样本量保持 150 例不变。根据期中分析重估的样本量或原始样本量,继续入组受试者直至试验结束,不对 I 类错误率进行调整。假设进行期中分析时求得的合并标准差为 13.66,根据标准差为 13.66,均数与初始设计一样为 6.00,重新计算的样本量为 220,考虑 20%的脱落率,则总样本量重估为 275 例。

示例 3: 非盲态下样本量重估

假想一个多中心、随机、双盲、阳性药平行对照的 III 期优效性临床试验。主要终点是 24 周时某标准分数较基线的变化值,并假设其近似服从正态分布,疗效分析采用 Z 检验。根据试验预期参数(如两组样本均值之差为 δ_0 ,标准差 $\sigma=1$)和其它所需要素(如 $\alpha=0.025$,检验效能 $1-\beta=90\%$),确定初始总样本量为 N 。

假设计划在完成 n_1 例时进行期中分析并重新估计样本量。用 $n_2=N-n_1$ 表示初始设计下第 2 阶段的样本量，而根据期中分析组间比较结果求得的第 2 阶段所需样本量表示为 n_2^* ，则 $n_2^*-n_2$ 即是需要增加的样本量， N^* 表示增加后的总样本量。

n_2^* 的确定需要采用合理的适应性修改方法。由于不同的适应性修改方法各有优劣，且有些方法在某些条件下可能是其它方法的特例，因此难以将方法进行明确划分，具体选用哪种方法应根据具体试验目的、假设和分析方法而定，也可以选择模拟分析的方法。以下简略地介绍这些常用方法的设计要点。需注意，鉴于研发费用的限制，很多方法在计算时会设置样本量上限；而考虑到最小检验疗效的临床意义，也有很多方法在计算时对其进行了考虑。以下介绍并未将这两点因素考虑在内。

1) 希望区域 (promising zone) 方法：根据期中分析结果计算条件检验效能 $CP(N, z_1)$ ，并将其分为三个区域，即 $(0, c_l]$ ， (c_l, c_u) 和 $[c_u, 1)$ ， z_1 为仅基于第 1 阶段 n_1 例数据得出的 Z 统计量。若 $CP(N, z_1) \leq c_l$ ，可终止试验；若 $CP(N, z_1) \geq c_u$ ，可继续试验；若 $c_l < CP(N, z_1) < c_u$ ，重新估计样本量：在满足 $CP(N, z_1) > 50\%$ 时，样本量的重新估计不会增加 I 类错误率。对于样本量的重新估计，可利用期中分析所得两组疗效之差代入样本量计算公式，计算新的样本量 N^* ，使之满足 $CP(N^*, z_1) = 1 - \beta$ 。最终分析时基于调整后样本量 N^* 计算 Z 检验统计量，与 $z_{1-\alpha}$ （不需要调整）进行比较以判断其有效性。

2) 加权统计量方法：也可视为一种逆正态分布组合函数方法，但更简便。根据期中分析结果，基于某种标准（如条件检验效能）计算出调整后样本量 N^* 。令 $w_1 = \sqrt{n_1/(n_1 + n_2)}$ ， $w_2 = \sqrt{n_2/(n_1 + n_2)}$ ，则最终的检验统计量为 $Z^* = w_1 Z_1 + w_2 Z_2^*$ ，并与 $z_{1-\alpha}$ （不需要调整）进行比较，这里， Z^* 服从标

准正态分布， Z_1 为仅基于第1阶段 n_1 例数据得出的 Z 统计量， Z_2^* 为仅基于调整后第2阶段 n_2^* 例数据得出的 Z 统计量。需注意上述权重的计算基于试验初始设计的 n_1 和 n_2 ，但这类加权统计量方法也可以只指定 n_1 而不需要确定初始总样本量 N 。

3) 极大似然比方法：设定拟达到的目标 $CP(N^*, z_1)$ 值，据此调整拒绝域的临界点 c ，其由公式 $1 - \alpha = \int_{-\infty}^k \Phi\left[\frac{c\sqrt{n_1+n_2^*}-z_1\sqrt{n_1}}{\sqrt{n_2^*}}\right]\phi(z_1)dz_1$ 决定（式中 $\Phi(\cdot)$ 和 $\phi(\cdot)$ 分别表示标准正态分布的累积分布函数和概率密度函数），这里 $z_1 < k$ ， k 基于所选 α 消耗函数方法（若有有效性检验）决定。最终分析基于调整后两阶段的样本计算 Z 统计量，并与调整后的临界点 c 进行比较。

4) 条件误差函数方法：选择满足一定条件的递增函数 $A(z)$ ，记 $z_A = \Phi^{-1}(1 - A(z_1))$ ，则 $n_2^* = 2(z_A + z_\beta)^2 / \delta_1^2$ ， δ_1 是期中分析所得疗效估计值， z_β 是标准正态分布的上 β 分位数，则 $n_2^* - n_2$ 即是需要增加的样本量。最终分析利用调整后第2阶段病例计算的 Z_2^* 值与 $(z_1^2 + (z_A + z_\beta)z_A) / \sqrt{z_1^2 + (z_A + z_\beta)^2}$ 进行比较，判断有效性。

5) 基于 P 值之和的方法： $N^* = |\delta_0 / \delta_1|^2 N$ ，这里 δ_1 是基于期中分析数据估计的疗效值。基于 n_1 和 n_2^* 分别计算两个阶段的 P 值 p_1 和 p_2 ，将 $p_1 + p_2$ 与0.2236进行比较，判断有效性。此方法要求期中分析数据所得的疗效估计值 δ_1 不能与初始设定的 δ_0 反方向，且不能减少样本量。

示例 4：两阶段无缝 II/III 期适应性设计试验

假想一个验证某新药用于改善某症状的多中心、随机、双盲、阳性药平行对照、优效性临床试验，主要终点是第8周时某种分数较基线的变化值，并假设其近似服从正态分布。计划采用两阶段无缝 II/III 期的适应性设计，II

期试验包含两个药物组（高和低剂量）和一个对照组，随机分组率为 1:1:1, 预期疗效为两组受试者间平均改进分数差值 δ ，设定单侧 $\alpha=0.025$ 和检验效能 $1-\beta$ （如 90%），计算出一个剂量组和对照组比较的样本量 N 。试验的第 $3n_1$ 个受试者完成 8 周随访后第 1 阶段结束，总样本量为 $N+n_1$ （两比较组总受试者数 N 加未选上剂量组在第 1 阶段的受试者数 n_1 ）。

对于第 1 阶段，令 p^{11} 和 p^{12} 分别表示低剂量组 vs.对照组(零假设为 H_0^{11} : 低剂量组与对照组无差别)和高剂量组 vs.对照组（零假设为 H_0^{12} : 高剂量组与对照组无差别）Z 检验的 P 值结果，采用闭合方法和 Hochberg 方法进行多重性调整，则检验低、高剂量组与对照组均无差别 ($H_0^{11} \cap H_0^{12}$) 的 P 值是 $p_{\text{int}}^1 = \min[2 * \min(p^{11}, p^{12}), \max(p^{11}, p^{12})]$ 。经多重性调整后，第 1 阶段所选剂量组与对照组疗效比较的 P 值应取 $p_1 = \max(p_{\text{int}}^1, \min(p^{11}, p^{12}))$ 。

对于第 2 阶段，入组病例数为原定两组样本量 N 减去第 1 阶段两组受试者数 $2n_1$ ，即 $N-2n_1$ ，将入组病例随机分配至选定的剂量组和对照组。试验终止时，分别根据第 1 阶段和第 2 阶段的受试者计算各阶段 P 值，分别记为 p_1 和 q 。采用逆正态组合检验，计算 $C(p_1, q) = 1 - \Phi[w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - q)]$ ，并与 $\alpha=0.025$ 进行比较判断有效性，这里权重 w_1 等于第 1 阶段三个组的受试者例数相对于试验入组总例数之比的平方根，即 $w_1 = \sqrt{3n_1/(N + n_1)}$ ；权重 w_2 等于第 2 阶段两个组的受试者例数相对于试验入组总例数之比的平方根，即 $w_2 = \sqrt{(N - 2n_1)/(N + n_1)}$ 。

示例 5：适应性富集设计

假想一个多中心、随机、双盲、阳性药平行对照、优效性两阶段临床试验。主要终点为总生存期(OS)，次要终点为无进展生存期（PFS）。假设全人群中风险比 HR（F）为 0.75，阳性亚组中风险比 HR（S）为 0.55。设定检

674 验水准为单侧 0.025，检验效能为 90%，并根据其它所需参数计算出在全人
675 群中所需要的总死亡数为 N_0 。

676 试验计划在入组 40%时利用期中分析进行目标人群选择。由于该试验受
677 试者总生存期较长，故用 PFS 作为期中分析时目标人群选择的依据，决策策
678 略为：①如果估计的 $HR(F) < 0.85$ 且 $HR(S) < 0.65$ ，第 2 阶段在阳性亚组
679 和全人群中继续试验；②如果 $HR(F) \geq 0.85$ 且 $HR(S) < 0.65$ ，第 2 阶段仅
680 入组阳性亚组；③如果 $HR(F) < 0.85$ 且 $HR(S) \geq 0.65$ ，第 2 阶段继续在全
681 人群中试验，且不对阳性亚组进行分析；④如果 $HR(F) \geq 0.85$ 且 $HR(S)$
682 ≥ 0.65 ，则以无效性终止试验。

683 用 p_1^F, p_1^S 和 p_1^{FS} ，以及 p_2^F, p_2^S 和 p_2^{FS} 来分别表示零假设 H_0^F, H_0^S 和
684 $H_0^F \cap H_0^S$ 在两个阶段对应的 P 值。对组合检验 $H_0^F \cap H_0^S$ 使用逆正态组合函数
685 和闭合方法，可将整体 I 类错误率控制在单侧 0.025 水平。如采用 Simes 方法，
686 则检验 $(H_0^F \cap H_0^S)$ 的 P 值为 $p^{FS} = \min(2\min(p^F, p^S), \max(p^F, p^S))$ 。记

687 $C(p, q) = 1 - \Phi[w_1 \Phi^{-1}(1 - p) + w_2 \Phi^{-1}(1 - q)]$ ，其中， $w_1 = \sqrt{\frac{n_1}{n_1 + n_2}}$ ， $w_2 =$

688 $\sqrt{\frac{n_2}{n_1 + n_2}}$ ， n_1 和 n_2 是分别表示两阶段的死亡病例数。在最终分析时，检验

689 $H_0^F \cap H_0^S$ 的 P 值为 $C(p_1^{FS}, p_2^{FS})$ ，当 $C(p_1^{FS}, p_2^{FS}) \leq 0.025$ 时，可拒绝 $H_0^F \cap H_0^S$ 。

690 如果在第 2 阶段仅入组亚组，则当 $C(p_1^{FS}, p_2^S) \leq 0.025$ 时，可拒绝 $H_0^F \cap H_0^S$ 。

691 同理，对 H_0^F 和 H_0^S 可分别利用 $C(p_1^F, p_2^F)$ 和 $C(p_1^S, p_2^S)$ 来检验。再基于闭合
692 方法，若同时拒绝 $H_0^F \cap H_0^S$ 和 H_0^F ，则可最终拒绝 H_0^F ；若同时拒绝 $H_0^F \cap H_0^S$
693 和 H_0^S ，则可最终拒绝 H_0^S 。

694 对于该试验，如果第 1 阶段的检验基于 PFS 而第 2 阶段的检验基于 OS，
695 则最终计算出的两阶段 P 值涵义难以解释，因此，试验设计为第 1 阶段的决

策基于 PFS 的描述性统计量结果，而最终计算两阶段 P 值时则均基于 OS 的 P 值结果。对于以生存期为终点的试验，无论第 1 阶段入组受试者的终点事件发生在哪个阶段，计算时均应将其归在第 1 阶段的结果内，否则，两阶段的独立性假设将不再成立，导致 I 类错误率增加。

示例 6：适应性主方案试验

假想一个验证新药治疗有 BRAF V^{600E} 正突变的罕见癌症患者的优效性临床试验，主要终点是经独立终点委员会确认的客观缓解率，并有持续缓解时间的记录。采用多中心、单臂篮式设计，试验入选受试者必须是经中心试验室确诊具有 BRAF V^{600E} 突变的晚期病例，包括 5 个队列，分别是间变性甲状腺癌、胆道癌、胃肠道间质瘤、毛细胞白血病和小肠腺癌。

在此试验中，虽然所有受试者按照同一个方案入组，但 5 个队列中的每一个队列都可被视为一个独立的试验，其结果都可单独用来支持相应队列的申报。由于试验目的是支持新药申报，因此必须事先确定样本量，应该根据优效性的决策规则分别计算每一队列所需的样本量。对于将两个或多个队列数据合并申报的考虑，由于本试验缺乏足够数据支持试验药对具有 BRAF V^{600E} 正突变的患者有同样的作用机理和相似疗效，因此，将任意两个或多个队列数据合并以支持对应合并队列新药申报的做法是不被接受的。