??????????????????????

????

与其他治疗领域一样，抗肿瘤药物在进入临床试验前，应该有足够的基于临床前实验或既往人体试验的科学证据显示某（些）剂量的试验药物在目标人群的安全性。临床试验的主要目的是针对药物研发提出相关的临床问题，通过恰当的试验设计和统计分析科学地回答这些问题。随机对照试验（Randomized Controlled Trial, RCT）是评价药物有效性和安全性的金标准，如果无法开展随机对照试验，则有效性和安全性结论的证据力度将会有所下降。由于肿瘤通常是严重危及生命的疾病，临床用药很大程度上存在未被满足的需求，所以抗肿瘤药物的临床研发有其特殊性。比如，早期临床试验以患者为研究对象，而不是健康受试者；某些情形下利用单臂试验结果申请注册上市等。针对不同肿瘤适应症，申办者应有不同的临床研发策略考虑，探索性试验和确证性试验在不同的研发项目计划中要达到的目的与作用也会不同。临床试验设计是决定研发成功与否的重要因素之一。良好的试验设计不仅有助于达到试验目的，同时还能提高研发效率。创新的临床试验设计类型和方法层出不穷，通过不断实践，抗肿瘤药物研发和审评的经验都在逐步丰富。本指导原则旨在针对抗肿瘤药物临床试验设计中的关键统计学技术问题，提供科学建议，为申办者开展抗肿瘤药物的临床研发提供参考。本指导原则仅代表当前的观点和认识，随着研究和认识的深入将不断修订和完善。

??????

抗肿瘤药物临床试验最常用的疗效终点有总生存期（Overall Survival, OS）、客观缓解率（Objective Response Rate, ORR）、无进展生存期（Progression Free Survival, PFS）等。

????????OS?

总生存期是指从随机化开始（或单臂试验中治疗开始）到任何原因导致死亡的时间。OS 相对客观并且精确可测，是随机对照临床试验中衡量抗肿瘤药物临床获益的最可靠终点。OS 通常应基于意向性治疗（Intention-To-Treat, ITT）的原则进行分析。ITT 分析应包括所有根据预先制定的研究方案参与随机化的受试者或单臂试验中接受过任何剂量药物的受试者，不考虑不依从、方案偏离、退出以及随机化或单臂试验中治疗开始后发生的任何事件。由于失访的受试者往往具有较高的死亡风险，如果两组之间的删失时间或删失比例不平衡有可能会导致分析结果产生偏倚，因此需对组间删失模式的均衡性进行评估，还应保证分析时所有受试者使用的是随访截止日期收集到的最及时更新的生存数据。

在安全性特征可接受的情况下，如果 OS 的改善具有统计学显著性和临床意义，可用来支持试验药物的常规新药上市申请。

对 OS的假设检验通常基于 log-rank检验，而 Cox 回归模型通常用于估计治疗效果（风险比）。生存率常用 Kaplan-Meier 方法计算，并用生存曲线呈现。无论事件何时发生，log-rank 检验对所有事件均赋予相同的权重。如果采用分层 log-rank

检验，分层因素一般需要从随机化分层因素中预先选定。如果有充分的依据认为风险函数呈非等比例性，也可以考虑采用其它加权方法。然而，对等比例风险假设的轻微偏离在实际中比较常见，且根据既往的临床经验预测随时间变化的风险比模式极为困难。因此，在采用加权方法之前，应充分考虑其利弊，并与监管机构沟通交流。比较不同试验的OS是不可靠的，因不同试验在患者选择、标准治疗（Standard of Care, SOC）以及最佳支持治疗（Best Supportive Care, BSC）等方面均可能存在异质性，因此在单臂试验中应谨慎使用和解读OS。

?????????ORR?

对于许多肿瘤类型，可以直接利用肿瘤影像学进行疾病评估，受试者的治疗策略通常基于肿瘤测量结果和临床症状。客观缓解率是指按照公认的缓解评价标准（如实体瘤RECIST 1.1版），肿瘤体积缩小达到预先规定值并能维持最低时限要求的患者比例，它是基于肿瘤测量的最普遍的终点。实体瘤的缓解可以是完全缓解（Complete Response, CR）或部分缓解（Partial Response, PR），对于非实体瘤的评估则有一些其它评价标准。单独使用ORR可能无法充分描述试验药物的抗肿瘤活性，故需要同时描述性分析缓解持续时间（即从初始肿瘤缓解到疾病进展或任何原因导致死亡的时间，以先发生者为准）和至缓解时间。对于通过稳定疾病病情使患者临床获益的药物，也可以分析疾病控制率（Disease Control Rate,DCR），该指标不仅考虑疾病缓解病例，还包括疾病维持稳定状态持续一定时间的病例。肿瘤大小随时间相对于基线的变化通常被视为连续变量，可利用瀑布图进行描述，帮助评估抗肿瘤活性。对于拟进行注册的试验（单臂或随机对照），肿瘤评估通常基于盲态独立中心审查委员会（Blinded Independent Central Review, BICR）的肿瘤测量和缓解评估。如ORR是主要疗效终点，初始缓解通常需要在后续的评估中确认。在临床实践中应考虑BICR和研究者评价不一致情况，因为研究者的评价结果会影响受试者的后续治疗，而这可能会对ORR分析带来偏倚。与OS一样，ORR一般也应基于ITT的原则进行分析。在ITT分析中，对于第一次肿瘤评估之前退出试验的受试者，不论何种退出原因，都被认为是非缓解者。ITT分析可以使当前试验结果与历史对照结果的比较更可靠，因为后者通常基于确证性试验中的ITT人群。原则上，缓解评估应基于与历史对照相同的缓解标准才具有可比性。如果由于历史原因不能使用相同的缓解标准时,需关注当前标准与历史标准的不同带来的影响。

??????????PFS?

无进展生存期是指从随机化开始（或单臂试验中治疗开始）至肿瘤进展或任何原因导致死亡（以先发生者为准）的时间。与PFS类似的终点还包括无病生存期（Disease Free Survival, DFS），是指从随机化开始（或单臂试验中治疗开始）至疾病复发或任何原因导致死亡的时间（以先发生者为准），多用于评价手术治疗或放疗后的辅助治疗。无事件生存期（Event Free Survival, EFS）是指从随机化开始（或单臂试验中治疗开始）到首次发生以下任何事件的时间：疾病进展而无法进行手术治疗、局部或远处复发、任何原因导致的死亡等，多用于评价手术治疗或放疗前的新辅助治疗。类似的终点还包括至疾病进展时间（Time To Progression, TTP）和至治疗失败时间（Time To Treatment Failure, TTF），这两种终点的分析通常考虑为敏感性分析，其结果不能作为确证性研究结论的主要证据，可以用于支持主要终点PFS的结果。

肿瘤进展的确切定义对基于肿瘤测量的终点至关重要，应事先在方案中予以明确，与ORR一样，进展的定义应遵循既定的缓解评估标准。PFS在单臂试验中难以解释，例如一些受试者即使没有积极治疗也可能维持很长时间的病情稳定，因此，以PFS为主要终点的注册试验应设置对照组。在随机双盲对照试验中，可依据安全性特征和肿瘤临床评估实践来决定是否需要BICR，但应保存肿瘤影像学资料以便稽查和核查。

PFS

PFS

PFS

PFS

BICR ③

④ PFS ITT

BICR PFS

PFS

PFS -

OS

PFS

PFS

Cox

PFS -

PFS

??????????PRO?

PRO

PRO

PRO

???????

?????????

I First in Human, FIH I

治疗剂量范围内治疗受试者），同时保证安全性和快速入组。I期临床试验的剂量递增方法分为两大类：一是基于规则的设计，包括传统的3+3设计及其衍生设计，不依赖于统计建模；二是基于模型的设计，如连续重新评估方法（Continuous Reassessment Method,CRM）。一些新兴的模型辅助方法如改良毒性概率区间（Modified Toxicity Probability Interval, mTPI）设计和贝叶斯最优区间（Bayesian Optimal Interval, BOIN）设计，虽然基于模型而建立，但这些方法预先指定剂量递增的规则，而且易于实施，具有选择目标毒性概率和队列大小的灵活性，还具有与基于模型的设计相当的性能。为了尽可能减少接受可能低于治疗剂量的受试者人数，I期剂量探索可从加速滴定设计开始，加速滴定部分通常在每个剂量水平招募1-3名受试者，并以发生2级或更高的非疾病相关毒性事件作为结束。加速滴定部分结束后，将采用正式的剂量递增方法进行剂量探索。在某些情况下，也可考虑采取患者内剂量递增（即受试者在后续治疗周期中的剂量水平高于其在第一个周期接受的剂量），但通常会导致第一个周期之后的安全性和耐受性数据难以解释。对于确定为候选II期推荐剂量（Recommended Phase 2 Dose, RP2D），应有足够数量的受试者接受了该剂量治疗。

????????????????

在抗肿瘤药物研发中，有时会在剂量探索阶段结束后在一个或多个肿瘤适应症中开展单臂试验，以进一步探索药物的安全性并初步研究药物的有效性。这些肿瘤适应症队列可以由同一治疗线次的不同肿瘤类型，或同一肿瘤类型的不同治疗线次，或两者的组合形成。队列中的受试者可以接受试验药物作为单药治疗或联合治疗（如与标准治疗或另一种试验药物联合）。单臂试验的研究方案应当包含足够的信息，以说明其基于队列研究目的所确定的样本量估计的合理性。在非随机队列中，抗肿瘤活性的评估通常采用多阶段设计来确定，以限制暴露于无效药物的患者数量。方案还应提供关于是否暂停入组的详细信息以及受试者参加期中分析的最短随访时间。若需要比较不同给药方案（如两种候选RP2D，或单药疗法和联合疗法）之间的安全性和抗肿瘤活性，则需开展更严格统计设计的随机队列。

若在FIH研究中开展单臂试验设计，开始时可能缺乏关于试验药物代谢动力学的足够数据，或未进行足够的安全性评估，此时如果快速入组，特别是在有令人兴奋的初步信号时，可能使大量受试者暴露于疗效未知和毒性特征不清楚的药物。为了减轻这种风险并保护受试者，申办者必须建立一套完善的操作流程，以方便数据收集，实时快速地评估新数据，向研究者、机构审查委员会（Institutional Review Board, IRB）及时公布期中分析结果。申办者应根据期中分析结果和统计分析计划中预定的决策规则，尽早暂停或结束抗肿瘤活性不足或安全性水平不可接受的队列入组，或提早终止失败的研究项目。

对于拟进行注册的扩展队列研究，应明确区分用于建立药物活性假设的患者群体和用于确认该假设的患者群体。为了达到验证试验假设的目的，建议对用于确认假设的患者群体开展独立的临床试验，特别是当FIH研究已对研究人群和样本量进行过多次变更时。如果试验未设置阳性对照组，则其数据必须非常有说服力，才能确证药物的有效性。因此，在设计单臂试验用于注册为目的时，需对已有数据证据和其样本量估算进行非常谨慎的评估。对于两种新型试验药物联合治疗的研究，除非对每种药物的贡献都有很好的理解，且能合理分离每种药物的单独贡献，否则不宜采用单臂试验。

???????

???????

在设计确证性试验时，申办者应根据临床试验的目的明确要估计的治疗效应。申办者在方案中应阐明研究人群、终点指标、治疗方案，应考虑试验过程中可能发生的影响治疗效应估计的伴发事件，如死亡、转组等，群体层面的汇总统计量、统计模型以及相应的敏感性分析也均应事先定义。虽然减少研究人群的异质性可能会提

高统计检验效能，但对目标人群的限制会使新药在实际应用中的效果难以评估。应基于证据选择最佳可用的治疗作为对照，因此，根据情况一般可以选择BSC、SOC或研究者选择性治疗作为对照。盲法设计是确证性试验控制偏倚的重要手段之一。如果临床试验只能使用开放设计（如因不同药物间毒性特征差异明显而使用开放设计），必须采取所有可能的措施来控制潜在的偏倚，比如对申办者试验团队遮蔽关键数据。无论采用开放设计还是双盲或单盲设计，对于重要且潜在的可能影响药物疗效的基线协变量，建议在随机化时予以考虑，对基线协变量的校正分析应在方案以及统计分析计划中事先规定。当使用预测生物标志物进行分层时，必须预先规定生物标志物及其确定生物标志物状态（阳性或阴性）的阈值，且阈值的确定方法必须经过科学验证并得到公认。确证性试验的整体I类错误率必须严格控制在一定水平。如果研究的主要目的中包括对多个人群（例如生物标志物阳性人群和所有患者人群）或多个终点（例如OS、PFS和 ORR）进行假设检验时，或者计划实施因有效而提前终止试验的期中分析时，应选择合适的多重性控制策略，并在方案以及统计分析计划中事先进行详细规定。计划因有效而提前终止试验的同时需要考虑安全性评价数据的充分性。确证性试验设计对统计学考虑要求较高，申办者应根据试验设计复杂度就确证性临床试验方案以及统计分析计划中关键技术问题与审评机构开展沟通交流。

???????

传统的研究设计用于抗肿瘤药物临床试验时，可参考如ICH E9等相关的指导原则。随着抗肿瘤药物研发的快速发展，一些新颖的试验设计在确证性试验中得以合理应用，包括适应性无缝剂量选择的设计、两阶段适应性设计、富集设计和主方案设计等，大大提高了临床研发的效率。

1. ??????

成组序贯设计通常用于按时间顺序进行的数据监测或对累积的数据进行统计推断。在设计成组序贯试验时，申办者应仔细考虑计划进行的期中分析次数和时间点，以及合适的α消耗函数。对于因有效性而提前停止的试验，鼓励申办者继续随访试验直至数据成熟，以更好地了解试验药物的长期临床获益。当期中分析或最终分析的时间点是由事件驱动时，主要数据集的确定应基于达到目标事件数量时的截止日期。在揭盲分析之前应确保在盲态状态下完成数据的收集和清理。由于可能存在收集偏倚，揭盲之后收集的数据将受到严格审查，甚至从分析集中排除。

2. ????????

传统药物研发遵循先进行II期试验，再进行III期试验的序贯方法。II期试验用于临床概念验证、剂量选择、人群选择甚至终点选择。在获得II期数据后会决定是否开始III期研究。III期试验需要时间来计划、启动和实施。适应性无缝II/III期设计作为两阶段适应性设计中的一个特例，试图消除II期和III期试验之间的空白期。可以采用操作无缝设计，将II期试验受试者排除在主要分析之外，也可以采用推断无缝设计，在主要分析中纳入II期试验受试者。前者不需要对I类错误的控制进行多重性调整，但对于后者，则可能需要根据适应性的性质和假设检验策略做出相应的调整。在决定采取无缝设计而不是序贯设计之前，应考虑两个重要因素。首先，从II期试验无缝过渡到III期试验时有足够的信息来支持合理决策。这通常取决于参与II期数据分析的受试者人数以及II期试验采用的终点指标是否对决策有帮助。第二，操作层面上可以顺利实施。无缝设计要求能够迅速地对数据进行清理和分析，快速增加III期入组人数，并能够加快药物上市的进程。决定采用操作无缝设计或者推断无缝设计的关键考虑因素在于II期试验中适应性决策的复杂性。一般来说，和操作无缝设计不同，II期和III期试验结果之间的一致性对推断无缝设计来说更为重要。虽然适应性无缝II/III期设计在加速药物研发方面具有很好的前景，但在采取此策略之前，需要全面权衡不同方法的优缺点。在开始

试验之前，需要解决试验设计、操作和统计分析等方面的问题，并与监管部门进行沟通。

3. ????

为了优化试验药物的获益-风险特征，确定适当的目标人群至关重要。合适的生物标志物可以通过各种不同的诊断方法（如转录物的表达谱分析、差异抗原表达、遗传诊断，包括下一代测序等）来识别和测量。由于多种可能性的存在，判断哪些生物标志物可以预测药物活性，以及如何在早期开发过程中确定生物标志物的阈值仍是一个挑战。为减少选择偏倚，应事先将研究受试者分成两组，并指定用于发现和确认生物标志物的训练集和验证集。每次调查新的生物标志物时，都需要重复这个产生和检验假设的过程。基于单臂试验发现的预测型生物标志物，不论其研究实施得如何严格，实际仍存在只是一种预后型生物标志物（可采用前瞻性流行病学研究以评估其预后效果）或者是仅能预测短期肿瘤反应的生物志记物（需要进行更长时间的随访）的可能。在随后的确证性试验设计中，必须考虑到上述不确定性。例如，当涉及到两个亚组人群之间的α分配时，若采取向下（step-down）法，则需要确定检验的层级性，而前期数据对此无法提供充分支持，此时最好能选取更恰当的方法进行α分配。此外，对于I类错误率控制的考虑，在涉及人群选择和试验扩展的统计设计中会更加复杂。在正式开展试验前，应衡量各种设计方案的利弊，并妥善解决监管部门所关注的问题。

4. ?????

在单一方案下同时检测多种试验药物和/或多个肿瘤适应症，且无需为每次试验制定新方案的试验设计，称为主方案设计。它包括篮式设计、伞式设计和平台设计。在有或无生物标志物富集的患者人群中同时研究一种试验药物在多个肿瘤适应症中的试验，被称为篮式试验。确证性篮式试验的主要研究人群通常包括具有独特分子标记的患者。考虑到无效的肿瘤队列可能会稀释整体治疗效果，因此肿瘤适应症的初步选择必须基于重要的科学和临床证据，以便为数据合并奠定坚实的基础，降低试验失败的风险。基于期中分析数据将疗效较差的肿瘤队列从最终的合并分析中去除可以进一步将风险降至最低，但可能会导致整体I类错误控制方面的问题，需要进行适当的多重性调整。去掉无效队列后，剩余肿瘤队列的样本量也将进行重新调整，以维持最终合并分析的统计效能。合并分析前还要考虑队列间的异质性。在这种情况下，样本量重新调整策略必须事先制定并与监管部门达成一致。如果I类错误能够得到适当控制，确证性篮式试验也可以考虑其它如贝叶斯等设计方法。无论采用哪种设计方法进行篮式试验，在合并分析中拒绝全局原假设并不意味着试验药物在所有参与合并分析的肿瘤适应症中同样有效，也不意味所有适应症均应获得批准。就基线特征对治疗效果的影响而言，与传统的III期试验相似，监管部门基于确证性篮式试验做出是否批准药物上市或同意说明书范围的决定将取决于额外分析的结果（例如合并分析中的治疗效果是否主要由某特定肿瘤适应症子集所决定，试验药物的获益-风险特征在单个肿瘤队列中是否有利）。另外也可能需要通过上市后研究进一步证实临床获益。作为篮式试验的补充，伞式试验可以在同一肿瘤适应症中同时研究多种试验药物。伞式试验中试验药物可以持续的加入或移出。当有多个试验组（或药物队列）开放入组时，应该采用随机化设计。随机化比率可以根据试验中新出现的数据进行调整，以倾向于更有前景的治疗组，并提前终止无效治疗组。由于试验药物的研究是在同一个平台上进行，且通常在某些特定的研究中心开展，因而不同药物队列之间患者群体的异质性可能较小，试验药物之间的比较也会比单独研究的结果更可信。随机对照伞式/平台试验可看作一种特殊类型的多臂III期试验，因此可遵循相同的原则进行多重性调整。如果试验的重点在于分别回答每一种治疗的疗效问题，而不是为了得到总体疗效的单个结论声明，则与单独的对照试验相比，采用共同对照的伞式/平台试验的总体I类错误率总是更低，原则上不需要进行多重性调整。但是，如果试验中包含了同一治疗的不同剂量组，则必需进行多重性调整以回答该治疗的疗效问题。若试验中同时采用了响应适应性随机化或其他适应性设计，多重性问题的控制将更为复杂。在随机对照伞式/

平台试验中，试验组和对照组之间的主要比较一般应基于同期参加试验被随机分配的受试者。

??????

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures[J]. Statistics in medicine, 2009, 28(4): 586-604.

2. Chapman P B, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation[J]. New England Journal of Medicine, 2011, 364(26): 2507-2516.

3. Chen C, Li X N, Li W, et al. Adaptive expansion of biomarker populations in phase 3 clinical trials[J]. Contemporary clinical trials, 2018, 71: 181-185.

4. Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients[J]. Clinical cancer research, 2005, 11(21): 7872-7878.

5. Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial[J]. Clinical trials, 2006, 3(1): 57-71.

6. Hobbs B P, Barata P C, Kanjanapan Y, et al. Seamless designs: current practice and considerations for early-phase drug development in oncology[J]. JNCI: Journal of the National Cancer Institute, 2019, 111(2): 118-128.

7. Howard D R, Brown J M, Todd S, et al. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group[J]. Statistical methods in medical research, 2018, 27(5): 1513-1530.

8. ICH. E9(R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2019.

9. Ji Y, Liu P, Li Y, et al. A modified toxicity probability interval method for dose-finding trials[J]. Clinical trials, 2010, 7(6): 653-663.

10. Kang S P, Gergich K, Lubiniecki G M, et al. Pembrolizumab KEYNOTE-001: an adaptive study leading to accelerated approval for two indications and a companion diagnostic[J]. Annals of oncology, 2017, 28(6): 1388-1398.

11. Mandrekar S J, Sargent D J. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges[J]. Journal of clinical oncology, 2009, 27(24): 4027.

12. Mayawala K, Tse A, Rubin E H, et al. Dose finding versus speed in seamless immune-oncology drug development[J]. The journal of clinical pharmacology, 2017, 57: S143-S145.

13. Proschan M A, Follmann D A. Multiple comparisons with control in a single experiment versus separate experiments: why do we feel differently?[J]. The american statistician, 1995, 49(2): 144-

149.

14. Schwartz L H, Litière S, de Vries E, et al. RECIST 1.1—update and clarification: from the RECIST committee[J]. European journal of cancer, 2016, 62: 132-137.

15. Seymour L, Bogaerts J, Perrone A, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics[J]. The lancet oncology, 2017, 18(3): e143-e152.

16. Sun L Z, Kang S P, Chen C. Testing monotherapy and combination therapy in one trial with biomarker consideration[J]. Contemporary clinical trials, 2019, 82: 53-59.

17. Yuan Y, Hess K R, Hilsenbeck S G, et al. Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials[J]. Clinical cancer research, 2016, 22(17): 4291-4301.

18. Zhou H, Yuan Y, Nie L. Accuracy, safety, and reliability of novel phase I trial design[J]. Clinical cancer research, 2018, 24(18): 4357-4364

??????????

中文	英文
盲态独立中心审查委员会	Blinded Independent Central Review, BICR
部分缓解	Partial Response, PR
患者报告结局	Patient Reported Outcome, PRO
机构审查委员会	Institutional Review Board, IRB
疾病控制率	Disease Control Rate, DCR
客观缓解率	Objective Response Rate, ORR
首次人体	First in Human, FIH
推荐 II期剂量	Recommended Phase 2 Dose, RP2D
完全缓解	Complete Response, CR
无病生存期	Disease Free Survival, DFS
无进展生存期	Progression Free Survival, PFS
无事件生存期	Event Free Survival, EFS
至疾病进展时间	Time to Progression, TTP
至治疗失败的时间	Time to Failure, TTF
总生存期	Overall Survival, OS
最佳支持治疗	Best Supportive Care, BSC

Revision #4
Created 2026-01-21 06:29:55 UTC by Colin Wang
Updated 2026-01-21 06:38:33 UTC by Colin Wang