

抗肿瘤药物临床试验终点技术指导原则

二〇一二年五月

抗肿瘤药物临床试验终点技术指导原则

一、概述

临床试验终点 (End Point) 服务于不同的研究目的。在传统的肿瘤药物的研发中, 早期的临床试验目的是评价安全性以及药物的生物活性, 如肿瘤缩小。后期的有效性研究通常评价药物是否能提供临床获益, 例如生存期延长或症状改善等。

用于支持药物批准的临床试验终点通常应当是反映临床获益的指标。在肿瘤领域, 生存期改善被认为是评估某种药物临床获益的合理标准。在 20 世纪 70 年代, 通常以影像检查或体检等肿瘤评估方法测得的客观缓解率 (Objective Response Rate ,ORR) 为依据批准抗肿瘤药物上市。在随后的数十年里, 逐渐认识到抗肿瘤药物的审批应该基于更直接的临床获益证据, 如生存期改善、患者生活质量提高、体力状况或肿瘤相关症状减轻等。这些临床获益很多时候并不能通过客观缓解率或与其相关的指标进行预测。

当某种药物用于治疗严重或威胁生命的疾病、对现有治疗有明显改进、或填补治疗空白时, 在一定条件下可采用替代终点 (Surrogate End Point) 支持该药物的上市申请。这些替代终点可能不像血压或血清胆固醇这类经过充分验证的指标, 但可能合理预测临床获益。此种情况下, 申请人必须承诺进行上市后临床试验以确证该药物的实际临床获益。如果上市后研究不能证明该药的临床获益, 或者申请人未按要求进行承诺的上市后研究, 则国家食品药品监督管理局 (以下简称 SFDA) 可将该药物从市场中撤出。

本指导原则的目的是为申请人开展抗肿瘤药物临床试验终点指标的选择提供参考，以使其符合某种药物上市申请的有效性评价要求。本指导原则主要适用于国内、外均未上市的抗肿瘤新化合物的临床试验研究，新生物制品也可参考部分内容。本指导原则中仅讨论用于治疗肿瘤患者的药物的终点,未讨论用于预防或降低肿瘤发生率的药物的终点。

二、关于临床试验终点的一般性考虑

本节回顾了抗肿瘤药物研发中的一般性问题。对常用的抗肿瘤药物临床试验终点进行了探讨，并对采用了这些终点的肿瘤临床试验设计中的相关问题进行了讨论。本节中将讨论的临床试验终点包括总生存期 (Overall Survival ,OS)、基于肿瘤测量的终点如无病生存期 (Disease - Free Survival ,DFS)、 ORR、完全缓解(Complete Response, CR)、疾病进展时间 (Time to Progression ,TTP)、无进展生存期 (Progression-Free Survival ,PFS)和基于症状评价的终点。抗肿瘤药物审批所用的重要临床试验终点比较见下表。

表：抗肿瘤药物审批所用重要临床试验终点的比较

终点	研究设计	优点	缺点
OS	<ul style="list-style-type: none"> 需随机研究 盲法不是必须的 	<ul style="list-style-type: none"> 广为接受的临床获益直接衡量方法 易于测量 可精确测量 	<ul style="list-style-type: none"> 可能需要大型研究 易受交叉治疗和后续治疗的影响 包括非癌症死亡
症状终点	<ul style="list-style-type: none"> 随机盲法研究 	<ul style="list-style-type: none"> 患者临床获益的直接感受 	<p>盲法通常难以进行</p> <p>数据缺失和不完整情况较普遍</p> <p>小变化的临床意义不清楚</p> <p>多元分析</p> <p>缺乏经过验证的测量工具</p>

终点	研究设计	优点	缺点
DFS	<ul style="list-style-type: none"> • 需随机研究 • 首选盲法研究 • 推荐进行盲态审查 	<ul style="list-style-type: none"> • 与生存研究相比所需病例少并且所需的随访时间短 	<ul style="list-style-type: none"> • 并非所有情况下在统计学上都是有效的生存期替代指标 • 非精确测量，存在评价偏倚，特别是在开放性研究中 • 不同研究存在不同定义
ORR	<ul style="list-style-type: none"> • 可用单臂或随机研究 • 比较性研究中首选盲法 • 推荐进行盲态审查 	<ul style="list-style-type: none"> • 可在单臂研究中评价 • 与生存研究相比，可较早并且在研究规模较小的研究中评价 • 有效性归因于药物，而非疾病的自然进程 	<ul style="list-style-type: none"> • 不是临床获益的直接测量 • 不是对药物活性的综合测量 • 受益仅局限于患者亚组
CR	<ul style="list-style-type: none"> • 可用于单臂或随机研究 • 比较性研究中首选盲法 • 推荐进行盲态审查 	<ul style="list-style-type: none"> • 可在单臂研究中评价 • 持续完全缓解可表明临床获益 • 与生存研究相比，可较早并且在研究规模较小的研究中评价 	<ul style="list-style-type: none"> • 并非全部病例获益的直接测量 • 不是对药物活性的综合测量 • 受益仅局限于患者亚组
PFS (包括全部死亡病例), 或 TTP (进展之前发生死亡病例被“删失 (censored)”))	<ul style="list-style-type: none"> • 随机研究 • 首选盲法 • 推荐进行盲态审查 	<ul style="list-style-type: none"> • 与生存研究相比所需病例少并且所需的随访时间短 • 包括对稳定疾病的测定 • 不受交叉治疗和后续治疗的影响 • 通常基于客观、定量评估 	<ul style="list-style-type: none"> • 不是所有情况下在统计学上都是有效的生存替代指标 • 非精确测量，受试者的评价存在偏倚，特别是在开放性研究中 • 在不同研究中存在不同定义 • 需频繁进行影像学和其他评估 • 包括各治疗组之间评估的时间平衡。

(一) 总生存期

总生存期定义为从随机化开始到因各种原因导致病人死亡之间的时间，且是按意向治疗人群(ITT)计算。这个终点精确可测，并有死亡日期提供依据。在终点评估时不会出现偏倚。生存期是迄今为止评价抗肿瘤药物最可靠的临床试验终点，当研究能充分评价生存期时，它通常是首选终点。

总生存期应在随机对照研究中评价。对于这类时间依赖性终点(例如 OS、PFS)的历史研究中的数据极少可信。历史研究中对照组和目前治疗组间除使用药物治疗不同外，其他因素的差异还包括病例的选择、影像技术或支持治疗的改善，都将导致结果出现显著差别。随机化研究通过进行直接结果的比较，可将这些差别最小化。如果药物的毒性可以接受，总生存期显著的改善可视为有临床意义，通常能支持新药的上市批准。

生存期研究实施和分析中存在的困难包括大型试验随访期较长，以及随后的抗肿瘤治疗可能会混淆生存期的分析。

(二) 基于肿瘤测量的临床试验终点

在本节，将讨论几种基于肿瘤测量的临床试验终点。这些终点包括无病生存期(DFS)、客观缓解率(ORR)、疾病进展时间(TTP)、无进展生存期(PFS)和治疗失败时间(TTF)。所有时间依赖性终点的数据收集和处理均基于间接的评价、计算或估算(如肿瘤的测量)。关于无进展生存期数据收集和讨论的讨论见附 3。

当选择基于肿瘤测量的临床试验终点时，应针对该终点在抗肿瘤药物临床获益评价中的不确定性和偏倚进行评估。在不同的肿瘤试验中，研究者对肿瘤测量的精确性相差甚远。此外，如果肿瘤没有明确边界(例

如恶性间皮瘤、胰腺癌和脑瘤)，使用肿瘤测量方法获得的缓解率可能是不精确的。药物上市申请时如果采用基于肿瘤测量的临床试验终点作为有效性的唯一证据，那么通常应提供来自第二个试验得到的确凿证据。

如果主要研究终点是选择基于肿瘤测量的终点指标（如 PFS 或 ORR），通常应由对研究治疗分配处于盲态的独立终点审查委员会验证该终点指标的评价（见附 4）。如果试验本身未设盲时，这种由独立的第三方进行的盲态下的测量特别重要。SFDA 可能会抽查部分资料进行监察以核实独立终点审查委员会的审阅程序。其他关于资料收集的细节见附件 1。当随机研究采用盲法（除非发生不良事件使得研究实际上已被揭盲）或者在大型随机研究中效应值是稳健的，并且敏感性分析证明未发现观察者的偏倚（尤其是针对 DFS），基于肿瘤测量的终点指标评价（尤其是 PFS 或 DFS）可不需要进行集中独立审核。

1.无病生存期

无病生存期(DFS)通常定义为患者从随机分组开始到出现肿瘤复发或由任何原因引起死亡之间的时间。该终点最常用于根治性手术或放疗后的辅助治疗的研究。如果某些疾病（例如血液肿瘤）在大部分患者化疗后达到完全缓解时，DFS 也可以作为一个重要终点。尽管在大多数辅助治疗的情况下，总生存期仍然是一个传统的终点指标，但是当生存期延长而使得选择生存期为临床试验终点不现实的时候，DFS 可以作为一个重要的终点指标。目前 DFS 已经成为乳腺癌辅助性激素治疗、结肠癌辅助治疗、以及乳腺癌的辅助性细胞毒治疗的主要审批基础。无疾病生存期可以是临床获益的替代终点或者可以为临床获益提供直接证据。这一决定取决于疗效大小、风险-效益关系以及疾病情况。

如果将 DFS 作为一个可能终点, 需要考虑的重要问题包括估计的疗效大小和已经证明的标准治疗方法的获益。试验方案应详细说明 DFS 的定义以及随访研究和随访的时间安排。许多原因可能导致计划外评价, 各试验组之间计划外评价的频率、时间和原因存在差异, 而这些差异可能会导致偏倚。如果可行, 可通过研究者和患者的双盲操作, 使这种潜在的偏倚最小化。可以通过对随访期间事件发生的总数 (不管事件发生的时间) 进行比较分析, 从而评估由于计划外的评价而产生的可能偏倚影响。

无病生存期的定义可能比较复杂, 尤其是当死亡发生而没有预先对肿瘤进展情况进行记录时。这些事件可记录为疾病复发, 或作为删失的事件。尽管所有关于死亡的统计分析方法均有一定的局限性, 但将所有原因导致的死亡均认为是由疾病复发所致可以将偏倚降至最低。这种定义的限制性在于高估了 DFS, 尤其是对于长期失访后死亡的患者。如果在各研究组中长期随访的频率不一致或是因为药物毒性而产生的非随机脱落, 将会产生偏倚。某些分析将肿瘤相关死亡认为是 DFS 事件, 并删失了非癌症死亡病例, 这种方法可能在判断死亡原因时产生偏倚。另外, 任何删失患者的方法 (无论是针对死亡还是末次随访) 都会认为被删失的患者与未被删失的患者具有相同的复发风险。

2. 疾病进展时间和无进展生存期

疾病进展时间 (TTP) 和无进展生存期 (PFS), TTP 定义为从随机分组开始至出现肿瘤客观进展之间的时间; TTP 不包括死亡。PFS 定义为从随机分组开始至出现肿瘤客观进展或死亡之间的时间。关于肿瘤进展的明确定义非常重要, 应在方案中进行详细描述。

2.1 TTP和PFS

与 TTP 相比，PFS 是更常选用的替代终点。因为 PFS 包括死亡，更好地反映了受试药物的毒副作用，因此与总生存期有更好的相关性。在对 TTP 的分析中，无论是在死亡时还是相对较早的随访期间，死亡均被删失（试验中的非随机脱落）。PFS 可假设患者的死亡与肿瘤进展有着随机的联系。然而，当大多数死亡与癌症不相关的情况下，TTP 也可以是一个合适的终点指标。

2.2 PFS作为支持药物审批的终点

上表列出了用 PFS 作为支持抗癌药物审批终点的优点和缺点。PFS 反映了肿瘤的生长，又可以在证实生存期获益之前进行评价。不会受到后续治疗的混淆。对于预定的样本量，PFS 受到的影响可大于总生存期受到的影响。然而对于许多不同种类的恶性肿瘤来说，正式确认 PFS 作为生存期的替代终点是比较困难的。通常没有足够的数据对生存期和 PFS 的相关性进行评价。抗肿瘤药物临床试验规模通常较小，已有药物的生存获益通常比较小。在不同的抗肿瘤治疗试验中，用于支持审批的 PFS 终点所起的作用是不同的。无进展生存期的延长是否直接代表临床获益或仅是临床获益的替代终点，取决于这种新治疗方法的疗效大小以及与现有治疗相比的风险-效益比。

2.3 PFS试验设计问题

在试验方案和统计分析计划（SAP）中应详细描述评价、测量和分析PFS的方法学。同样，在试验方案中仔细对肿瘤进展标准进行定义也很重要。现在还没有法定标准对肿瘤进展作出定义，申请人可能使用不同的标准，包括RECIST标准。除了公认的PFS标准中列出的大纲，试验

方案和统计分析计划应增加其他细节。在两个试验组中，随访和影像学评价必须是均衡的，以避免系统偏倚。研究应尽量采用盲法。如果将患者或研究者评价作为进展终点的一个因素，则设盲显得尤其重要。至少，应由处于盲态的独立裁定小组（independent adjudication team，一般包括影像学家和临床医师）进行评价。SFDA和申请人应在以下方面提前达成一致意见：

- 研究设计；
- 疾病进展的定义；
- 记录在CRF表上的数据；
- 统计分析计划（SAP）；
- 缺失数据的处理办法和数据删失办法；
- 如适用，独立终点审核委员会（IRC）的操作规程（见附件4）。

2.4 PFS的分析

由于有数据缺失等现象，PFS的分析变得比较困难。试验方案中应针对每位患者对“一个充分的评价随访（即，在此次随访中完成了所有约定的肿瘤评价）”给予定义。分析计划应总结比较各治疗组的随访充分性。方案应详细说明如何分析不完整和/或缺失的随访数据，以及数据删失的方法。分析计划应明确说明主要分析，以及一个或多个敏感性分析，以评价结果的可靠性。虽然所有含缺失数据的分析都存在一定问题，但只要敏感性分析和主要分析都支持结论，那么结果就是正确的。评价中应该包括长期失访患者的死亡人数。这类死亡会造成高估了随访较少的组的PFS，从而导致PFS的测定产生偏倚。

由于可以从多个时间、多种来源获得病情进展数据（包含计划外随

访的身体检查和各种类型的影像学扫描)，因此每次评价随访进行的数据收集有必要在随访前后限定的较短时间内进行。在较长时间内收集数据时，难以确定事件日期和删失日期。建议：如果之前不存在缺失的评价，将最早观察到病情进展的时间标记为进展日期，并将确定无进展的末次影像学评价日期确定为截止日期。附件3提供了一套主要分析或敏感性分析中可用于PFS的表格，建议在II期临床试验结束后与SFDA讨论PFS数据的收集和分析计划，在专门的方案评估中进行核实。

3.客观缓解率

客观缓解率(ORR)是指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。缓解期通常是指从开始出现疗效直至证实出现肿瘤进展的这段时间。一般定义客观缓解率为完全缓解加上部分缓解之和。客观缓解率是一种直接衡量药物抗肿瘤活性的指标，可以在单臂试验中进行评价。疾病稳定不应该是客观缓解率的组成部分。疾病稳定可以反映疾病的自然进程，而肿瘤缩小则是直接疗效。同样，疾病稳定可通过TTP或PFS分析进行更精确的评价。如果可能，应采用标准化的指标以确定疗效，如RECIST标准。缓解标准应在试验开始前的方案中提前定义。客观缓解率的评估包括缓解程度、缓解持续时间以及完全缓解率(没有可测量到的肿瘤)。

4.治疗失败时间

治疗失败时间(TTF)是一个复合的终点指标，即从随机化开始到无论何种原因(包括疾病进展、治疗毒性和死亡)导致治疗终止之间的时间。一个合理的支持审批的终点指标应当能清楚地将有效性和药物毒性、患者或医师退出、或患者不耐受区分开，TTF不能将有效性与其他

变量进行充分区分。因此，不建议将TTF作为支持药物批准的终点。

（三）基于症状评估的临床试验终点

症状和体征的改善通常被认为是临床受益，如体重的增加、疼痛的减轻或止痛药用量减少等。主要可用于盲法、多数患者有症状、无有效治疗药物和较少做影像评估的试验的疗效评价指标。在非盲法试验中则容易受到主观因素的影响，导致结果偏倚。当以症状和体征的改善作为支持抗肿瘤药物审批的主要终点时，应当能够区分是肿瘤相关症状的改善还是药物毒性的减小或缺失。

患者自评结果（Patient Reported Outcome ,PRO）是直接来自患者的关于其健康状况的报告，而非来自临床医生或其他任何人，可作为反映症状获益的恰当评价方法。但有一定局限性，研究者和受试患者报告中可能存在很大差别，问卷信息收集的时间点也会有影响，语言因素也会导致不能准确评估。生活质量评分（Quality of Life, QOL）也可以用来评估与健康相关的生活质量。但应当注意，以 QOL 来衡量药物的结果可能只能说明某种药物相对其他药物来说毒性较小，但并非其有效性更好。

设计适宜的详细的评估量表是准确评估药物作用的基础，不应仅仅提供“出现或未出现”这样的数据。用于抗肿瘤药物临床试验效果评价的量表必须经信度（reliability）和效度（validity）分析，且是学术界认可的。量表中各项目的评价应尽可能采用定量或等级来反映观察项目变化的程度，应尽可能避免采用“是或否”、“出现或未出现”这样的二分类数据。

由于毒性或肿瘤进展而终止对受试者的评价，导致数据丢失，这种

情况在抗肿瘤药物临床试验中较为常见。数据的丢失会导致评价的困难。因此，在试验方案中应有相应的措施尽可能避免或减少数据的缺失。

1.特定症状终点

与“疾病进展时间 (TTP)”相似，“癌症症状进展时间 (Time to progression of cancer symptoms)”是临床获益的直接测量指标而不是替代终点。如上文所述，疾病进展测量中存在的问题（例如评估缺失）同样存在于“症状进展时间”的评估中。在非盲的试验中可能导致评价偏倚。其次，肿瘤的发展和肿瘤相关症状开始之间会出现间隔。往往在达到症状终点前，治疗就进行了调整，从而混淆了分析结果。此外，许多抗肿瘤药物的试验中采用的是肿瘤相关症状很轻的患者。而且很难将肿瘤相关症状和药物毒性区分开。

复合症状终点的中的某一症状指标应当具有相似的临床重要性，并且其结果不应当仅仅归因于一个指标。例如，基于复合终点批准用于治疗肿瘤骨转移患者的药物，与骨骼相关的事件被定义为病理性骨折、骨放射治疗、骨外科手术和脊髓压迫。

选择恰当的研究人群是证明症状方面获益的关键。对于在研究基线时即有症状的患者，可利用分类的症状缓解分析评价。对于在基线时无症状的患者，应分析“首次发生症状的时间”。即使患者终止使用研究药物或开始使用一种新药，如果随访持续到首次出现症状，仍可评价症状的进展。

2.症状数据面临的问题

数据缺失以及评价不充分可能使症状数据的评价更复杂，特别是对于开放性研究而言。因药物毒性或肿瘤进展而退出研究是症状数据丢失的一个原因。理想的情况是，当患者停止治疗时应该继续收集可供分析

的信息。基于多样性考虑，应进行多种症状的前瞻性数据收集，并且需要在SAP中详细说明必要的统计学修正。

（四）生物标志物（Biomarker）

尽管目前许多生物标志物已经作为临床观察肿瘤反应和进展的监测指标，比如CA-125用于卵巢癌，PSA用于前列腺癌的观察，血液和尿液中异常蛋白水平用于骨髓瘤缓解评价。但尚需要做进一步的研究证实现有测试方法的可靠性，并确定生物标志物改善是否能预测临床获益。因此，目前生物标志物不能单独作为上市批准的依据，SFDA可以接受肿瘤标志物作为复合终点的一个指标。例如在卵巢癌患者中，伴随CA-125上升的某些特定临床事件（如体力状况明显下降或肠梗阻）可反映患者病情进展。此外，生物标志物还用于确定预后因素、患者选择，以及在试验设计中需要考虑的分层因素。提倡保留肿瘤部分组织、体液或血清样本等以进行相关标志物的研究检查。

三、临床试验设计考虑

新药上市必须基于“充分且良好对照的研究”的有效性证据支持。研究必须与一个对照组进行比较，必须就药物的疗效提供足够充分的评价。用于证明有效性的最可靠的方法是，在盲法随机对照试验中，显示出有统计学意义和临床意义终点指标的改善。下面我们将讨论几个关于支持药物上市的抗肿瘤药物临床试验设计的问题。

（一）单臂试验

在当前没有其他治疗方法的情况下，并且据推测肿瘤的明显缩小可归因于试验药物，SFDA有时会接受在单臂研究中所观察到的客观缓解率和缓解持续时间作为支持审批的证据。例如在急性白血病等疾病中，

可以使用缓解率作为支持审批的终点指标，因为在这些疾病中，完全缓解与输血需求量减少、感染降低和生存期增加相关。因为各种类型肿瘤的自然进程变异性很大，单臂试验不能充分体现时间-事件终点，如生存期、TTP和PFS。因此如采用时间-事件终点指标时,需要进行随机对照研究。

（二）非劣效性研究

非劣效性 (Non-inferiority) 试验的目的是通过一个预先规定值 (非劣效性界值) 论证一个新药的疗效不低于标准治疗药物。非劣效性界值是指疗效的降低必须在临床可接受的范围内，且不得超出阳性对照药的效果。标准治疗药物必须有明确的临床获益 (生存获益)。如果一个新药的疗效低于阳性对照药疗效的程度超出了非劣效界值，则可推断该新药是无效的。

非劣效性试验是基于外部 (历史性) 数据确定阳性对照药物的疗效大小。在抗肿瘤药物试验中，通常没有充分数据描述该疗效。非劣效性试验还依赖于一个恒定的假设。该假设包括：历史研究和当前研究中阳性对照药的疗效是恒定的。阳性对照药的有效性来源于历史数据和当前试验，这一假设包括两者均具有同样的患者人群特点、支持治疗方法、评价方法。阳性对照药疗效的估计值应基于对历史研究进行全面的荟萃分析。这些研究应重新证实与安慰剂相比阳性对照药的有效性。实施非劣效性试验的困难包括如何评估阳性对照药疗效大小以及确定需要保留的效应值 (非劣效界值)。与优效性试验相比，非劣效性试验通常需要更多的病例，并且涉及到临床试验结果的重现性。此外，后续治疗和交叉到阳性对照药组可以干扰各种非劣效性分析。使用生存期之外的其他终点的非劣效性试验，对结果的解释是存有疑问的。

（三）放疗保护剂和化疗保护剂的试验设计

放疗保护剂和化疗保护剂是专门用于减轻放疗或化疗毒性的药物。这些药物的试验评价通常有两个目标。一是评价保护剂是否达到预期减轻放疗或化疗毒性的目的。二是确定保护剂是否危及抗肿瘤疗效。可采用替代终点检验第二个目的，如客观缓解率(ORR)或者至肿瘤进展时间(TTP)，而非总生存期。

四、结语

本指导原则的总体原则是帮助申请人选择上市申请的终点，同时鼓励申请人在申报支持新药上市方案的之前和SFDA进行讨论，以探讨用于支持药物上市申请的终点和方案设计的可行性。

最后，上市审批并不仅仅取决于临床试验设计，还取决于药物上市申请中所有研究结果和数据。

附 1:

肿瘤测量数据采集

以下是有关肿瘤测量数据的重要注意事项。

- CRF 表和电子数据记录治疗前基线随访时鉴定的靶病灶。这类病灶的回顾性鉴定是不可靠的。
- 肿瘤病灶可用指定的字母或数字标识。这样的标识可区分同一解剖位置发生的多个肿瘤，并可对基线和随访期观测的肿瘤作配对比较。
- 要有一套机制确保随访期间关键时间点的全部数据收集。CRF 表应保证每次随访时对受试者所有靶病灶进行评估，并且在基线和随访期间采用同样的影像学或测量方法完成所要求的全部检测。
- CRF 包括的数据应显示每次随访中是否进行了影像学扫描检查。
- 当病灶完全治愈时，应标记为“0”。否则，病灶消失不能与缺失值区别。
- 随访期检查应对在病灶的原发部位以及新发部位出现的新发病变进行及时检测。新发病变的出现和部位应记录在 CRF 表中，并以电子数据提交。

附 2:

PFS 分析中需考虑的问题

研究方案和统计分析计划应能详细地体现无进展生存期 (PFS) 的主要分析。其中包括终点的详细描述, 评价肿瘤的方法, 评价肿瘤进展情况时使偏倚最小化的方法, 例如建立独立终点评审委员会 (independent endpoint review committee, IRC)。通过一两次的次要分析评价临床实验中预期的问题以及评价结果是否可靠。以下列举一些需要考虑的重要因素。

- **进展日期的定义。**生存分析使用确切的死亡日期进行分析。但是在 PFS 分析中, 确切的进展日期往往不可知。以下是在 PFS 分析中用来定义已记录进展日期 (PDate) 的两种方法。

1. 一个方法是将能够证明进展的第一时间定为 PDate:

- 以新病灶的出现作为进展的标准, PDate 指的是首次观察到可测定的新病灶的日期。
- 如果在不同的时间进行多次基于靶病灶测量值总和的评价, PDate 是指对靶病灶进行最后一次观察或影像学检查的日期, 可以显示出靶病灶测量值总和的增加达到预定的阈值。

2. 另一个方法是将 PDate 定义为在所有影像学检查完成后 (综合证明存在进展), 即按方案计划进行的临床随访的日期。

- **删失/截尾日期的定义。**删失日期的定义适用于在数据截止或中途退出之前没有进展记录的患者。在这些患者中, 删失日期定义为有足

够证据证明进展情况的最终日期。一个可行的方法是采用最后评价的日期。也可以用多次影像学检查来确定进展。另一个可行的方法是采用与影像学检查相应的临床随访日期。

- **充分的 PFS 评价的定义。**对于无证据证明病情进展的病人，PFS 的删失通常取决于最后对肿瘤进行充分评价的日期。充分的肿瘤评价的精确定义包括充分的靶病灶的评估及用于评价非靶病灶和寻找新病灶的充分的影像学检查。
- **部分缺失肿瘤数据的分析。**当在充分肿瘤评价随访期间出现数据部分缺失时，分析计划应当描述计算进展状态的方法。
- **肿瘤数据完全缺失。**有时在出现数据缺失的随访之后紧随着的是随访发现死亡或发现进展。另一种情况是后续评价显示没有进展。对于后一种情况，继续治疗和继续监测进展证据看起来是妥当的，但是处理缺失数据的方法因后续事件而不同，并能呈现信息的删失。另一个可行的方法是包括来自后续 PFS 评价之后的数据。当评价频率较高并且仅有一次随访缺失时是可行的，但是当有两个或更多的随访缺失时，应在最后一次充分肿瘤评估后删失可能更合理。为评价缺失数据的潜在影响，统计分析计划 (SAP) 详述主要和次要的 PFS 分析至关重要。在确定删失和进展状态的过程中应综合考虑脱落原因。例如就主要分析而言，因为未被证实的临床进展、抗肿瘤治疗的改变或体力状况的恶化而退出研究的患者应在最后充分的肿瘤评价时删失。次要的敏感性分析会将这些脱落视为进展事件。尽管进展后的随访可能很困难，但应尽全力保证能对发生疾病进展的患者进行随访，而不必在意缺失的随访次数。
- **无法测量疾病的进展。**如果可能，进展标准应对每一个评价形态特

征的方法进行描述（如 CT 扫描、骨扫描）。通过一个处于盲态的复审委员会，核查无法测量而影像学扫描证实进展的疾病，并经过 SFDA 的核实是非常重要的。

- **可疑病灶。** 在分析期间，对于指定的进展状态，应对不明确的病灶提供评价和随访的计算方法。

附 3:

PFS 分析示范表

正如第二部分（二）2.4 所讨论的，敏感性分析有助于确定 PFS 分析是否可靠。但是，敏感性分析是探索性的，支持主要分析的结果，并且有效性不可能仅仅基于敏感性分析来确定。下表描述了不同的敏感性分析方法以说明如何指定进展日期以及进展数据的删失日期。下列 3 个表分别描述了 3 种不同的敏感性分析方法：

a.表 A 给出了对于仅仅包括有充分记录的，并能够被证实的进展事件的敏感分析。其他数据被删失。表 A 进展日期为：

- 仅基于由独立终点审查委员会（IRC）核实的放射性评价。临床进展不能作为进展终点。
- 第一次记录肿瘤进展的时间。
- 密切随访患者的死亡时间。但是，遗漏两次或两次以上随访后发生的死亡，在最后一次随访时被删失。

表 A

PFS1 (仅包括有记录的进展)

情况	进展或删失日期	结果
无基线肿瘤评估	随机化	删失
预定随访的过程中证实有进展	最初的发现： <ul style="list-style-type: none"> 影像学检查发现新的病灶（如果进展的标准是新的病灶）日期；或 多个可测量病灶的末次的影像学检查日期（如果进展的标准是可测量的多个病灶面积之和） 	进展
无进展	可测量病灶的末次影像学检查日期	删失
无证据表明进展而停止治疗	可测量病灶的末次影像学检查日期	删失
因毒性或其他原因停止治疗	可测量病灶的末次影像学检查日期	删失
开始采用新的抗癌治疗	可测量病灶的末次影像学检查日期	删失
首次疾病进展评估之前死亡	死亡日期	进展
充分的评估过程中死亡	死亡日期	进展
遗漏一次以上随访后死亡或进展	可测量病灶的末次影像学检查日期	删失

表 B 中的敏感性分析通过指定删失日期以及预定随访日期的事件以纠正肿瘤评价随访进度中的潜在偏倚。但是，如果在邻近的最后一次随访时发生进展，这一方法可能引入偏倚，特别是对于开放试验而言。这一方法最好应用于随机盲法的研究中。

表 B

PFS2 (进展和评估日期一致)

情况	进展或删失日期	结果
无基线肿瘤评估	随机化	删失
预定随访的过程中证实有进展	下次预定随访日期	进展
无进展	有适当评估的末次随访日期	删失
无证据表明进展而停止治疗	有适当评估的末次随访日期	删失
因毒性或其他原因停止治疗	有适当评估的末次随访日期	删失
开始采用新的抗癌治疗	有适当评估的末次随访日期	删失
首次疾病进展评估之前死亡	死亡日期	进展
充分的评估过程中死亡	死亡日期	进展
遗漏一次以上随访后死亡或进展	有适当评估的末次随访日期	删失

b. 表C中的敏感性分析则根据研究者的评价来评价PFS。但是，如果在邻近的最后一次随访时发生进展，这一方法可能引入偏倚，特别是对于开放试验而言。这一方法最好应用于随机盲法的研究中。

表C

PES3 (包含研究者要求)

情况	进展或删除日期	结果
无基线肿瘤评估	随机化	删失
预定随访的过程中证实有进展	下一次预定随访日期	进展
无进展	有适当评估的末次随访日期	删失
研究者声称有临床进展	预定随访（如在随访之间则为下次预定随访）	进展
因毒性或其他原因中止	有适当评估的末次随访日期	删失
新的抗癌治疗开始并且无临床进展	有适当评估的末次随访日期	删失
首次疾病进展评估之前死亡	死亡日期	进展
充分的评估随访或者患者缺失一次 评估随访后死亡	死亡日期	进展
长期失访（二次或以上错过评估） 后死亡	有适当评估的末次随访日期	删失

附 4:

肿瘤终点的独立审核

申请人和 SFDA 应能够核实支持药品审批的临床试验结果。在单臂研究中确定的客观缓解率可以通过对有限数量的影像学检查予以核实。但是，当药物批准是基于无进展生存期的测量时，需要缜密的计划以减小偏倚，并允许申请人和 SFDA 核实结果。独立终点审查委员会 (IRC) 在进行影像学判定解释和独立评价时，应将偏倚降低到最小。在启动研究之前，应拟定一个清晰描述 IRC 职责和进行过程的书面计划大纲（独立评价的章程）。计划应规定如何保证委员会的独立性，如何收集、贮藏和运送结果。还应包括如何消除不同解释的差异，以及临床数据的最终解释和监察程序。