

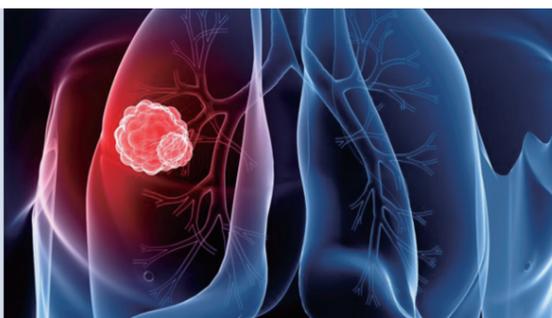
非小细胞肺癌试验终点 30 年变迁

OS 作为主要终点的试验减少, 而 PFS 作为主要终点的试验增多

▲ 本报记者 李玉梅

总生存 (OS) 与无进展生存 (PFS) 是评估非小细胞肺癌 (NSCLC) 治疗疗效的两个重要指标。OS 是临床试验终点的金标准, 而 PFS 是一个替代终点指标。PFS 能否替代 OS 一直是饱受争议的话题。

近期, 美国《临床肿瘤学杂志》发表的一项回顾性分析显示, 过去 30 年间, 晚期 NSCLC 的 III 期临床试验的设计和解读发生明显改变; 将 OS 作为主要终点指标试验比例正在下降。



专家观点

无进展生存作为试验主要终点优缺点兼在

▲ 北京大学肿瘤医院 张力建

多伦多大学 Natasha B. Leigh 等报告的回顾性分析显示, 过去 30 年间, 晚期 NSCLC 临床试验的设计和解读发生了显著改变。

尽管总生存 (OS) 仍是晚期 NSCLC 新药 III 期试验最常用的主要终点指标, 但是过去 10 年间使用无进展生存 (PFS) 作为主要终点的试验逐渐增多。

NSCLC 临床试验终点的选择发生改变, 而 PFS 能否替代 OS? 应基于 PFS 作为试验终点的优缺点进行审度。

PFS 作为试验终点优点

☆ PFS 反应患者肿瘤的增长情况, 而肿瘤增长在很大程度上是造成肿瘤相关死亡的原因, 因此 PFS 具有作为替代终点的良好特性。

☆ 评估 PFS, 无需非要等到出现死亡事件, 因此可使试验随访时间会大大缩短。

☆ 与 OS 相比, PFS 的获益程度 (magnitude) 会更大, 意思是同样的两组治疗, 在两组间 PFS 的差异往往会比两组间 OS 差异更大, 也就是说可通过更少的事件数来检验出差异常, 这也就意味着病例数大大减少。

☆ PFS 较少受非研究药物治疗的影响。因为患者往往是在药物治疗无效即疾病出现进展



张力建 教授

后才开始考虑换药, 因此对于 PFS 的评估往往是在患者换药之前, 这就使得 PFS 不像 OS 那样容易受非研究药物治疗的影响。

PFS 作为试验终点缺点

☆ PFS 的确认对于许多种类的肿瘤比较困难。

☆ 对 PFS 的评估不如 OS 评估精确, 因为 PFS 评估主要依赖于一些放射学检查, 某个患者在某次放射学检查时出现进展, 那么其进展日期就记录为放射学检查的这一天, 其实很可能患者在这次放射学检查前早就出现进展了, 只是没有进行放射学检查确定而已。因此, 在时间精度方面肯定不如死亡时间那么精确, 特别是当放射学检查时间间隔的比较长时。

☆ 存在主观偏倚, 因为进展与否最终还是由研究者根据放射学检查主观判定的。

☆ 需要频繁的放射学检查。



回顾性分析数据

OS 作为主要终点试验比例逐渐下降

多伦多大学 Natasha B. Leigh 等对 200 篇 NSCLC 临床试验相关文献进行回顾性分析, 发现把 OS 作为主要终点的试验明显减少。(J Clin Oncol. 2014 年 3 月 3 日在线版)

研究者旨在分析过

去 30 年晚期非小细胞肺癌 III 期试验设计和解读的变化。研究者最初筛选了 1980-2010 年间发表的 248 项试验, 排除不符合晚期 NSCLC 全身治疗标准的试验后纳入 203 项独立研究: 32 项发表于 20 世纪 80 年代,

53 项发表于 20 世纪 90 年代, 118 项发表于最近 10 年。

结果显示, 过去 30 年间试验终点的选择发生明显变化, 在 2000 年以前几乎所有 III 期 NSCLC 临床试验都将 OS 作为主要终点; 但是 2001-2010 年, 19% 的 III 期试验开始使用 OS 之外其他指标作为主要终点 (与之前的数十年相比, $P < 0.001$), 多数使用 PFS (13%)。

试验样本量随时间推移明显增加, 从首个 10 年的中位 152 例增至 2001-2010 年的中位 413 例 ($P < 0.001$)。

终点指标具有统计学意义改善的试验比例随着时间推移仍比较稳定, 然而报告阳性结果的试验 (主要终点指标

未显著改善) 比例明显增加, 从首个 10 年的 30% 增至 2001-2010 年的 53% ($P < 0.001$)。

对报告终点指标具统计学意义改善的 60 项试验分析显示, 患者获益幅度呈下降趋势, 1981-1990 年试验的中位生存获益为 3.9 个月, 在 1991-2000 年试验的中位生存获益为 2.4 个月, 最近 10 年试验的中位生存获益为 2.5 个月; 对所有报道“结果阳性”的试验进行分析显示, 首个 10 年间生存获益为 3.9 个月, 第二个 10 年间生存获益为 2.0 个月, 2001-2010 年生存获益为 0.9 个月 (图 1)。

只有早于 1990 年的试验将存在统计学差异但临床生存获益不显著的结果报告为阴性。

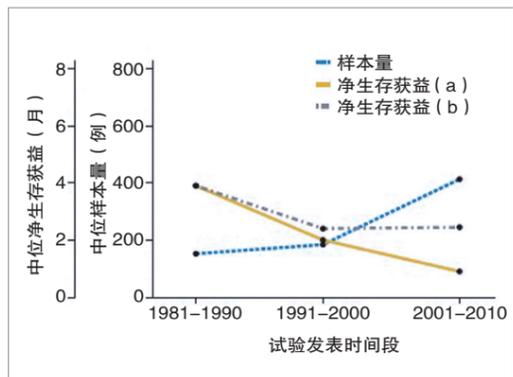


图 1 随时间推移中位样本量变化以及所有阳性结果试验 (a) 和统计学上达主要终点的试验 (b) 报告的净生存获益

研究者说

FDA 认可将 PFS 作为药物审批的潜在试验主要终点

自 1990 年至 2010 年, 使用 PFS 作为主要终点的试验比例从 0% 上升至 13%。虽然达到其设定的主要终点的试验比例仍保持不变, 但是报告阳性结果的试验比例却上升了。尽管随时间推移生存获益程度有所下降, 但是不具有统计学意义。这些发现引出了许多关于晚期非小细胞肺癌 III 期试验设计和解读的问题。

多阶段的临床试验旨在将资源投入到有希

望的药物上, 每个阶段都淘汰掉有毒性或药效最低的, 而该回顾性分析提示, 在晚期非小细胞肺癌试验中研究人员完成这一工作的效率日益下降。

晚期 NSCLC 临床试验推动了一线、二线化疗药物以及用于化疗失败后患者的靶向治疗药物的临床应用, 然而这些治疗药物的疗效一般。

随着肿瘤学团体和制药企业对 NSCLC 兴趣的增加, 试验设计和解

读也正在发生改变。这些改变包括主要终点和次要终点的选择、统计学功效和样本量大小。因为越来越多的患者在疾病进展后就会改用其他治疗方案, 并且治疗选择也越来越多, 所以 PFS 作为主要终点被提倡使用。尽管美国食品药品监督管理局 (FDA) 已认可将 PFS 作为晚期 NSCLC 治疗药物审批的潜在主要终点, 但是 FDA 对这种做法持保留态度。

除了临床试验设计

的改变, 试验解读方式也会发生改变。曾有报道称, 尽管一种新治疗方案的疗效极其有限或不确切, 但是越来越多的研究者愿意将这些结果解读为阳性。

该回顾性分析证实, 晚期非小细胞肺癌 III 期试验设计发生了令人堪忧的改变。在过去 30 年间, 尽管通过扩大样本量、提高统计功效使其设计水平得到提高, 但是判断肺癌新药临床获益的标准似乎正在降低。