

がん臨床試験で発生する区間打ち切りデータが ログランク検定の検出力に与える影響の評価

中外製薬株式会社 吉本拓矢* 杉谷康雄 浅川誉 仲川慎太郎 山本英晴

進行癌における臨床試験において、真の評価項目である全生存期間 (Overall Survival; OS) の代替評価項目として、無増悪生存期間 (Progression Free Survival; PFS) を主要評価項目に設定することは珍しくない。一般に PFS はランダム化から病勢進行又は死亡のうちいずれか早い方が最初に確認された日までの期間と定義され、群間で PFS の分布に差があるか否かを検定する手法として、多くの場合ログランク検定が選択される。この際、必要イベント数及び症例数の算出には PFS が指数分布すること、及び PFS イベントの発現とその結果の観察にタイムラグがないことを仮定することが一般的である。しかしながら、倫理的な配慮を含む実施上の理由から、実際には患者の病勢進行を確認するための検査は一定の間隔において定期的な実施され、RECIST ガイドライン¹では、治療期間中は 6~8 週ごと、治療終了後は 3~4 カ月ごとを検査間隔の例として挙げている。このことから、観察される PFS は実際に病勢進行が発生した日ではなく病勢進行が観察された検査日となり、区間打ち切りデータとなるため、検査間隔や検査回数が PFS のハザード比や生存時間の中央値の推定値或いはログランク検定の検出力に影響すると考えられる。Carroll²は検査間隔の影響を生存関数に指数分布を仮定したパラメトリックな手法で尤度を構成することで確認したが、全尤度をベースにした検討が現実には即しておらず一般化は困難であると言える。また、早期癌における術前・術後療法の評価に際しては無病生存期間 (Disease Free Survival; DFS) を主要評価項目とすることが一般的であるが、本質的に PFS と同様の問題を孕んでいる。

そこで、実際の臨床試験に則して症例登録期間や解析時期を考慮したシミュレーション研究によって、検査間隔や検査回数がログランク検定の検出力、ハザード比や Kaplan Meier 推定の中央値にどの程度影響するかを評価した。さらに、進行癌の末期治療では、PFS が短く、病勢進行前に死亡する症例も少なくない。PFS のイベントが死亡による場合は、区間打ち切りデータとならずイベント発現時点を正確に観察可能である。そのためこの時 PFS を、死亡までの時間 (Overall Survival; OS) と病勢進行までの時間の 2 変量の指数分布に従うと仮定した場合に、検査間隔や検査回数がどのように上述の項目に影響するかを評価した。

その結果、PFS が単一の指数分布に従うとした場合現実的な範囲の検査間隔や検査回数であれば検出力への影響は少なかったが、その一方で病勢進行までの期間と OS が 2 変量の指数分布に従うことを仮定し PFS が短いケースでは、検査間隔や検査回数が検出力に大きな影響を及ぼすことが確認された。

[1] New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1), E.A. Eisenhauer et al, European Journal of Cancer 45 (2009) 228-247.

[2] Analysis of progression-free survival in oncology trials: Some common statistical issues, Kevin J. Carroll, Pharmaceut. Statist. 2007; 6; 99-113