

Power comparison of tests of restricted mean survival time with log-rank test and generalized Wilcoxon test under various survival distributions

【背景】

Restricted Mean Survival Time (RMST) の差や比がハザード比に替わる効果の指標として注目されている。RMST の差や比の推定値は比例ハザード性が成立していない場合でも解釈が可能であり、臨床的にも理解しやすいという点で治療効果の推定方法としては優れているといえる。一方で、RMST の差や比の検定についての議論は十分になされておらず、ハザード比の推定に対応する検定手法としてこれまで用いられてきたログランク検定との性能評価も限られている。

【目的】

RMST の差と比による検定の検出力を、比例ハザード性が不成立の場合も含めた様々な生存分布の下でログランク検定および一般化 Wilcoxon 検定と比較する。

【方法】

比例ハザード性の下では、生存関数が指数分布に従うとし、ランダム打ち切りの割合・登録期間の長さ・対照群のハザードを変化させた場合の検出力への影響を検討する。

比例ハザード性が不成立の下では、生存関数が区分指数分布、ワイブル分布、対数ロジスティック分布などに従うとし、生存曲線が早期もしくは後期に開く場合を想定する。

それぞれのシナリオごとに RMST の差と比のワルド検定の検出力をログランク検定および一般化 Wilcoxon 検定と比較する。

【結果】

指数分布に基づく比例ハザード性の下では、RMST の差と比による検定はログランク検定より検出力が低下した。検出力の低下の程度は、ランダム打ち切りの割合・登録期間の長さ・対照群のハザードの大きさによって変化し、特に対照群のハザードが非常に小さく登録期間が長い場合には 30~40%程度の低下が確認された。

比例ハザード性が不成立の下では、生存曲線が早期に開く場合に RMST の差と比の検出力はログランク検定よりも高く、逆に生存曲線が遅れて開く場合には RMST の差と比の検出力はログランク検定よりも低いことが確認された。

【考察】

RMST の差と比の検定は、一般化 Wilcoxon 検定と同様に、イベントが起りやすく早期に生存曲線の群間差がつくような状況で有用であることが示唆された。逆にイベントが起りにくい状況や治療効果が遅れて見られるような状況においては、群間差の検定として RMST を用いることは検出力の観点では推奨されないといえる。