

論 文 要 旨

制限付き平均生存時間の推定手法間の性能比較

生物統計情報学コース

49-186603

木村 流星

背景

比例ハザード性（以下 PH）が成り立たない状況（以下 Non-PH）でも適切な解釈が可能な制限付き平均生存時間（以下 RMST）が注目されている。RMST の推定手法は、 Kaplan-Meier の直接積分法（以下 KM 積分）、Pseudo-observation（以下 PO）、Inverse probability censoring weighted（以下 IPCW）、Flexible parametric models（以下 FPM）が提案されているが、手法間の性能比較がほぼされていない。パラメトリックに推定する FPM は他手法と挙動が異なると考え、特定状況下で手法間の性能差を示すことを目的とした。

方法

登録 3 年、追跡 5 年の time-to-event 型のランダム化 2 群比較臨床試験を想定したシミュレーション研究を行った。PH およびがん免疫療法の臨床試験を想定した Non-PH の各状況において、各群の症例数、期待イベント発生割合を変えた複数のシナリオを検討した。境界時点 τ を 7 年に設定し、評価指標として、バイアス、標準誤差、被覆確率、平均二乗誤差および検出力または α エラーを用いた。PH のシナリオにおいてはスプライン関数の knot 数が 2 の FPM モデル、Non-PH のシナリオにおいては knot 数が 1 の交互作用項を加え、Non-PH に拡張した FPM モデルを用いた。また、KM 積分による RMST の推定を行うことができない場合、必要に応じて追跡期間を延長することで対処した。

結果

リスク集合が少ないシナリオにおいては、IPCW の標準誤差が他手法と比較して大きく推定された (KM 積分と比較して最大 0.026 の差)。イベントがある程度発生するシナリオにおいて、PH の状況では FPM の標準誤差が比較的小さく推定されたため検出力が高くなる傾向 (KM 積分と比較して最大 2.4% の差)、被覆確率が高くなる傾向 (0.950–0.961)、平均二乗誤差が小さくなる傾向 (KM 積分と比較して最大 0.022 の差)、そして α エラーが小さく、一部シナリオで 0.05 を下回る傾向がみられた (0.038–0.053)。同様にイベントがある程度発生するシナリオにおいて、治癒患者を含む遅発性効果型 Non-PH の場合、FPM の群間差に最大 0.017 の過大評価のバイアスと比較的高い検出力 (KM 積分と比較して最大 2.1% の差) がみられた。イベント数が極端に少ないシナリオにおいては、FPM のシミュレーションエラーが発生し群間差の平均二乗誤差が非常に大きくなる傾向がみられ (KM 積分と比較して最大 2.251 の差)、FPM のエラーがほぼ発生しないシナリオであっても各群の平均二乗誤差が大きくなる傾向がみられた (KM 積分と比較して最大 0.017 の差)。

結論

本研究によっていくつか手法間の性能差が示された。一般的な状況の臨床試験において極端にリスク集合が少ないときには IPCW の使用は控えるべきであり、PO、FPM の使用を推奨するが、共変量を調整する必要がない場合には KM 積分も選択肢に含まれるだろう。遅発性効果型が想定されるがん免疫療法の臨床試験においてイベント数が極端に少ない場合には FPM の使用は推奨できない。