

主 論 文 要 旨

No.1

報告番号	甲 乙 第	号	氏 名	木村 流星
主 論 文 題 名： 最小化法を用いたランダム化比較試験における再ランダム化検定とその信頼区間構成法に関する統計的考察				
(内容の要旨) 【第1章 序論】 臨床試験では治療群間の有効性や安全性を適切に比較するためにランダム化が実施される。Pocock-Simon の最小化法を用いた試験では、症例数に対して予後因子が多い場合があり、通常の漸近検定では第一種の過誤確率を名義有意水準に保持することが困難である。こうした状況でも第一種過誤確率に制御しつつ、検出力の低下を防ぐ方法として、再ランダム化検定が注目されている。 生存時間データの評価には、比例ハザード性の仮定が一般的に用いられるが、この仮定が成立しない状況がしばしばあり、比例ハザード性の仮定を前提としない解析手法が数多く提案されている。しかし、これらの解析手法を再ランダム化検定と組み合わせた場合の性能評価は不十分である。また、Garthwaite の方法として知られる再ランダム化検定に基づく信頼区間の構成法は、生存時間データには適用できない。生存時間データを用いる臨床試験では最小化法が用いられることが多いため、これらは解決すべき重要な課題である。本研究では、比例ハザード性の仮定に依存しない検定統計量に基づく再ランダム化検定を提案し、その性能を評価する（第2章）。また、幅広いアウトカムに適用可能な再ランダム化検定に基づく信頼区間を新たに提案する（第3章）。 【第2章 再ランダム化検定の性能評価】 再ランダム化検定は、最小化法を用いたランダム化比較試験において、漸近検定の代替または補助的手法として注目されている。しかし、比例ハザード性が成立しない場合、その性能や適切な検定統計量の選択については十分な検討がされてこなかった。 本研究では、比例ハザード性の仮定に依存しない統計量に着目し、逆確率重み付け法を用いて共変量を調整した制限付き平均生存時間の群間差（調整 dRMST）および Lin ら（2020）が用いた組み合わせである層別重み付き log-rank 検定統計量の最大値（層別 MaxCombo）を再ランダム化検定に拡張した手法を提案した。これら手法および既存法の性能を、非比例ハザードの状況下において最小化法を用いたランダム化比較試験を想定した数値実験を通じて検証した。その結果、調整 dRMST に基づく再ランダム化検定は、設定した境界時間の値に影響を受けやすい一方、臨床的観点から設定可能であり				

解釈しやすい効果指標を提供する有用なアプローチであることが示唆された。層別 MaxCombo に基づく再ランダム化検定は、広範なシナリオ下で一貫して優れた検出力を示し、ロバストな手法であった。臨床試験をデザインする際には解析手法を事前規定する必要があるため、非比例ハザード性が予想される臨床試験では、層別 MaxCombo が有用であると考えられる。

【第3章 再ランダム化検定に基づく信頼区間構成法の提案】

一般に、臨床試験では検定の結果に加えて推定値およびその信頼区間も報告することが求められる。再ランダム化検定に基づく信頼区間の構成法として Garthwaite の方法が知られているが、この方法は生存時間データには適用できない。

本研究では Garthwaite の方法を修正し、生存時間を含む多様なアウトカムおよび推定量に対して統一的に適用可能な再ランダム化検定に基づく信頼区間の構成法を提案した。数値実験を通じて、提案手法は一部のシナリオで打ち切りの影響により被覆確率が若干低下する可能性があるものの、漸近法と比較して優れた性能を示した。今後、再ランダム化検定を適用する際には、提案した信頼区間を併せて報告することが推奨される。

【第4章 考察】

再ランダム化検定の適用事例は限られているが、結果の解釈や意思決定における補助的なツールとして活用可能な手法である。再ランダム化検定の主解析および補助的な解析としての採用が進んでいない主な理由としては、検定統計量の選択指針や性能が十分に示されていないこと、さらにハザード比や dRMST に対しては再ランダム化検定に基づく信頼区間の構成法である Garthwaite の方法が適用できないことが考えられる。本研究はこれらの課題に取り組み、一部の問題を解決した。今後、実用化に向けて本研究の解析手法を R パッケージなどの形で一般公開することを検討する。また、提案した信頼区間構成法は、生存時間以外のアウトカムや多様な推定量にも適用可能であることから、生存時間以外のアウトカムや回帰に基づく条件付き期待値以外を含む様々な推定量に対する検討の拡充が可能である。

【第5章 結論】

再ランダム化検定に関する本研究の成果は、実務における再ランダム化検定の重要性を強調し、その適切な運用を通じて質の高いエビデンスの創出に寄与するものと考えられる。今後、提案手法の改良や統計ソフトウェアでの実装を進めるとともに、幅広いアウトカムおよび多様な推定量に対する性能評価が望まれる。