

レセプトデータにおける外部情報を用いた交絡調整法の実践

東京大学大学院 医学系研究科 公共健康医学専攻 生物統計学分野

竹内 由則

【研究の背景】

レセプトデータのような既存の医療情報データを二次利用して疫学研究を行なう場合、潜在的な交絡要因の情報が十分に取得されていないという限界がある。これに対処するため、小規模だが交絡要因の情報が豊富な部分集団 (Validation subsample, VS) を用いた交絡調整が行われる場合がある。最近、このような方法の一つとして、生存時間解析で頻用される Cox 比例ハザードモデルに基づく Martingale 残差を用いて欠測している交絡要因を補完する手法が提案された。一方、Aalen 加法ハザードモデルは、比例ハザード性の仮定が不要であり、時間依存性曝露効果を容易に推定できるなど、Cox モデルよりも多様な状況下で利用可能である。

【目的】

本研究では、Martingale 残差に基づく補完法を、より柔軟な生存時間解析において利用可能とするため、Aalen 加法ハザードモデルを用いる場合への拡張を行う。

【方法】

Cox モデルに対する Martingale 残差に基づく補完法では、欠測した交絡要因の補完モデルの説明変数として、集団の全てで測定されている変数のみを用いた Cox モデルから算出した Martingale 残差を用いる。提案法における変更点は、Aalen モデルによって算出された Martingale 残差を補完モデルの説明変数として用いる点である。

仮想データを用いたシミュレーション実験により、Aalen モデルに対する Martingale 残差に基づく補完法 (提案法) の性能評価を行った。まず、発生さ

せた共変量および曝露データに基づく加法ハザードモデルにより、仮想的な生存時間データ (N = 10,000) を生成し、全ての共変量の情報が得られた「VS (約 10%)」と、一部の共変量を意図的に欠測させた「Remaining sample, RS (約 90%)」に分割した。打ち切り時間は指数分布に基づいて発生させた。曝露は全てのデータで観測されているものとした。その後、両データセットで観測されている共変量のみを用いた Aalen モデル (Reduced model) および、提案法を用いて欠測データを補完した後に、全ての共変量を用いた Aalen モデル (Imputed model) を用いて曝露による累積ハザード差を推定した。想定されるいくつかの状況を再現した各シナリオについて、1,000 回の試行を実施し、それぞれのモデルの推定結果について、真値からのバイアス、最小二乗誤差、95%信頼区間の被覆確率を算出し、性能評価を行った。

さらに、提案法の実データへのアプリケーション事例を示すため、(株)JMDC のレセプトデータを用いて、スタチン系医薬品の糖代謝異常発症リスクを評価した。スタチン系医薬品は既存の糖代謝異常発症リスクに基づいて、高リスクスタチンと低リスクスタチンに分割し、それぞれ推定を行った。比較対象はスタチン系以外の脂質異常症治療薬とした。共変量として、性別、年齢、併用薬、既往歴および健診結果を用いた。レセプトデータセットのうち健診結果データが得られている集団を VS、得られていない集団を RS として扱った。コホート (VS+RS) および VS のみに対して、健診結果以外の共変量を用いた Aalen モデル (Reduced model) および、全ての共変量を用いた Aalen モデル (Full model) によって、いずれかの脂質異常症治療薬の初回処方か

ら糖代謝異常発症までの時間に対する生存時間解析を行った。

【結果】

シミュレーション実験の結果、Reduced model の推定値と比較して Imputed model の推定値のバイアスおよび最小二乗誤差は小さくなり、提案法によって RS における未測定交絡要因による残差交絡を軽減できることが示された。Reduced model の信頼区間被覆確率は名義水準を大きく下回ったが、Imputed model の被覆確率は名義水準を満たした。この結果は、曝露効果が null でない場合、曝露効果が null である場合、VS の割合が非常に小さい場合、VS の選択が Missing at random あるいは Missing not at random の場合、打ち切りがワイブル分布に従う場合、共変量の分布が RS と VS で異なる場合のいずれの状況下においても認められた。

実データ解析では、高リスクスタチンについては、VS・コホートいずれにおいても、Reduced model と比較して Full model で高い累積ハザード差が推定された。VS に対する Full model では、ほぼ全期間で累積ハザード差の点推定値が 0 以上だったが、信頼区間は概ね 0 を含んでいた。一方、提案法により補完後のコホートに対する Full model では信頼区間が狭くなり、全期間で糖代謝異常発症リスクの統計学的に有意な上昇が認められた。低リスクスタチンでは、VS・コホートいずれにおいても、Reduced model では、期間を経るごとに問代謝異常発症リスクが低下していく傾向が示された。しかし、Full model によって健診結果データの影響を考慮することでこの低下傾向は消失し、0 前後を推移する累積ハザード差が推定された。

【考察】

本研究の結果から、提案法によって RS で欠測している交絡要因を補完することで、集団全体を用いて多くの交絡要因を調整した解析を集団全体で行う

ことが可能となり、Reduced model による残差交絡によるバイアスや、VS のみを対象とした場合の検出力不足などの課題を克服できる可能性が示された。

しかし、提案法によって妥当な推定を行うためには、データの発生構造や、VS でのみ測定される共変量がアウトカムに与える効果の強さに関連した強い仮定が必要となる。そのため、提案法を利用するためには、これらの仮定をどの程度満たしているかについての留意が必要となる。

【結論】

小規模な VS の交絡要因情報を用いて RS における未測定交絡要因を調整する方法の一つとして Cox モデルに対して提案されていた、Martingale 残差に基づく欠測交絡要因の補完法は Aalen モデルにおいても利用できる可能性が示された。一方、提案法の正当化には依然多くの強い仮定が必要であることから、推定に必要な仮定を軽減するとともに、より幅広い状況で利用可能になることを目指した、更なる手法の改善を試みるべきと考えられた。