

1. 序文

被験者個人でなく、被験者の集団をランダムに分ける比較試験をクラスターランダム化試験 (Cluster Randomized Trial, CRT) という。感染リスクの共有や教育的介入後の被験者間の情報交換などにより被験者の独立性を保つことができない場合や、大規模な予防医学的研究や救急医療現場における介入研究のような、クラスター内で同じ介入を行うことで実施可能性が高まる場合に採用されるデザインで、同じ病院に通う患者や地域ごとの住人の集団をクラスターとすることが多い。

同一クラスターに属する被験者のエンドポイントには類似性が存在し、これをクラスター内相関と呼ぶ。クラスター内相関を無視すると、エンドポイントのばらつきを過小評価することになり α エラーが増大するため、クラスター内相関を考慮した計画と解析を行う必要がある。クラスター内相関を考慮する解析方法は「クラスターレベルの解析」と「個人レベルの解析」に大別できる。「クラスターレベルの解析」は、クラスターごとに割合や平均値などエンドポイントの要約指標を算出し、それを用いて群間比較を行う二段階の解析方法で、中でも最も仮定の少ない比較方法の一つが並べ替え検定である。「個人レベルの解析」は各被験者のデータを直接扱い、モデルによって被験者が属するクラスターを考慮する解析方法である。CRT の解析手法として標準的な手法が周辺モデルで、一般化推定方程式 (Generalized Estimating Equation, GEE) により集団全体の介入効果を推定するものである。条件付きモデルはクラスター効果を潜在変数としてモデルに導入した変量効果モデルのことで、クラスターごとの介入効果を推定する。両者はクラスター内相関のモデル化が異なるため、エンドポイントが連続変数でない場合には結果は一致しない。

標準的な解析手法である GEE の漸近性質はクラスター数がクラスター内被験者数に比べて十分多いことに依存する。そのためクラスター数が少ないほど介入効果の分散が過小評価され α エラーが増大することが問題となっており、修正方法が複数提案されている。また、変量効果モデルによる解析やクラスターレベルの解析は、検出力が低いとの指摘やバイアスが生じるとの指摘もある。並べ替え検定に対しては、GEE から導出した検定統計量を用いることで検出力を改良する方法も提案されているが、実用された報告はない。

近年、相関を考慮した計画と解析を行った CRT 研究は増加傾向にあるが、実施規模はクラスター数が 30 以下の研究が半数を占めると言われている。つまり GEE を中心に不適切な解析が行われている可能性があり、複数ある解析手法のいずれがどのような状況下で妥当な解析手法となるのか評価が必要である。

本研究の目的は、とくに漸近性質の成り立ちにくい 2 値データのエンドポイントの解析において、GEE、分散の過小評価を修正した GEE、変量効果モデル、並べ替え検定の性質をシミュレーション実験を通じて評価することで、クラスター数やクラスター内相関の異なる状況ごとに適した解析手法を明らかにすることである。

2. 方法

2.1. GEE による解析

クラスター i ($i=1, \dots, K$) の被験者 j ($j=1, \dots, n_i$) で観察される結果変数を Y_{ij} で表す。 Y_{ij} の期待値 μ_{ij} を割付群の指示変数 X_{ij} 、切片 λ 、介入効果 β によってロジスティックモデルでモデル化する。

$$\text{logit}(\mu_{ij}) = \lambda + \beta X_{ij}$$

CRT で興味のあるパラメータは介入効果 β で、 β の分散にはロバスト分散を用いることが通常である。しかしクラスター数が少ないほどロバスト分散は過小評価になる。そこで本研究では従来の GEE (ロバスト分散法) だけでなく、過小評価を修正するために提案された分散修正方法である Mancl 法と Kauerman 法の評価も行った。

2.2. 変量効果モデルによる解析

変量効果モデルはロジスティックモデルの線形予測子に正規分布に従うクラスター効果 u_i を変量として追加するモデルが一般的で、本研究でもこれを採用した。

$$\text{logit}(\mu_{ij}) = \lambda + \beta X_{ij} + u_i, \quad u_i \sim N(0, \sigma^2)$$

クラスター効果を与えたもとでクラスター内被験者を独立として扱い、介入効果を推定する。本研究における介入効果 β の推定には、結果変数 Y を 1 次近似して残差に正規残差を仮定するペナルティ付き擬似尤度法を用いた。

2.3. 並べ替え検定による解析

並べ替え検定は「どのクラスターもどちらの群に割り付けられても同じ結果が観察される」という帰無仮説を検定するものである。各クラスターの代表値に対し、どのような再割付も等確率で起こりうると考えて全割付パターンを発生させ、パターンごとの群間差の統計量で構成した分布を帰無分布とする。この帰無分布において実際観測された以上の群間差があるパターンの数を数え、その割合が名義水準以下のときに有意と判断する。本研究ではまずクラスター内被験者数で重み付けした割合を用いた並べ替え検定 (並べ替え 1) の評価を行った。さらに GEE から導出した修正並べ替え検定統計量による並べ替え検定 (並べ替え 2) も行い検出力改善の程度を検討した。

2.4. シミュレーション実験

本研究では2値のエンドポイントがクラスター内被験者それぞれに横断的に観測される状況において上記の解析手法をシミュレーション実験により評価した。評価指標には α エラー、検出力、推定値の平均バイアスと平均二乗誤差 (Mean Squared Error, MSE)、95%信頼区間の被覆確率を用いた。 α エラーを評価するデータは両群のイベント発生確率 μ_i を30%, 20%, 10%, 5%, 2%として発生させた。検出力を評価するデータは、両群の発生確率の組み合わせを(30%, 20%), (20%, 10%), (10%, 5%), (5%, 2%)として発生させた。クラスター内相関係数は $\rho=0, 0.05, 0.10, 0.15$ とし、クラスター内のすべての被験者同士の相関が共通になるようデータを発生させた。総被験者数は過去のCRTを参考に600人で固定し、クラスター数を6, 8, 10, 20, 30, 50, 100に設定した状況と、クラスター数を10, 20に固定し被験者数を300, 600, 900に設定した状況を検討した。さらに、クラスター数と総被験者数は変えず、クラスター内被験者数の分布の変動係数が0.5となるようにクラスター間で不均等にした状況下でも同様の評価を行った。

3. 結果

3.1. α エラー

両群のイベント発生確率が10%、クラスター内被験者数が均等な場合の α エラーを図1に示す。横軸Kがクラスター数、折れ線でつながれた4点は左からクラスター内相関係数 $\rho=0, 0.05, 0.10, 0.15$ の場合である。GEEのうちロバスト分散法はクラスター数が少なく、クラスター内相関が高いほど α エラーが増大した。GEEの修正方法はKauerman法よりもMancl法で α エラーの増大が抑制されたがクラスター数が20以下では5%を上回った。変量効果モデルはクラスター数、クラスター内相関に依存せず安定した挙動を示したのに対し、並べ替え検定は1, 2ともクラスター数が少ない場合、過度に保守的になる傾向が見られた。クラスター内被験者数が不均等な状況下ではロバスト分散法の α エラーが増加したが、特に変量効果モデルや並べ替え検定は影響を受けなかった。クラスター数を固定し被験者数を変化させた場合も、全ての手法で α エラーの増減は殆ど見られなかった。

3.2. 検出力

検出力は、クラスター内相関が大きいほど全手法で減少し、クラスター数が少ないほど低下が顕著であった(図2)。GEEのうち α エラーが名義水準に近いMancl法は相関に伴う検出力の低下が大きく、変量効果モデルを10%以上下回ることがあった。並べ替え検定は各手法の中で最も検出力が低く、クラスター数が6の場合は0%であった。クラスター内被験者数が不均等な場合、クラスター数が多くても並べ替え検定1の検出力は他より10%近く低い結果を示した。クラスター数を固定し被験者数を変化させた場合、被験者数の増加に伴い、検出力は高くなったが、手法間の相対的な関係は不変であった。

4. 考察

ロバスト分散法は、クラスター数が少ない場合だけでなく、多い場合でもクラスター内相関が存在すると名義水準通りの検定ができないことが示された。全体を通じて名義水準に近い検定ができる手法は Mancl 法、変量効果モデルと並べ替え検定で、この中で検出力が最も高いのは変量効果モデルであった。並べ替え検定は α エラーを増大させない手法として知られており、修正方法によって検出力に多少の改善はみられたが、クラスター数が少ない場合は過度に保守的になり実用的でないことが示された。クラスター数が 50 以上の場合はロバスト分散法以外の手法はいずれも同様の挙動を示した。クラスター内被験者数が不均等な CRT では検出力の低下や GEE に基づく手法で α エラーの増大が見られるため、試験開始時期を揃える、大きな地域を分割するといったクラスター内被験者数の不均等を防ぐ手法も重要であると考えられた。被験者数を変化させた場合の α エラーの挙動はどの手法でもほとんど変化がみられなかったことから、本研究の結果はある程度の一般化が可能であると考えられた。

5. 結論

クラスター数、クラスター内相関の程度によらず変量効果モデルを用いた解析が望ましいことが示唆された。研究対象集団全体における効果を推定する場合は、GEE の分散修正法である Mancl 法または並べ替え検定法による解析も適切な手法であることが示されたが、ロバスト分散法は用いるべきでないことが明らかとなった。

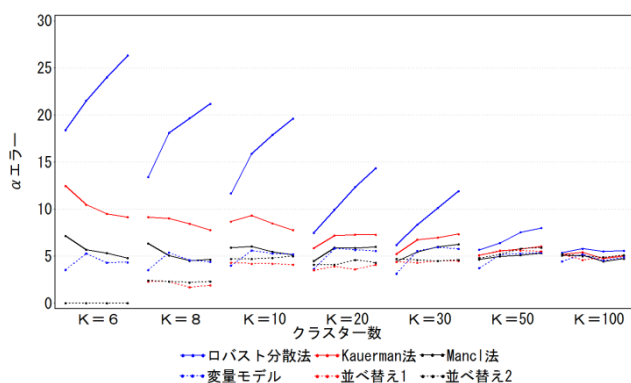


図 1. 被験者数が均等な場合の α エラー

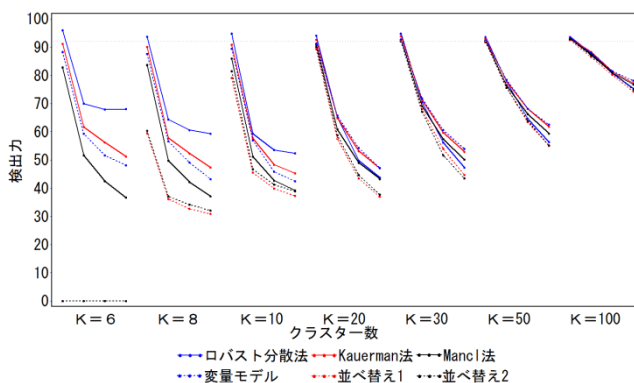


図 2. 被験者数が均等な場合の検出力