【企画セッション】欠測のあるデータにおける主解析の検討

(5)セッションのまとめ

土居正明[1)],
日本製薬工業協会 医薬品評価委員会　データサイエンス部会　タスクフォース4
欠測のあるデータの解析検討チーム.
1)東レ株式会社

【Planning session】Study to the candidates for the primary analysis for the longitudinal data with missing observations.

Summary of the session.

Masaaki Doi[1)],
The team for statistical analysis of data analysis with missing data, task force 4, data science expert committee, drug evaluation committee, Japan Pharmaceutical Manufacturers Association,
1) Toray Industries, Inc.

要旨：

企画セッションのまとめと，全体を通しての質疑応答を行う.

キーワード：MI, wGEE, Doubly Robust

今回の解析の結果（うつ病）(赤字：MMRMと同等)

MAR	αエラー（%）	検出力（%）	群間差の推定値の平均	群間差に対するMSE
MMRM	4.79	90.40	-2.9988	0.8250
MI (100回補完)	4.73	90.34	-2.9971	0.8168
wGEE(OS)	5.36	89.85	-2.9961	0.8606
DR	3.79	86.79	-2.9972	0.9373

MNAR	αエラー（%）	検出力（%）	群間差の推定値の平均	群間差に対するMSE
MMRM	4.84	90.53	-2.9688	0.8055
MI (100回補完)	4.68	90.49	-2.9650	0.7981
wGEE(OS)	5.23	89.83	-2.9446	0.8303
DR	3.52	87.04	-2.9487	0.9414

今回の解析結果（慢性疼痛）（赤字：MMRMと同等）

MCAR + MAR	αエラー（%）	検出力（%）	推定値の平均（真値：-1.00）	推定値のSD	MSE
MMRM	4.98	81.99	-1.0168	0.3526	0.1246
MI（補完100回）	4.20	77.94	-1.0035	0.3585	0.1285
wGEE(OS)	6.31	74.51	-0.9534	0.3754	0.1431
DR	3.93	76.66	-1.0306	0.3864	0.1502

MCAR + MNAR	αエラー（%）	検出力（%）	推定値の平均（真値：-1.00）	推定値のSD	MSE
MMRM	4.96	79.37	-0.9696	0.3488	0.1226
MI（補完100回）	4.18	74.62	-0.9533	0.3538	0.1274
wGEE(OS)	6.04	73.01	-0.9216	0.3665	0.1412
DR	3.57	75.31	-0.9818	0.3777	0.1429

MMRM と MI, wGEE, DRの比較（まとめ）

	αエラー		検出力		バイアス（検出力算出時）		推定値のSD（検出力算出時）	
	うつ病	慢性疼痛	うつ病	慢性疼痛	うつ病	慢性疼痛	うつ病	慢性疼痛
MMRM	5%弱	5%弱	90%程度	80%前後	0.00～0.04	-0.02～0.04	0.90前後	0.35前後
MI	5%弱	5%弱	同程度	-2～-3%	同程度	同程度	同程度	+0.01程度
wGEE(OS)	5%強	6%強	同程度	-6%程度	同程度	+0.05程度	+0.01～0.03	+0.02程度
DR	4%未満	4%未満	-3～4%	-4～5%	同程度	同程度	+0.04～0.07	+0.03程度

※うつ病はMAR, MNAR共通．慢性疼痛はMCAR+MNAR, MCAR+MNAR共通

◎MMRMとの比較

◎MI：慢性疼痛データで検出力がMMRMにやや劣る

◎wGEE (OS)：慢性疼痛データで　観測確率モデルの誤特定
→ バイアス　→　αエラー増大，検出力低下か？

◎DR：バイアスはMMRMと同程度。検出力はややMMRMに劣る.
→　推定量のばらつきが大きいのが原因か？

その他の特徴・検討結果

- **MI**
  - **補完回数**：5，10，20，50，100，200，500，1,000，1,500を検討
    - 今回のデータでは100回くらいまでは検出力単調増加
    - 100回くらいで飽和
    - データによる違いは大きいため，データ毎に検討が必要
- **wGEE**
  - **重み**：OS (Observation Specific) と SS (Subject Specific)の比較
    - 今回のデータでは，OSの方がαエラー，検出力の観点からよい性能
- **DR**
  - **分散推定**：Bootstrap
    - 時間の都合でBootstarp回数20回．回数を増やしたときの影響は未検討

現在のまとめ

- 今回想定した状況下ではMMRMの性能が最もよかった．しかし，
  - 検討している状況が非常に限られている．
  - 2つのデータも，違いが多く単純比較できない．
  - 現在扱ったデータは，MMRMに都合がよさそう．
    - 応答変数モデルは正しく特定
    - 観測確率モデルのみ誤特定されうる

今後の課題

・より詳細な状況設定のもとでの，各手法の性能評価
　例1)　MMRMの性能が落ちるときは？
　　そのときのその他の手法の性能は？
　例2) 他の疾患・プロトコールに合わせたデータでは？

・実データの解析

・非単調な欠測の場合の対応

→ JPMAで引き続き検討